Some examples of compact supermanifolds with non-Abelian fundamental group

Alice Rogers[a)]
Theoretical Physics Group, Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, England

(Received 5 December 1979; accepted for publication 8 February 1980)

A detailed construction is given of a supermanifold which is compact and has far from trivial topology. Higher dimensional examples are also described. The Grassmann algebra from which the supermanifolds are constructed may have finite or infinite dimension.

PACS numbers: 02.20.Nq, 02.40.Vh

1. INTRODUCTION

The purpose of constructing manifolds is to combine a fixed and manageable local structure with a variety of possible global topologies. A differentiable manifold is locally diffeomorphic to $\mathbb{R}^m$, while a supermanifold is locally "superdiffeomorphic" to ${B_L}^{m,n}$ (the Cartesian product of m copies of the even part and n copies of the odd part of B_L, the Grassmann algebra over $\mathbb{R}^L$). In the context of physics this local structure is motivated by "superspace", originally proposed by Salam and Strathdee.[1]

Precise definitions of supermanifolds vary, both in the topology used on ${B_L}^{m,n}$ and in the definition of "superdifferentiability" of the transition functions between overlapping coordinate neighborhoods. The definition used here is that developed by the author in Ref. 2, which includes other definitions[3] but also allows many further possibilities. The natural definition of "superdifferentiability" of functions of Grassmann elements, summarized below, is in effect much more restrictive than that for functions of real numbers, basically because Taylor series in nilpotent elements terminate. As a result the class of possible global topologies for supermanifolds (even with the broader definition) is more restricted than that for ordinary differentiable manifolds. (A comparable situation exists for complex manifolds.)

It is the purpose of the present paper to demonstrate that supermanifolds exist which are far from topologically trivial, and also that the broader definition of supermanifold adopted in Ref. 2 does substantially increase the possible global structures.

An (m,n) dimensional G^∞ supermanifold over B_L is defined[2] to be a Hausdorff topological space Y together with a set of charts (U_α,Ψ_α) such that (a) $\cup_\alpha U_\alpha = Y$; (b) Ψ_α is a homeomorphism of Y onto an open subset of ${B_L}^{m,n}$; and (c) $\Psi_\alpha \circ \Psi_\beta{}^{-1}: \Psi_\beta(U_\alpha \cap U_\beta) \to \Psi_\alpha(U_\alpha \cap U_\beta)$ is "G^∞" or "superdifferentiable," where in (b) the topology on ${B_L}^{m,n}$ is the usual topology on ${B_L}^{m,n}$ regarded as a finite-dimensional vector space and in (c) a function $f: U \subset {B_L}^{m,n} \to B_L$ is G^∞ if, given $(a_1,\ldots,a_{m+n})$ and $(a_1+h_1,\ldots,a_{m+n}+h_{m+n})$ in U,

$$f(a_1+h_1,\ldots,a_{m+n}+h_{m+n}) = f(a_1,\ldots,a_{m+n}) + \sum_{k=1}^{m+n} h_k (G_k f)(a_1,\ldots,a_{m+n})) + o(\|(h_1,\ldots,h_{m+n})\|)^2,$$

where the partial derivatives $G_k f$ are in turn differentiable functions of U into B_L. (More details of this definition can be found in Ref. 2).

2. A FIRST EXAMPLE

The first example is a (1,2) dimensional supermanifold over B_1 (constructed from ${B_1}^{1,2}$ very much as an Iwasawa manifold is constructed from C^3, Ref. 4). Suppose ${B_1}^{1,2}$ is identified with the set of 3×3 matrices of the form

$$\begin{pmatrix} 1 & x & y_1 \\ 0 & 1 & y_2 \\ 0 & 0 & 1 \end{pmatrix},$$

where x is an even element of B_1, and y_1 and y_2 are odd elements. This set of matrices forms a (non-Abelian) group under matrix multiplication, and thus ${B_1}^{1,2}$ acquires the structure of a Group G, summarized by

$$(x,y_1,y_2)(u,v_1,v_2) = (x+u, y_1+v_1+xv_2, y_2+v_2).$$

Letting 1 denote the unit element of B_1 and β the odd generator of B_1, an element of ${B_1}^{1,2}$ can be written $(a1,b_1\beta,b_2\beta)$ where a, b_1, and b_2 are real numbers. Define D to be the discrete subgroup of G consisting of elements $(m1,n_1\beta,n_2\beta)$ where m, n_1, n_2 are integers. Then G/D, the space of left cosets of D in G (with the quotient topology) can be given the structure of a G^∞ supermanifold, as is now described.

Let $[(x,y_1,y_2)]$ denote the coset containing (x,y_1,y_2); then $[(x,y_1,y_2)] = [(x',y_1',y_2')]$ if and only if there exist integers m, n_1, n_2 such that

$$x = x' + m1, \quad y_2 = y_2' + n_2\beta,$$
$$y_1 = y_1' + n_1\beta + x'n_2\beta.$$

Six subsets of ${B_1}^{1,2}$ are now specified, in terms of which eight coordinate neighborhoods on G/D are defined:

$$S_1 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, \tfrac{1}{5} < a < \tfrac{4}{5}\},$$
$$S_2 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, \tfrac{1}{5} < b_1 < \tfrac{4}{5}\},$$
$$S_3 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, \tfrac{1}{5} < b_2 < \tfrac{4}{5}\},$$

$$T_1 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, -\tfrac{2}{5} < a < \tfrac{2}{5}\},$$
$$T_2 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, -\tfrac{2}{5} < b_1 < \tfrac{2}{5}\},$$
$$T_3 = \{(a1,b_1\beta,b_2\beta) | a,b_1,b_2 \in \mathbb{R}, -\tfrac{2}{5} < b_2 < \tfrac{2}{5}\}.$$

Letting $V_1 = S_1 \cap S_2 \cap S_3$, $V_2 = T_1 \cap S_2 \cap S_3$, $V_3 = S_1 \cap T_2 \cap S_3$, $V_4 = S_1 \cap S_2 \cap T_3$, $V_5 = T_1 \cap T_2 \cap S_3$, $V_6 = T_1 \cap S_2 \cap T_3$,

a)Supported by the Science Research Council.

 022-2488/81/030443-02$1.00

$V_7 = S_1 \cap T_2 \cap T_3$, $V_8 = T_1 \cap T_2 \cap T_3$, the eight coordinate neighborhoods $U_i \subset G/D$ are defined to be $U_i := [V_i]$ $(i = 1,...,8)$ and the eight coordinate maps $\Psi_i : U_i \to V_i \subset B_1^{1,2}$ by $\Psi_i([p]) = p$ $(i = 1,...,8)$. The U_i cover G/D since, given any (x,y_1,y_2) in $B_1^{1,2}$, there exist integers m, n_1, n_2 (not necessarily unique) such that

$$x = a1 + m1, \quad y_1 = b_1\beta + n_1\beta + an_2\beta, \quad y_2 = b_2\beta + n_2\beta$$

with $-2/5 < a, b_1, b_2 < 4/5$. The maps Ψ_i are well defined because no individual V_i contains two distinct members of the same coset, and they are evidently homeomorphisms of the U_i onto the V_i which are open subsets of $B_1^{1,2}$.

It must now be established that the transition functions $\Psi_i \circ \Psi_j^{-1} : \Psi_j(U_i \cap U_j) \to \Psi_i(U_i \cap U_j)$, $i, j = 1,...,8$ (of which there are 56 to be considered) are all G^∞ (superdifferentiable). One example, $\Psi_6 \circ \Psi_1^{-1}$, is calculated in detail here. Suppose $[(x,y_1,y_2)] \in U_1 \cap U_6$. Then $(x,y_1,y_2) = ((a+m)1, b_1\beta + n_1\beta + an_2\beta, b_2\beta + n_2\beta)$, where $\frac{1}{5} < a < \frac{4}{5}, \frac{1}{5} < b_1 < \frac{4}{5}$, and $\frac{1}{5} < b_2 < \frac{4}{5}$ and also $(x,y_1,y_2) = ((c+m')1, d_1\beta + n_1'\beta + cn_2'\beta, d_2\beta + n_2'\beta)$, where $-\frac{2}{5} < c < \frac{2}{5}$, $\frac{1}{5} < d_1 < \frac{4}{5}$, and $-\frac{2}{5} < d_2 < \frac{2}{5}$. Thus $(x,y_1,y_2) = ((a+m)1, b_1\beta + n_1\beta + an_2\beta, b_2\beta + n_2\beta)$, where either

$$\text{case (1): } \tfrac{1}{5} < a < \tfrac{2}{5}, \quad \tfrac{1}{5} < b_1 < \tfrac{4}{5}, \quad \text{and} \quad \tfrac{1}{5} < b_2 < \tfrac{2}{5};$$

$$\text{or case (2): } \tfrac{3}{5} < a < \tfrac{4}{5}, \quad \tfrac{1}{5} < b_1 < \tfrac{4}{5}, \quad \text{and} \quad \tfrac{1}{5} < b_2 < \tfrac{2}{5};$$

$$\text{or case (3): } \tfrac{1}{5} < a < \tfrac{2}{5}, \quad \tfrac{1}{5} + a < b_1 < \tfrac{4}{5}, \quad \text{and} \quad \tfrac{3}{5} < b_2 < \tfrac{4}{5};$$

$$\text{or case (4): } \tfrac{3}{5} < a < \tfrac{4}{5}, \quad \tfrac{1}{5} < b_1 < a - \tfrac{1}{5}, \quad \text{and} \quad \tfrac{3}{5} < b_2 < \tfrac{4}{5}.$$

Therefore the set $\Psi_1\{U_1 \cap U_6\}$ is the union of four disjoint subsets of $B_1^{1,2}$, these sets consisting of points $(a_1, b_1\beta, b_2\beta)$ in $B_1^{1,2}$ with a, b_1, and b_2 satisfying one of the four sets of conditions listed above; also

$$\text{in case (1) } \Psi_6 \circ \Psi_1^{-1}(p,q_1,q_2) = (p,q_1,q_2);$$

$$\text{in case (2) } \Psi_6 \circ \Psi_1^{-1}(p,q_1,q_2) = (p-1,q_1,q_2);$$

$$\text{in case (3) } \Psi_6 \circ \Psi_1^{-1}(p,q_1,q_2) = (p,q_1 - p\beta, q_2 - \beta);$$

$$\text{in case (4) } \Psi_6 \circ \Psi_1^{-1}(p,q_1,q_2) = (p-1, q_1 - (p+1)\beta, q_2 - \beta).$$

The transition function $\Psi_6 \circ \Psi_1^{-1}$ is thus clearly G^∞, and in fact G^ω (superanalytic). The other transition functions have been calculated and also proved to be G^ω. Thus G/D has the structure of a G^ω supermanifold.

Since the group G can also be regarded as a simply connected three-dimensional Lie group, with D a discrete subgroup, G/D can be regarded as a three-dimensional real analytic manifold, (with fundamental group isomorphic to D, which is non-Abelian). However, G/D regarded as a supermanifold (with supermanifold structure defined as above) will have the same topology as G/D regarded as a manifold, and thus will also have fundamental group isomorphic to D. Also G/D can be mapped homeomorphically onto a closed and bounded subset of $\mathbb{R}^6$, and thus shown to be compact.

3. FURTHER POSSIBILITIES

(1,2) dimensional supermanifolds over B_L for any finite value of L, and also over the infinite-dimensional algebra B_∞ defined in Ref. 2, can easily be constructed in a similar manner by starting with the basic group structure

$$(x,y_1,y_2)(u,v_1,v_2) := (x+u, v_1+y_1+xv_2, y_2+v_2)$$

once again and considering the quotient of this group by the discrete subgroup D consisting of elements (p,q_1,q_2) of $B_L^{1,2}$ with p, q_1, and q_2 "Grassmann integers," that is

$$p = m_1 1 + m_2\beta_1\beta_2 + \cdots,$$

$$q_1 = n_1\beta_1 + n_2\beta_2 + n_3\beta_3 + n_4\beta_1\beta_2\beta_3 + \cdots,$$

$$q_2 = n_1'\beta_1 + n_2'\beta_2 + n_3'\beta_3 + n_4'\beta_1\beta_2\beta_3 + \cdots,$$

where the m_i, n_i, and n_i' are integers and $\beta_1, \beta_2, ..., \beta_L$ are the odd generators of the Grassmann algebra. In each case the supermanifold constructed is compact and has fundamental group isomorphic to D.

Higher dimensional supermanifolds can also be constructed in a similar manner by considering multiplicative groups of upper triangular matrices of higher order. For instance there is a (3,3) dimensional supermanifold G/D, where G is the group of 4×4 matrices of the form

$$\begin{pmatrix} 1 & u_1 & v_1 & u_2 \\ 0 & 1 & v_2 & u_3 \\ 0 & 0 & 1 & v_3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

under matrix multiplication, with u_1, u_2, u_3 even elements of B_L and v_1, v_2, v_3 odd elements, and D is the discrete subgroup of G consisting of matrices whose elements are all Grassmann integers. Again, all such supermanifolds will be compact and have non-Abelian fundamental group.

The definition of supermanifold used here is broader than the others[3] largely because a finer topology is used on $B_L^{m,n}$; in fact the coarser topology precludes the existence of compact, Hausdorff supermanifolds such as those constructed here (see Ref. 2, proposition 3.4), and it is thus demonstrated that the broader definition does considerably extend the range of possible global topologies for supermanifolds.

ACKNOWLEDGMENT

I am grateful to Chris Isham for helpful discussions.

[1] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974).

[2] A. Rogers, ICTP/78-79/15, J. Math. Phys. (to be published).

[3] M. Batchelor, Trans. Am. Math. Soc. (to be published); F.A. Berezin and D.A. Leites, Sov. Math. Dokl. **16**, 1218 (1975); B. DeWitt (in preparation); B. Kostant, in *Differential Geometric Methods in Mathematical Physics*, Proc. symposium, Bonn, July 1975, Vol. 570 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1977).

[4] S.S. Chern, *Complex Manifolds Without Potential Theory* (Van Nostrand, Princeton, New Jersey, 1967).

Conserved densities for nonlinear evolution equations. II. Odd order case

L. Abellanas and A. Galindo
Department of Theoretical Physics, Universidad Complutense de Madrid, Madrid-3, Spain

(Received 7 August 1979; accepted for publication 15 November 1979)

This paper extends to nonlinear evolution equations of odd order the analysis of existence and structure of the polynomial conserved densities. The results for low order densities are similar to the case of even order. The situation for densities with high order derivatives is now radically different. An asymptotic algorithm is presented for the search of such densities, which are shown to be quadratic in the highest derivatives. The very existence of just one high order conserved density is shown to severely restrict the evolution equation, and in the third order case it leads, with some minor additional hypothesis, to the KdV family.

PACS numbers: 02.30.Jr

1. INTRODUCTION

In a previous paper[1] the polynomial conserved densities for general evolution equations of even order were analyzed in detail. As indicated therein, the case of odd order requires a special treatment since the existence of nontrivial conserved densities of arbitrarily high order in the field derivatives cannot be excluded in this case (recall, for instance, the Korteweg–de-Vries equation[2]).

The present paper discusses some general properties of the polynomial evolution equations $u_t = P(u,...,u_M)$ of odd order. Concerning densities of low order, the situation is much alike to the even order case,[1] as proved in Sec. 2, where criteria are given for the existence of one or two nontrivial conserved densities of this class. Sections 3 and 4 are devoted to analyze high order nontrivial densities. In Sec. 3 some necessary conditions on P are obtained for at least one such density $\rho(u,...,u_N)$ to exist, and the quadratic dependence of ρ on the two highest derivatives u_N, u_{N-1} is established by direct computation. Section 4 presents an asymptotic algorithm which allows us, first, to show this quadratic character in full generality, secondly, to limit the order of derivatives which may enter the coefficients of these quadratic terms, and finally, to give an explicit differential system for such coefficients. The integrability conditions for this system impose strong restrictions on P. As an illustration, we prove in Sec. 5 that the KdV equation or its modified form are singled out among the third order nonlinear evolution equations $u_t = u_3 + f(u,u_1,u_2)$ as simultaneously possessing one very low and one high order nontrivial conserved densities.

Notation

We will keep to the notation in Ref. 1. The symbol u_M denotes the M-derivative of $u = u(x,t)$ with respect to the real variable x: $u_M = \partial^M u/\partial x^M = D^M u$, where D, in general, stands for the total x-derivative. By $\mathcal{F}_M$ we shall denote the set of sufficiently smooth functions $F(u,u_1,...,u_M)$. Similarly

$$\mathcal{F}'_M \equiv \{F \in \mathcal{F}_M : F_{u_M} \equiv \partial F/\partial u_M \neq 0\},$$
$$\mathcal{F}''_M \equiv \{F \in \mathcal{F}_M : F_{u_M u_M} \equiv \partial^2 F/\partial u_M^2 \neq 0\}.$$

Two densities $\rho, \bar{\rho}$ will be called equivalent ($\rho \sim \bar{\rho}$) whenever $\rho - \bar{\rho} = Q_1$ ($\equiv DQ$), i.e., if $\rho - \bar{\rho}$ is trivial. Finally $C_N(P)$ will stand for the set of polynomial conserved densities $\rho \in \mathcal{F}''_N$ under $u_t = P$.

2. CONSERVED DENSITIES OF LOW ORDER

Let us consider the (nonlinear) evolution equation

$$u_t = P(u,...,u_M), \quad M \text{ odd}, \tag{1}$$

where P is a polynomial in $\mathcal{F}'_M$. We are going to see that, concerning the low order ($N < M/2$) conserved densities of (1), many of the results obtained in Ref. 1 for even M still hold in the case of odd M. To begin with, we state the following

Theorem 1: $P_{u_M u_M} \neq 0 \Rightarrow \nexists\, \rho \in C_N(P)$, $N < M/2$.

Proof: $\rho_t \sim (\delta\rho/\delta u)P = Q(\cdots,u_{M-1})_1$ requires P to be linear in u_M.

Remark: The major difference between the odd and the even case lies in the fact that the existence of $\rho(\cdots,u_M)$, $N > M/2$, is only possible for odd M (remember the KdV equation!). To further emphasize this point let us observe that even the conclusion of Theorem 1 fails for $N > M/2$. As a matter of fact the equation $u_t = u_x^2$ admits the conserved densities $\rho = u_x^k$, $\forall k \geqslant 1$.

Given a polynomial P, linear in u_M, let us decompose it in the form

$$P = au_M + b = A_1 + B, \quad a,b,B \in \mathcal{F}_{M-1}, \; A \in \mathcal{F}'_{M-1}. \tag{2}$$

Then we have the following:

Criteria:

(i) $0 \neq d(B) \neq d(A)$, $M \geqslant 3$
$\Rightarrow \forall$ nontrivial conserved $\rho(\cdots,u_{(M-3)/2}) \sim \lambda u$.

(ii) $A_{u_{M-1}u_{M-1}} \neq 0 = B_{u_{M-1}u_{M-1}}$, $M \geqslant 3$
$\Rightarrow \forall$ nontrivial conserved $\rho(\cdots,u_{(M-3)/2}) \sim \lambda u$,

where $d(\cdot)$ denotes the degree of the corresponding polynomial in the variable u_{M-1}.

Proof: (i) The conservation of ρ requires the existence of $Q \in \mathcal{F}'_{M-2}$ such that

 022-2488/81/030445-04$1.00

$$\rho_t \sim \frac{\delta\rho}{\delta u} A - \left(\frac{\delta\rho}{\delta u}\right)_1 B = Q_1 . \tag{3}$$

By comparing the powers of u_{M-1} we conclude (i).

(ii) The coefficient of u_{2M-2} in

$$\frac{\delta}{\delta u}\left[\frac{\delta\rho}{\delta u} A - \left(\frac{\delta\rho}{\delta u}\right)_1 B\right] = 0, \tag{4}$$

leads to the condition

$$\left(\ln \frac{\delta\rho}{\delta u}\right)_1 = B_{u_{M-1}u_{M-1}} / A_{u_{M-1}u_{M-1}} \tag{5}$$

so that (ii) follows at once. ∎

Finally we state a necessary condition for the coexistence of (at least) two conserved densities of low order.

Theorem 2: Let $\rho, \bar{\rho}$ be two nontrivial conserved densities for (1), $M \geqslant 3$, of orders $\leqslant (M-3)/2$, and such that $\bar{\rho} \not\sim \lambda\rho$, $\forall \lambda \in \mathbb{R}$. Then

(a) $P_{u_M u_{M-1}} = 0$,

(b) $b_{u_{M-1}u_{M-1}} = 2a_{u_{M-2}}$.

Proof: (a) Since $\rho, \bar{\rho}$ are nontrivial and conserved under $u_t = P$ there must exist two functions $F, G \in \mathscr{F}'_{M-1}$ such that

$$P = \frac{F_1}{\delta\rho/\delta u} = \frac{G_1}{\delta\bar{\rho}/\delta u} . \tag{6}$$

Let us define

$$g \equiv \alpha F - G, \quad \alpha \equiv \frac{\delta\bar{\rho}}{\delta u} \Big/ \frac{\delta\rho}{\delta u} . \tag{7}$$

Obviously $\alpha \in \mathscr{F}_{M-3}$ and moreover, since $\bar{\rho} \not\sim \lambda\rho$,

$$g_1 = \alpha_1 F \in \mathscr{F}'_{M-1} . \tag{8}$$

Hence $g \in \mathscr{F}'_{M-2}$ and F must be linear in u_{M-1}. It follows that

$$P = F_1 \Big/ \frac{\delta\rho}{\delta u} = h(\cdots, u_{M-2}) u_M + k(\cdots, u_{M-1}), \tag{9}$$

and thus (a) is proved.

(b) Let $F = S(\cdots, u_{M-2}) u_{M-1} + T(\cdots, u_{M-2})$, $P = a(\cdots, u_{M-2}) u_M + b(\cdots, u_{M-1})$. From $F_1 = (\delta\rho/\delta u)P$ we get

$$a \frac{\delta\rho}{\delta u} = S \Rightarrow a_{u_{M-2}} \frac{\delta\rho}{\delta u} = S_{u_{M-2}} , \tag{10}$$

$$b \frac{\delta\rho}{\delta u} = S_{u_{M-2}} u_{M-1}^2 + V(\cdots, u_{M-2}) u_{M-1} + W(\cdots, u_{M-2}).$$

So the conclusion follows at once. ∎

3. CONSERVED DENSITIES OF HIGH ORDER

Let $\rho \in \mathscr{F}''_N$, $N > M+1 \geqslant 4$, be a conserved density under (1). As it is well known this implies $(\delta/\delta u)\rho_t = 0$. Now, it is not difficult to show by direct computation that

$$\rho_t = \sum_{j=0}^{N} \rho_{u_j} P_j = A u_{M+N} + B u_{M+N-1} + \hat{C}, \tag{11}$$

with

$$A \equiv \rho_{u_N} P_{u_M}, \quad B \equiv \rho_{u_N}[P_{u_{M-1}} + N(P_{u_M})_1] + \rho_{u_{N-1}} P_{u_M} . \tag{12}$$

The terms included in $\hat{C}$ do not contribute to the coefficients of u_{M+2N}, u_{M+2N-1} in $(\delta/\delta u)\rho_t$, so that the only nontrivial contributions to these two coefficients come from:

$$(-1)^{M+N} D^{M+N} \frac{\partial}{\partial u_{M+N}} \to (-1)^{M+N} A_{u_N} u_{M+2N} + (-1)^{M+N}[A_{u_{N-1}} + (M+N)(A_{u_N})_1] u_{M+2N-1} ,$$

$$(-1)^{M+N-1} D^{M+N-1} \frac{\partial}{\partial u_{M+N-1}} \to (-1)^{M+N-1} B_{u_N} u_{M+2N-1} ,$$

$$(-1)^N D^N \frac{\partial}{\partial u_N} \to (-1)^N A_{u_N} u_{M+2N} + (-1)^N [B_{u_N} + N(A_{u_N})_1] u_{M+2N-1} ,$$

$$(-1)^{N-1} D^{N-1} \frac{\partial}{\partial u_{N-1}} \to (-1)^{N-1} A_{u_{N-1}} u_{M+2N-1} .$$

The coefficient of u_{M+2N} turns out to be $[1 + (-1)^M] A_{u_N}$, which is automatically zero. On the other hand, the vanishing of the coefficient of u_{M+2N-1} requires

$$M(\rho_{u_N u_N})_1 P_{u_M} - 2\rho_{u_N u_N} P_{u_{M-1}} = (2N-M)\rho_{u_N u_N}(P_{u_M})_1 . \tag{13}$$

Consequently $\rho_{u_N u_N} \in \mathscr{F}_M$, and therefore

$$\rho \sim \tfrac{1}{2} a^{(0)}(\cdots, u_M) u_N^2 + b(\cdots, u_{N-1}), \quad N > M+1 \geqslant 4. \tag{14}$$

The next proposition quotes some other interesting consequences of (13):

Proposition 1: Let $N > M+1 \geqslant 4$. Then:

(i) $\dfrac{\delta}{\delta u}\dfrac{P_{u_{M-1}}}{P_{u_M}} \neq 0 \Rightarrow \nexists\, \rho \in C_N(P)$;

(ii) $\dfrac{\delta}{\delta u}\dfrac{P_{u_{M-1}}}{P_{u_M}} = 0 \Rightarrow a^{(0)} = k Q^{2/M} P_{u_M}^{2N/M-1}$,

where k is a constant and Q is defined by $(\ln Q)_1 = P_{u_{M-1}} / P_{u_M}$;

(iii) $P_{u_M} = \text{constant} (\neq 0)$, $P_{u_{M-1}} \neq 0 \Rightarrow \nexists\, \rho \in C_N(P)$.

Assertion (iii) follows from the polynomial character of ρ.

It is a simple exercise to prove that for $N > M+4 \geqslant 9$

$$\rho_t = A u_{M+N} + B u_{M+N-1} + C u_{M+N-2} + D u_{M+N-3} + \hat{E}, \tag{15}$$

with

$$C \equiv \rho_{u_N} \sum_{j=0}^{2} \binom{N}{j} (P_{u_{M+j-2}})_j + \rho_{u_{N-1}}[P_{u_{M-1}} + (N-1)(P_{u_M})_1] + \rho_{u_{N-2}} P_{u_M} ,$$

$$D \equiv \rho_{u_N} \sum_{j=0}^{3} \binom{N}{j} (P_{u_{M+j-3}})_j + \rho_{u_{N-1}} \tag{16}$$

$$\times \sum_{j=0}^{2} \binom{N-1}{j} (P_{u_{M+j-2}})_j$$
$$+ \rho_{u_{N-2}} [P_{u_{M-1}} + (N-2)(P_{u_M})_1] + \rho_{u_{N-3}} P_{u_M}.$$

The terms included in $\hat{E}$ do not contribute to the coefficients of $u_{M+2N},\dots,u_{M+2N-3}$ in $(\delta/\delta u)\rho_t$. The vanishing of the coefficient of u_{M+2N-2} provides no new information. As to the coefficient of u_{M+2N-3} it leads to the condition:

$$\frac{M^2+2}{6}[A_{u_{N-1}} - B_{u_N}]_2 + \frac{M}{2}[A_{u_{N-1}} + B_{u_N}]_2$$
$$+ M[A_{u_{N-2}} - B_{u_{N-1}} + C_{u_N}]_1 + 2[(A_{u_{N-2}})_1 - (C_{u_N})_1$$
$$+ A_{u_{N-3}} - B_{u_{N-2}} + C_{u_{N-1}} - D_{u_N}] = 0. \tag{17}$$

The vanishing of the coefficient of u_N in (6) requires, having in mind (14), that:

$$P_{u_M} b_{u_{N-1}u_{N-1}u_{N-1}} = 0. \tag{18}$$

In other words

$$\rho \sim \tfrac{1}{2} a^{(0)}(\dots,u_M)u_N^2 + \tfrac{1}{2} a^{(1)}(\dots,u_{M+2})u_{N-1}^2 + c(\dots,u_{N-2}). \tag{19}$$

In fact, one can easily prove that $a^{(1)} = \beta(\dots,u_M)u_{M+2} + \gamma(\dots,u_{M+1})$, but we omit the details.

4. A PRACTICAL ALGORITHM OF ASYMPTOTIC CHARACTER ($N \gg 1$)

Two remarkable suggestions arise from the preceding section. First of all, one suspects that for $N \gg 1$ every conserved density $\rho \in C_N(P)$ should be quadratic in the highest derivatives. And on the other hand formula (15) could possibly admit a simple generalization when $N \gg 1$, which might prove useful when computing $\delta\rho_t/\delta u$.

Indeed it is rather easy to show that given $q \geqslant 0$, there exists $m_0(q)$ such that

$$P_m - \sum_{j=0}^{q} u_{M+m-j}\left[\sum_{k=0}^{j}\binom{m}{k}(P_{u_{M+k-j}})_k\right]$$
$$\in \mathscr{F}_{M+m-q-1}, \quad \forall m \geqslant m_0(q). \tag{20}$$

Relations of the form (20), which express equality only in the dominant terms containing $u_{M+m}, u_{M+m-1},\dots, m \gg 1$, will be simply written in the sequel as

$$P_m \doteq \sum_{j>0} u_{M+m-j}\left[\sum_{k=0}^{j}\binom{m}{k}(P_{u_{M+k-j}})_k\right]. \tag{20'}$$

With this notation, the generalization of (15) for a $\rho \in C_N(P)$, $N \gg 1$, turns out to be

$$\rho_t \doteq \sum_{j>0} A^{(j)} u_{M+N-j},$$
$$A^{(j)} \equiv \sum_{i=0}^{j} \rho_{u_{N-i}} \sum_{k=0}^{j-i}\binom{N-i}{k}(P_{u_{M+i-j+k}})_k. \tag{21}$$

Therefore, the vanishing of $\delta\rho_t/\delta u$ leads to the equation

$$\sum_{n>0} u_{M+2N-n}\left[\sum_{j=0}^{n}\sum_{m=0}^{n-j} c_{j,m}^{(n)}\{A^{(j)}_{u_{N-n+j+m}}\}_m\right] \doteq 0,$$
$$c_{j,m}^{(n)} \equiv (-1)^j\left[\binom{M+N-j}{m} + (-1)^{M+n+m}\right.$$
$$\left.\times\binom{N-n+j+m}{m}\right]. \tag{22}$$

The coefficients of u_{M+2N-n}, $n = 0,1,2,\cdots(n \ll N)$ in (22) must be zero, so that after substitution of $A^{(j)}$ by their expressions (21) one finds the conditions

$$\sum_{j=0}^{n}\sum_{m=0}^{n-j} c_{j,m}^{(n)}\left[\left\{\sum_{i=0}^{j}\rho_{u_{N-i}}\sum_{k=0}^{j-i}\binom{N-i}{k}\right.\right.$$
$$\left.\left.[P_{u_{M+i-j+k}}]_k\right\}_{u_{N+j+m-n}}\right]_m = h^{(n)}, \tag{23}_n$$

with

$$h^{(n)} \equiv \begin{cases} 0, & n = 1,2,\dots,M-1,M+1,M+3,\dots, \\ & M+2r+1,\cdots, \\ (-1)^r \sum_{i>0} a_{u_i}^{(r)} P_i, & a^{(r)} \equiv \rho_{u_N u_{N-r}}, \\ & n = M+2r,\ r \geqslant 0 \text{ sufficiently small.} \end{cases}$$

Remark that $(23)_n$ holds only for $n \ll N$. More explicitly, given n, $\exists N_0(n)$ such that $(23)_n$ holds whenever $N \geqslant N_0(n)$. We are now in position to prove that ρ must be quadratic in its highest derivatives. We proceed by induction. Since it is known to be true for the first two terms [see (19)], let us suppose that

$$\rho = \tfrac{1}{2}\sum_{j=0}^{q-1} a^{(j)} u_{N-j}^2 + R(\dots,u_{N-q}), \quad q \ll N, \tag{24}$$

where the functions $a^{(j)}$ depend on derivatives of low orders ($\ll N$). We want to prove that R must be quadratic in u_{N-q}.

First of all we observe that $h^{(n)}$ depends only on low order derivatives for $n \ll N$, so we can forget them in this argument.

Let us see which are the contributions in $(23)_{2q+1}$ to the coefficient of u_{N-q+1}. Their sum must be zero, because it is nothing but the coefficient of $u_{N-q+1}u_{M+2N-2q-1}$ in $\delta\rho_t/\delta u$. Now, the quadratic terms $\frac{1}{2}a^{(j)}u_{N-j}^2$, $0 \leqslant j \leqslant q-1$, make no contribution since $[a^{(j)}]_m$ has only low order derivatives. Hence the only contribution will come from R, due to the terms $i = q, q+1,\dots,2q+1$, in $(23)_n$. Let us put $i = q+\alpha$, $0 \leqslant \alpha \leqslant q+1$. Then the summation in k forces j to be $\geqslant q+\alpha$. And furthermore, the partial derivative with respect to $u_{N+j+m-n}$ requires $2q+1-j-m \geqslant q$, so that finally we conclude

$$q+\alpha+m \leqslant j+m \leqslant q+1, \quad \text{i.e.,}\ \alpha+m \leqslant 1.$$

Moreover the only way to get u_{N-q+1} in $(23)_{2q+1}$ is by means of the final D^m derivative with $m > 0$. Thus $\alpha = 0$, $m = 1$, $i = j = q$, which obviously contributes $c_{q,1}^{(2q+1)} P_{u_M} R_{u_{N-q}u_{N-q}u_{N-q}}$ to the u_{N-q+1} term. Therefore $R_{u_{N-q}u_{N-q}u_{N-q}} = 0$, as announced. (Remark that $c_{q,1}^{(2q+1)} = (-1)^q M \neq 0$, $\forall q$.)

In the discussion above the precise range of $a^{(j)}$ in the derivatives has not been fixed. We are now able to show that $a^{(j)} \in \mathscr{F}_{M+2j}$, a fact which was already made plausible in the preceding section. Indeed, for $a^{(p)}$ to appear in $(23)_n$ the only possibilities are:

(1) $i = N-s$, $s \ll N$. But then $n \simeq N$.

(2) $i = p$ and hence $n \geqslant j \geqslant p$. Now the partial derivative

with respect to $u_{N+j+m-n}$ vanishes unless $N+j+m-n=N-p$, i.e., $n=j+m+p\geqslant 2p+m$.

As a consequence, the lowest n such that $a^{(p)}$ may appear in $(23)_n$ is $n=2p$. However, since $c_{p,0}^{(2p)}=0$, this is not really the case. Thus we must consider the next one, $n=2p+1$. Now $(23)_{2p+1}$ contains both $a^{(p)}$ and $a_1^{(p)}$ separatedly in nonzero terms. If we make the induction hypothesis that $a^{(j)}\in\mathcal{F}_{M+2j}$, $0\leqslant j\leqslant p-1$, then it is very easy to see that the highest derivative in $(23)_{2p+1}$ is u_{M+2p+1}, so that $a^{(p)}\in\mathcal{F}_{M+2p}$, which completes the argument.

We may summarize the conclusions of this section in the following way:

Proposition 2: Given $J\geqslant 0$, there exists $N_0(J)$ such that every $\rho\in C_N(P)$, $N\geqslant N_0(J)$, satisfies

$$\rho\sim\tfrac{1}{2}\sum_{j=0}^{J}a^{(j)}u_{N-j}^2+R,\quad R\in\mathcal{F}_{N-J-1},$$

with $a^{(j)}\in\mathcal{F}_{M+2j}$. Moreover, the coefficients $a^{(j)}$ satisfy the differential equations

$$\sum_{r=0}^{[(n-1)/2]}\sum_{m=0}^{n-2r}[E_{r,m}^{(n)}a^{(r)}]_m=h^{(n)},\quad n=1,2,\cdots(\ll N),\qquad (23')_n$$

with

$$E_{r,m}^{(n)}\equiv c_{n-r-m,m}^{(n)}\sum_{k=n-2r-m-M}^{n-2r-m}\binom{N-r}{k}[P_{u_{M+n-2r-m+k}}]_k.$$

Note: Equation $(23')_n$ follows from $(23)_n$ by suitably changing the indices. In principle, $(23')_n$ allows to calculate $a^{(0)}$, $a^{(1)}$,..., in terms of P whenever the function P does not preclude the existence of solutions. (See, for instance, Proposition 3 in the next section.)

5. APPLICATION TO THIRD-ORDER EQUATIONS

In order to illustrate the power of the asymptotic algorithm, we are going to prove the following.

Proposition 3: If the polynomial P in (1) is such that:

(a) $P=u_3+f(u,u_1,u_2)$;

(b) $\exists a(u)\in\mathcal{F}_0'$ conserved under (1);

(c) $\exists\rho\in C_N(P)$ for some $N\gg 1$;

then

$$P=u_3+(\alpha_0+\alpha_1u+\alpha_2u^2)u_1,\quad \alpha_j\in\mathbb{R}.\qquad (25)$$

Proof: Hypothesis (c) and the Proposition 1 (iii)$\Rightarrow f_{u_2}=0$, i.e., $f=f(u,u_1)$. The conservation of $a(u)$ requires

$$0\sim a_t=a'(u)P\sim(a'''/2)u_1^3+a'f.\qquad (26)$$

Therefore, $\exists H(u)$ such that

$$f=-(a'''/2a')u_1^3+H(u)u_1.\qquad (27)$$

The polynomial character of P (and hence f) forces $a'''=0$, and thus

$$P=u_3+H(u)u_1.\qquad (28)$$

Let us finally apply the asymptotic algorithm $(23')_n$ to the density ρ in (c):

$$n=1\Rightarrow a^{(0)}=\text{constant. We choose } a^{(0)}=1.$$

$$n=3\Rightarrow a^{(1)}=-\frac{2N+1}{3}[H(u)+\alpha],\ \alpha=\text{constant}.\qquad (29)$$

$$n=5\Rightarrow 3a_1^{(2)}-\tfrac{1}{6}(2N+1)(N^2+N-2)H_3-\frac{2N+1}{3}H_1[(2N-1)(H+\alpha)-NH]$$
$$=h^{(5)}=\frac{2N+1}{3}\{[H'u_2-\tfrac{1}{2}H''u_1^2+\tfrac{1}{2}H^2]_1+H'''u_1^3\}.\qquad (30)$$

The integrability of (30) requires $H'''u_1^3\sim 0$, i.e., $H'''=0$, which completes the proof. ■

Note that (25) covers those $f(u,u_1,u_2)$ which are linear combinations of u_1 (linear), uu_1 (KdV) and u^2u_1 (modified KdV). It is also remarkable that other sets of weak hypothesis lead to the same conclusion. For instance, Proposition 3 holds if (a), (b), and (c) are replaced by

(a') $P_{uu_1}=0$;

(b') $\exists a(u), b(u)\in\mathcal{F}_0'$, conserved and inequivalent $(b(u)\not\sim\lambda a(u))$;

(c') $\exists\rho\in C_N(P)$, some $N\gg 1$, with $a^{(0)}\in\mathcal{F}_0$.

The proof follows similar lines, by using Proposition 1, (ii).

When $H'''=0$, the integration of (30) leads to

$$a^{(2)}=\frac{2N+1}{18}[N(N+1)H'u_2+(N^2+N-3)H''u_1^2+(2N-1)(H+\alpha)^2+\beta],\qquad (31)$$

with β constant. One might proceed this way step by step to determine the coefficients $a^{(r)}$. As $H'''=0$, conserved densities of arbitrarily high order are known to exist[3] and thus no further obstruction would be met. The succesive integration constants $\alpha,\beta,\cdots$, cannot be determined since their values can be changed by adding to our ρ other conserved densities of lower order.

If $\alpha=\beta=\cdots=0$, $H(u)=u$, and $N=9$, then

$$\rho=u_9^2-\tfrac{19}{3}uu_8^2+\tfrac{19}{18}[90u_2+17u^2]u_7^2+\cdots,\qquad (32)$$

which coincides, up to an irrelevant global factor $-2519424/46189$, with the corresponding density given in Ref. 2.

It should finally be noted that our Proposition 3 generalizes a previous result by Estabrook and Wahlquist,[4] who reached the same conclusion (25) under the stronger hypothesis $P=u_3+f(u)u_1$, by using their prolongation structure techniques.

ACKNOWLEDGMENTS

We thank G. García Alcaine, F. Guil, L. Martínez, and F. Ramirez for many useful discussions. The financial support of the Junta de Energía Nuclear is also acknowledged.

[1] L. Abellanas and A. Galindo, J. Math. Phys. **20**, 1239 (1979); **21**, 1267 (1980).

[2] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).

[3] R. M. Miura, "The Korteweg–de-Vries equation: A model equation for nonlinear dispersive waves," in *Nonlinear Waves*, edited by S. Leibovich and A. Seebass (Cornell U. P., Ithaca, New York, 1974).

[4] F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **17**, 1293 (1976). The authors are indebted to the referee for bringing this reference to their attention.

On the nature of the Gardner transformation

B. A. Kupershmidt
Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 April 1980; accepted for publication 17 October 1980)

It is shown that every higher Korteveg-de Vries equation can be included in a one-parameter family of integrable equations.

PACS numbers: 02.30.Jr, 02.30.Qy

1. INTRODUCTION

The initial analysis of the Korteveg–de Vries (KdV) equation by Gardner, Green, Kruskal, Miura, and Zabusky, which later developed into a theory of the so-called "integrable systems", appears now to be a combination of some ingenious tricks (which have only been applied in the case of the KdV equation) and the general concept of "inverse scattering". While the general developments of the theory of inverse scattering are fairly well known (see, e.g., Refs. 1–4), the applicability of the ingenious tricks has not been examined in general; it is the goal of this note to discuss the degree of generality of the particular trick that has been called the "Gardner transformation" (see the historical remarks in Ref. 5, p. 422), which led to the discovery of an infinite number of conservation laws for the KdV equation.

2. THE GARDNER TRANSFORMATION FOR THE KdV EQUATION

We recall briefly the actual derivation, which is taken from Ref. 5. If v satisfies the modified KdV (MKdV) equation

$$v_t = 6v^2v_x - v_{xxx}, \tag{1}$$

then

$$u = v^2 + v_x \tag{2}$$

satisfies the KdV equation

$$u_t = 6uu_x - u_{xxx}. \tag{3}$$

Equation (2) is called the "Miura transformation" (a general interpretation of it can be found in Ref. 6). Now (3) admits Galilean symmetry

$$t' = t, \qquad x' = x + 6ct, \qquad u' = u + c, \tag{4}$$

while (1) does not. Then a suitable combination of (2) and (4) shows that

$$u = w + \epsilon w_x + \epsilon^2 w^2, \tag{5}$$

the Gardner transformation, is a solution of (3) if w is a solution of

$$\begin{aligned} w_t &= 6ww_x - w_{xxx} + 6\epsilon^2 w^2 w_x \\ &= \partial(3w^2 - w_{xx} + 2\epsilon^2 w^3), \qquad \partial \equiv d/dx, \end{aligned} \tag{6}$$

which can be considered as a deformation of Eq. (3). Moreover, Eq. (6) also possesses an infinite number of conservation laws [because it becomes equivalent to (1) after a change of variables]. So we have an integrable deformation. Since such deformations are only very rarely integrable, the situation is quite intriguing. And even if all that were not enough, formula (5) tells us that Eqs. (3) and (6) are in fact equivalent as evolution fields [in other words, using (5) we can express w as a formal power series in ϵ with coefficients that are polynominals in $u, u_x, \cdots$ $w = u - \epsilon u_x + \epsilon^2(u_{xx} - u^2) + \cdots$. Since $w_t = \partial(3w^2 - w_{xx} + 2\epsilon^2 w^3)$ we obtain an infinite number of conservation laws for (3) by inverting (5); half of them are nontrivial and this completes the "classical" story of the Gardner transformation[7]].

3. INTERPRETATION

The Galilean invariance of the KdV equation which was used in an essential way in the construction of the deformation (5), is no longer available for the higher KdV equations which correspond to the Lax representation

$$L_t = [P, L], \tag{7}$$

with $L = -\partial^2 + u$. Therefore, if any deformation exists for the higher KdV equations it must be based on something other than Galilean invariance. We shall examine this in what follows.

First, we recall that one of the most important properties of the Miura map (2) is the fact that this map is a "canonical transformation" from the natural Hamiltonian structure $\partial(\delta/\delta v)$,

$$v_t = \partial(\delta H/\delta v), \tag{8}$$

of Eq. (1), namely,

$$v_t = 6v^2v_x - v_{xxx} = \partial(\delta/\delta v)(\tfrac{1}{2}v^4 + \tfrac{1}{2}v_x^2), \tag{1'}$$

into the second Hamiltonian structure
$B_2 = (-\partial^3 + 2\partial u + 2u\partial)(\delta/\delta u)$,

$$u_t = (-\partial^3 + 2u\partial + 2\partial u)(\delta h/\delta u) \tag{9}$$

of the KdV equation, namely

$$u_t = 6uu_x - u_{xxx} = (-\partial^3 + 2u\partial + 2\partial u)(\delta/\delta u)\tfrac{1}{2}u^2. \tag{3'}$$

The word canonical means that the corresponding Poisson brackets are compatible with the Miura map (2). Technically, this statement is equivalent to the equality

$$(2v + \partial)\partial(2v - \partial) = -\partial^3 + 2u\partial + 2\partial u, \tag{10}$$

where $2v + \partial = Du/Dv$ is the corresponding Fréchet derivative and $2v - \partial = (2v + \partial)^+$ is its adjoint operator (for details see Ref. 2 or 6).

Keeping this in mind it is natural to assume that the map (5) is also a canonical map from the natural Hamilton-

 022-2488/81/030449-03$1.00

ian structure $\partial(\delta/\delta w)$ of Eq. (6).

$$w_t = 6ww_x - w_{xxx} + 6\epsilon^2 w^2 w_x = \partial(\delta/\delta w)(w^3 + w_x^2/2 + \epsilon^2 w^4/2), \tag{11}$$

into some other Hamiltonian structure, which must be nothing else but $[(Du/Dw)\partial(Du/Dw)^+]\delta/\delta u$. We have $(Du/Dw)\,\partial(Du/Dw)^+ = (1 + \epsilon\partial + 2\epsilon^2 w)\partial(1 - \epsilon\partial + 2\epsilon^2 w) = \partial - \epsilon^2\partial^3 + 2\epsilon^2[(w + \epsilon w_x + \epsilon^2 w^2)\partial + \partial(w + \epsilon w_x + \epsilon^2 w^2)] = \partial + \epsilon^2(-\partial^3 + 2u\partial + 2\partial u)$. Therefore the map (5) is indeed canonical into the Hamiltonian structure B:

$$B = [\partial + \epsilon^2(-\partial^3 + 2u\partial + 2\partial u)](\delta/\delta u), \tag{12}$$

which is just the first Hamiltonian structure $B_1 = \partial(\delta/\delta u)$ of the KdV equation,

$$u_t = 6uu_x - u_{xxx} = \partial(\delta/\delta u)(u^3 + \tfrac{1}{2}u_x^2)$$

plus ϵ^2 times the second Hamiltonian structure B_2 of the KdV equation: $B = B_1 + \epsilon^2 B_2$.

Thus we have arrived at the combination (12) of two Hamiltonian structures for the higher KdV equations. This means that the Hamiltonians for these equations, if they exist, are formed in some way from the regular Hamiltonians H_n of the KdV equations. Recall that the sequence H_n of Hamiltonians is such that the equation

$$u_t = B_2(H_n) = B_1(H_{n+1}) \tag{13}$$

is the higher KdV equation number n (see, e.g., Ref. 2), and all the Hamiltonians H_n commute in both the Hamiltonian structures B_1 and B_2.

Because the natural initial term $H_0 = u$ is such that $B_1(H_0) = 0$, we see that Eq. (13) can be written in terms of our mixed Hamiltonian structure (12) as

$$u_t (= B_2(H_n)) = B(\tilde{H}_n), \tag{14}$$

where

$$\tilde{H}_n = \epsilon^{-2} \sum_{k=0}^{n} (-\epsilon^{-2})^k H_{n-k}. \tag{15}$$

Notice that $\tilde{H}_n = \tilde{H}_n(\epsilon)$ is a *singular* function of ϵ. For example, the KdV equation could be written as

$$u_t = 6uu_x - u_{xxx} = [\partial + \epsilon^2(-\partial^3 + 2u\partial + 2\partial u)]\frac{\delta}{\delta u}\left(\frac{1}{\epsilon^2}\frac{u^2}{2} - \frac{1}{\epsilon^4}\frac{u}{2}\right),$$

and if we consider the linear combination of the KdV fields which corresponds to $H = \Sigma\, c_i H_i$,

$$u_t = B_2(H), \tag{16}$$

then the same Eq. (16) in our mixed Hamiltonian structure B has the Hamiltonian

$$\tilde{H} = \sum_i c_i \epsilon^{-2} \sum_{k=0}^{i} (-\epsilon^{-2})^k H_{i-k}. \tag{17}$$

4. DEFORMATION

The next step is to make sure that every deformed equation

$$w_t = \partial(\delta/\delta w)\tilde{H}^*, \tag{18}$$

where $\tilde{H}^*(w) = \tilde{H}(u)$ at $u = w + \epsilon^2 w^2 + \epsilon w_x$ and H is taken from (17), is indeed a deformation of the "unperturbed" Eq. (16) in exactly the same manner as (6) is a deformation of (3). Of course, $B|_{\epsilon=0} = B_1$ but there is a potential source of difficulties in the singular dependence of $\tilde{H}$ upon ϵ in (17).

Theorem. The r.h.s. of the modified Eq. (18) is a polynomial in ϵ for $\tilde{H}$ taken from (17).

Proof: It suffices to check the claim for $\tilde{H} = \tilde{H}_n$ from (15). In this case the r.h.s. of (18) is clearly a finite polynomial in ϵ and ϵ^{-1}; we wish to show that this polynomial contains no terms which involve negative powers of ϵ. This follows from the *regular* invertibility of the deformation (5): We can write $w = u + \Sigma_{k=1}^{\infty} \epsilon^k P_k$, P_k being finite polynomial in $u, u_x, \ldots$, and therefore $w_t = \partial_t(u + \Sigma\, \epsilon^k P_k) = F + \Sigma\, \epsilon^k\, \Sigma(\partial P_k/\partial u^{(s)})\partial^s F$, where F is the r.h.s. of (16) and $u^{(s)} = d^s u/dx^s$. Substituting $u = w + \epsilon^2 w^2 + \epsilon w_x$ in the last expression we find w_t as a formal series in non-negative powers of ϵ only. Q.E.D.

Example: The next KdV equation after (3) is

$$u_t = (-\partial^3 + 2u\partial + 2\partial u)(\delta/\delta u)(u^3 + u_x^2/2) = \partial[u^{(4)} - 10uu^{(2)} - 5u^{(1)2} + 10u^3] = \partial(\delta/\delta u)(5/2u^4 + 5uu^{(1)2} + \tfrac{1}{2}u^{(2)2}). \tag{19}$$

Corresponding $\tilde{H} = \epsilon^{-2}(u^3 + u_x^2/2) - \epsilon^{-4}(u^2/2) + \epsilon^{-6}(u/2)$; this gives a deformation of Eq. (19) in the form

$$w_t = \partial(\delta/\delta w)\{\tfrac{5}{2}w^4 + 5ww^{(1)2} + \tfrac{1}{2}w^{(2)2} + \epsilon^2[5w^2 w^{(1)2} + 3w^5] + \epsilon^4 w^6\}. \tag{20}$$

Note that the (usual) modified equation associated with (19) via the Miura map (2) is

$$v_t = \partial(\delta/\delta v)\{v^6 + 5v^2 v^{(1)2} + \tfrac{1}{2}v^{(2)2}\}. \tag{21}$$

5. RELATIONS BETWEEN THE GARDNER AND MIURA TRANSFORMATIONS

In Secs. 3 and 4 we showed that the deformation (5) is indeed valid for the whole KdV hierarchy. We now wish to understand its relation with the Miura map (2).

The map (2) is canonical between the Hamiltonian structures $\partial(\delta/\delta v)$ and B_2 and therefore it is canonical for the Hamiltonian structures $\epsilon^2\partial(\delta/\delta v)$ and $\epsilon^2 B_2$. If we now make the translation $\tilde{u} = u + c\epsilon^{-2}$, then the map

$$\tilde{u} = c\epsilon^{-2} + v_x + v^2 \tag{2'}$$

is canonical between $\epsilon^2\partial(\delta/\delta v)$ and $\epsilon^2[-\partial^3 + 2(\tilde{u} - c\epsilon^{-2})\partial + 2\partial(u - c\epsilon^{-2})]\delta/\delta\tilde{u} = [\epsilon^2(-\partial^3 + 2\tilde{u}\partial + 2\partial\tilde{u}) - 4c\partial](\delta/\delta\tilde{u})$, which is exactly $B(\tilde{u})$ when $c = -\frac{1}{4}$. To eliminate ϵ^2 in $\epsilon^2\partial(\delta/\delta v)$ we set $v = \epsilon\tilde{v}$; then (2') becomes

$$\tilde{u} = -\epsilon^{-2}/4 + \epsilon^2\tilde{v}^2 + \epsilon\tilde{v}_x, \tag{2''}$$

which is a canonical map between $\partial(\delta/\delta\tilde{v})$ and $B(\tilde{u})$. To convert (2'') into a regular map we observe that the Hamiltonian structure $\partial(\delta/\delta\tilde{v})$ has *constant* coefficients and hence is invariant under translations of $\tilde{v}$. So if we let $\tilde{v} = w + b$, (2'') will become

$$\tilde{u} = -\epsilon^{-2}/4 + \epsilon^2 w^2 + 2\epsilon^2 wb + \epsilon^2 b^2 + \epsilon w_x. \tag{2'''}$$

Now the regularity condition for (2‴),

$$\epsilon^2 b^2 = 1/4\epsilon^2, \tag{22}$$

yields

$$b = 1/2\epsilon^2, \tag{23}$$

and (2‴) becomes (5).

6. DISCUSSION

For each higher KdV Eq. (13) we constructed its deformation (18) which has the following properties:

(i) There exists the *reduction map* (5) of the deformed equation into the undeformed one. Therefore the *deformed system is also integrable* (meaning: has an infinity of integrals), because all conservation laws (c.l.'s) of the undeformed equation become c.l.'s of the deformed equation after pull back;

(ii) The deformed equation (18) is Hamiltonian; it has now *only one* Hamiltonian structure, $\partial(\delta/\delta w)$. In this structure, all integrals of the deformed equations commute, since they are preimages of the c.l.'s which were in involution already, and the reduction map is *canonical*. Note that there is no such thing as "Lenard relations" (13) for the deformed equations.

It will be important to understand whether there exists any general "integrable deformations" pattern in the theory of integrable systems. The answer is undoubtedly yes and will be dealt with elsewhere. Here I shall make brief remarks.[8]

A) If one begins with the arbitrary scalar Lax equation (7) with

$$L = \partial^{n+2} + \sum_{i=0}^{n} u_i \partial^i \tag{24}$$

then one can construct the deformation theory, and generalizations of both properties (i) and (ii) from the above discussion remain true.

B) When an integrable equation is not bi-Hamiltonian, Hamiltonian formalism is of little help either to find a deformation or to interpret it.

Examples:

1) If

$$p_t = 6p_x C(1 + \epsilon^2 C) - p_{xxx} + 2\epsilon^2 \nu^2 p_x^3, \tag{25}$$

where

$$C = (\mathrm{sh}2\epsilon\nu p)/(2\epsilon\nu) + (\mathrm{ch}2\epsilon\nu p - 1)/(2\epsilon^2), \tag{26}$$

then

$$W = C + \nu p_x \tag{27}$$

satisfies (6). Thus (25) represents the *second* deformation of the KdV equation (3).

(2) If

$$q_t = \partial[2q^3 - q_{xx} + 6\epsilon^2 q q_x^2/(1 + 4\epsilon^2 q^2)], \tag{28}$$

then

$$w = [(1 + 4\epsilon^2 q^2)^{1/2} - 1]/2\epsilon^2 + q_x(1 + 4\epsilon^2 q^2)^{-1/2} \tag{29}$$

satisfies (6),

$$v = q + \epsilon q_x(1 + 4\epsilon^2 q^2)^{-1/2} \tag{30}$$

satisfies (1), and we have the commutative diagram (2)·(30) = (5)·(27). Thus (28) is the deformation of the MKdV equation (1), (30) is the reduction map, and (29) is the *deformation of the Miura map*. This suggests that not only integrable systems but also their relationships are objects of deformations.

(3) Deformations phenomenon is not the privilege of only the Lax equations. Consider, e.g., the Benney equations for long waves on a two-dimensional surface[9]

$$a_{n,t} = a_{n+1,x} + n a_{n-1} a_{0,x}, \qquad n = 0,1,2,\cdots \tag{31}$$

for the sequence of functions $a_n(x,t)$. This system has an infinity of integrals $h_n \in a_n + \mathbb{Z}[a_0,\ldots,a_{n-2}]$ (see Ref. 2).

Proposition 32: Let

$$A_{n,t} = A_{n+1,x} + nA_{n-1}A_{0,x} + \epsilon[A_0 A_{n,x} + (n+1)A_n A_{0,x} + nA_{n-1}(A_{1,x} - \epsilon A_0 A_{0,x}/2)]/2, \qquad n = 0,1,2,\cdots. \tag{33}$$

Denote by $H_n \in A_n + \mathbb{Q}[A_0,\ldots,A_{n-1}]$ the integral $\#n$ for (33). Then the map

$$a_n = A_n + O(\epsilon), \tag{34}$$

such that

$$h_n = H_n + \epsilon H_{n+1}, \tag{35}$$

maps solutions of (33) into solutions of (31).

C) Evidently conservation laws survive deformations, i.e., remain nontrivial under deformations. Therefore it is important to know which integrals of the undeformed equations were nontrivial in the first place.

Let us consider, as an example, the well-known case of the KdV hierarchy (13). Then the r.h.s. of (18) shows that for the deformed equation w is the c.l. Therefore inverting the Gardner transformation (5)

$$w = \sum_{n=0}^{\infty} h_n \epsilon^n \tag{36}$$

one gets an infinity of c.l.'s $h_n \in \mathscr{A}$ where $\mathscr{A}$ is the ring of polynomials in $u,u_x,\cdots,$. If $f,g,\in\mathscr{A}$, let us write $f \approx g$ is $f(u,0,0,\cdots) = g(u,0,\cdots)$, and $f \sim 0$ if $f = \partial g$.

Proposition 37: $h_{2n+1} \sim 0, h_{2n} \not\sim 0$.

Proof: 1) Write $w = w^+ + w^-$ where $w^+ = \Sigma\, h_{2n}\epsilon^{2n}$, $w^- = \Sigma\, h_{2n+1}\epsilon^{2n+1}$ and substitute this into (5). Then the part which is odd in ϵ yields $w^- - \epsilon w_x^+ - 2\epsilon^2 w^+ w^- = 0$, or $w^- = -(2\epsilon)^{-1}\partial \ln(1 - 2\epsilon^2 w^+)$. 2) From (5) one gets $u \approx w + \epsilon^2 w^2$, so $w \approx (2\epsilon^2)^{-1}[1 - (1 - 4\epsilon^2 u)^{1/2}] = \Sigma_{\infty}^{n=0} c_n \epsilon^{2n} \mu^{n+1}$, all c_n's are different from zero. Thus $h_{2n} \approx c_n u^{n+1}$. Note now that if $f \sim 0$ then $f \approx 0$. Q.E.D.

[1] I. M. Krichever, Russ. Math. Surv. **32**, 185–213 (1977).
[2] Yu. I. Manin, J. Sov. Math. **11**, 1–122 (1979).
[3] G. Wilson, Math. Proc. Cambridge Philos. Soc. **86**, 131–143 (1979).
[4] I. M. Gel'fand and L. A. Dikii, Russ. Math. Surv. **30**, 67 (1975).
[5] R. M. Miura, SIAM Review **18**, 412 (1976).
[6] B. A. Kupershmidt and G. Wilson, "Modifying Lax Equations and the second Hamiltonian structure," Invent. Math. (to appear).
[7] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[8] B. A. Kupershmidt, "Deformations of Integrable Systems" (to appear).
[9] D. J. Benney, Stud. Appl. Math. **52**, 45 (1973).
[10] B. A. Kupershmidt, "Integrable Two-Dimensional Systems" (to appear).

Analytical expansions for Fermi–Dirac functions

Alicia Bañuelos [a)]
Laboratorio de Física del Plasma, Dirección General de Investigación y Desarrollo, Ministerio de Defensa Departamento de Física-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

Ricardo Angel Depine [a)]
Laboratorio de Optica, Departamento de Física-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

Roberto Claudio Mancini [a)]
Laboratorio de Física del Plasma, Dirrección General de Investigación y Desarrollo, Ministerio de Defensa Departamento de Física-Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1428 Buenos Aires, Argentina

(Received 15 August 1980; accepted for publication 18 September 1980)

We obtain a fast convergent series expansion for the Fermi–Dirac function $F_\sigma(\alpha)$ for $-10 \leqslant \alpha \leqslant -1$. We give values of $F_\sigma(\alpha)$ for $\sigma = n + \frac{1}{2}$ $(n = 0,1,\cdots,6)$ with α in the same range.

PACS numbers: 02.30.Mv

I. INTRODUCTION

The Fermi–Dirac functions $F_\sigma(\alpha)$, where σ is a positive real parameter, is defined for all real numbers α by

$$F_\sigma(\alpha) = \frac{1}{\Gamma(\sigma)} \int_0^\infty \frac{x^{\sigma-1}}{e^{x+\alpha}+1}\, dx.$$

When σ is an integer, this integral may be easily evaluated by a power series; a complete discussion of this case is due to Rhodes.[1] For arbitrary σ, there are several expansions depending on the range of values of α.[2–4] The calculation of $F_\sigma(\alpha)$ for $\alpha < 0$ is needed in many questions of quantum statistical mechanics; for example, to solve the equations of state corresponding to extreme conditions (high pressure and nonzero temperature). Analytical expansions are available in all ranges, except when $-10 \leqslant \alpha \leqslant -1$. Previous evaluations of $F_\sigma(\alpha)$ for this range were made by numerical integration[4,5] or by polynomial approximation.[6] In this paper we obtain a fast convergent series expansion of $F_\sigma(\alpha)$ for $-10 \leqslant \alpha \leqslant -1$.

II. SERIES EXPANSION FOR $-10 \leqslant \alpha \leqslant -1$

For simplicity, let us define

$$I = I_\sigma(\alpha) = \Gamma(\sigma) F_\sigma(\alpha) = \int_0^\infty \frac{x^{\sigma-1}}{e^{x+\alpha}+1}\, dx.$$

Substituting $y = x + \alpha$ in this integral gives

$$I = \int_{-|\alpha|}^{\infty} \frac{(y+|\alpha|)^{\sigma-1}}{e^y+1}\, dy.$$

Now I will be calculated as

$$I = I_1 + I_2 + I_3$$

by dividing the integration interval by the points $-p$ and p, where $0 < p < |\alpha|$. Another restriction on the values of p and convenient numerical suggestions will appear later.

A. Evaluation of I_1

First we expand the integrand denominator of

$$I_1 = \int_{-|\alpha|}^{-p} \frac{(y+|\alpha|)^{\sigma-1}}{e^y+1}\, dy \tag{1}$$

in a series of powers of e^y

$$\frac{1}{e^y+1} = \sum_{n\geqslant 0} (-1)^n e^{ny}.$$

This series converges uniformly for $|e^y| < 1$, that is, for $y < 0$. Now, by expanding e^{ny} at a convenient point y_0, we obtain

$$e^{ny} = e^{ny_0} \sum_{k\geqslant 0} \frac{n^k (y-y_0)^k}{k!}.$$

By replacing successively in (1), taking into account the uniform convergence of the series to exchange the order of integrals and summations, it follows that

$$I_1 = \sum_{n\geqslant 0} (-1)^n e^{ny_0} \sum_{k\geqslant 0} \frac{n^k}{k!} \times \int_{-|\alpha|}^{-p} (y+|\alpha|)^{\sigma-1} (y-y_0)^k\, dy.$$

The integrals involved in this expression may be evaluated using the formula

$$\int (a+bx)^{\sigma-1} x^k\, dx = \frac{(a+bx)^\sigma}{b^{k+1}} \times \sum_{0\leqslant j\leqslant k} \frac{(-1)^j \binom{k}{j} (a+bx)^{k-j} a^j}{k-j+\sigma}.$$

Thus,

$$I_1 = (|\alpha|-p)^\sigma \sum_{n\geqslant 0} (-1)^n e^{ny_0} \sum_{k\geqslant 0} \frac{n^k}{k!} A_k,$$

where

$$A_k = \sum_{0\leqslant j\leqslant k} \frac{(-1)^j \binom{k}{j} (|\alpha|-p)^{k-j} (y_0+|\alpha|)^j}{k-j+\sigma}.$$

[a)] Fellow of CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina).

 022-2488/81/030452-04$1.00

B. Evaluation of I_2

We notice that

$$I_2 = |\alpha|^{\sigma-1}\int_{-p}^{p}\frac{(y/|\alpha|+1)^{\sigma-1}}{e^y+1}dy.$$

Since $0<p<|\alpha|$, the series

$$\left(\frac{y}{|\alpha|}+1\right)^{\sigma-1}=\sum_{n\geqslant 0}\binom{\sigma-1}{n}\left(\frac{y}{|\alpha|}\right)^n$$

converges uniformly. The same statement holds[7] for the series

$$\frac{1}{e^y+1}=\frac{1}{2}+\frac{1}{2}\sum_{k\geqslant 1}\frac{(1-4^k)B_k y^{2k-1}}{k(2k-1)!}$$

when $p<\pi$, B_k being the nonzero Bernoulli numbers. Therefore, arguments used in Sec. (A) apply here, yielding

$$I_2=|\alpha|^{\sigma-1}\sum_{n\geqslant 1}C_n\left(\frac{1}{\sigma-2n+1}+\frac{1}{|\alpha|}\sum_{k\geqslant 1}\frac{D_k}{2k+2n-1}\right),$$

where

$$C_1=(\sigma-1)p,$$

$$C_{n+1}=C_n\frac{(\sigma-2n)(\sigma-2n-1)}{(2n+1)2n}\left(\frac{p}{|\alpha|}\right)^2,$$

$$D_1=-3B_1p^2,$$

and

$$D_{k+1}=D_k\frac{(1-4^{k+1})B_{k+1}p^2}{(1-4^k)B_k 2(k+1)(2k+1)}.$$

C. Evaluation of I_3

Recall that

$$I_3=\int_p^\infty\frac{(y+|\alpha|)^{\sigma-1}}{e^y+1}dy.$$

Since the expansion of the integrand denominator in a series of powers of e^{-y}

$$\frac{1}{e^y+1}=-\sum_{n\geqslant 1}(-1)^n e^{-ny},$$

is uniformly convergent for $y>0$, by exchanging the integral and the summation, with the substitution $z=n(y+|\alpha|)$ we obtain

$$I_3=\sum_{n\geqslant 1}\frac{(-1)^{n+1}e^{n|\alpha|}}{n^\sigma}\int_{n(|\alpha|+p)}^\infty z^{\sigma-1}e^{-z}dz.$$

Thus, I_3 can be expressed in terms of incomplete Gamma functions as

$$I_3=\sum_{n\geqslant 1}\frac{(-1)^{n+1}e^{n|\alpha|}}{n^\sigma}\Gamma(\sigma,n(p+|\alpha|)).$$

From our numerical investigations we conclude that, in order to achieve a fast convergence, the values of p and y_0 may be chosen as follows:

$$p=|\alpha|/2,\ \text{if } 1\leqslant|\alpha|<5;$$

$$p=2.5,\ \text{if } 5\leqslant|\alpha|\leqslant 10$$

$$y_0=-(|\alpha|+p)/2.$$

As an application, values of $F_\sigma(\alpha)$ for $-10\leqslant\alpha\leqslant -1$ and $\sigma=n+\frac{1}{2}$ $(n=0,1,...,6)$ were computed with a maximum relative error of 10^{-5}. In particular, we have checked the accuracy of all previously tabulated values.[4,6] In the course of the computation we have made use of Abramowitz's tables[7] for Bernoulli numbers. The corresponding computing program is to be published elsewhere.[8]

TABLE I. Values of $F_\sigma(\alpha)$ for $-10\leqslant\alpha\leqslant -1$ and $\sigma=n+\frac{1}{2}$ with $n=0,1,...,6$.

ALPHA	F1/2	F3/2	F5/2	F7/2	F9/2	F11/2	F13/2
-1.0	1.027050 00	1.575650 00	2.002250 00	2.294840 00	2.478670 00	2.587180 00	2.648360 00
-1.1	1.071660 00	1.668580 00	2.165020 00	2.503120 00	2.718420 00	2.846840 00	2.919800 00
-1.2	1.116230 00	1.760990 00	2.338520 00	2.728200 00	2.979860 00	3.131580 00	3.218490 00
-1.3	1.160820 00	1.853840 00	2.523170 00	2.971190 00	3.264670 00	3.443600 00	3.547010 00
-1.4	1.205190 00	2.022130 00	2.719430 00	3.233230 00	3.574720 00	3.785360 00	3.908200 00
-1.5	1.249360 00	2.144870 00	2.927750 00	3.515470 00	3.911980 00	4.159460 00	4.305160 00
-1.6	1.293150 00	2.271990 00	3.148550 00	3.819180 00	4.278550 00	4.568720 00	4.741270 00
-1.7	1.336600 00	2.403480 00	3.382290 00	4.145630 00	4.676590 00	5.016200 00	5.220190 00
-1.8	1.379640 00	2.539290 00	3.629400 00	4.496100 00	5.108450 00	5.505160 00	5.745900 00
-1.9	1.422220 00	2.679390 00	3.890300 00	4.871950 00	5.576640 00	6.039130 00	6.322730 00
-2.0	1.464290 00	2.823730 00	4.165410 00	5.274620 00	6.083740 00	6.621810 00	6.955340 00
-2.1	1.505840 00	2.972230 00	4.455170 00	5.705520 00	6.632530 00	7.257260 00	7.648860 00
-2.2	1.546830 00	3.124870 00	4.760000 00	6.166170 00	7.225860 00	7.949790 00	8.408660 00
-2.3	1.587200 00	3.281580 00	5.080290 00	6.658050 00	7.866800 00	8.704000 00	9.240840 00
-2.4	1.627060 00	3.442300 00	5.416450 00	7.182750 00	8.558550 00	9.524830 00	1.015170 01
-2.5	1.666310 00	3.606980 00	5.768880 00	7.741880 00	9.304480 00	1.041750 01	1.114820 01
-2.6	1.704940 00	3.775550 00	6.137970 00	8.337070 00	1.010810 01	1.138770 01	1.223760 01
-2.7	1.742980 00	3.947950 00	6.524110 00	8.970030 00	1.097320 01	1.244120 01	1.342850 01

(*continued*)

TABLE I. (*Continued*).

-2.8	1.780416 00	4.124120 00	6.927693 00	9.642483 00	1.19034D 01	1.358450 01	1.47294D 01
-2.9	1.817246 00	4.304016 00	7.349C6D 00	1.03562D 01	1.29030D 01	1.48242D 01	1.61486D 01
-3.0	1.853496 00	4.487550 00	7.788616 00	1.11129D 01	1.39761D 01	1.61675D 01	1.76973D 01
-3.1	1.889156 00	4.674696 00	8.246696 00	1.19145D 01	1.51271D 01	1.76220D 01	1.93858D 01
-3.2	1.924246 00	4.865366 00	8.723676 00	1.27629D 01	1.63606D 01	1.91957D 01	2.12257D 01
-3.3	1.958776 00	5.059526 00	9.219886 00	1.36599D 01	1.76813D 01	2.089700 01	2.32292D 01
-3.4	1.992756 00	5.257106 00	9.735686 00	1.46075D 01	1.90942D 01	2.27350D 01	2.54097D 01
-3.5	2.026196 00	5.458056 00	1.027146 01	1.56077D 01	2.06046D 01	2.47191D 01	2.77811D 01
-3.6	2.059126 00	5.662326 00	1.082746 01	1.66625D 01	2.22176D 01	2.68593D 01	3.03587D 01
-3.7	2.091536 00	5.869866 00	1.140466 01	1.77739D 01	2.39389D 01	2.91662D 01	3.31585D 01
-3.8	2.123456 00	6.080616 00	1.200156 01	1.89439D 01	2.57743D 01	3.165090 01	3.61979D 01
-3.9	2.154896 00	6.294536 00	1.262026 01	2.01749D 01	2.77297D 01	3.43251D 01	3.94956D 01
-4.0	2.185876 00	6.511576 00	1.326056 01	2.14667D 01	2.98114D 01	3.72011D 01	4.30696D 01
-4.1	2.216396 00	6.731696 00	1.392266 01	2.28277D 01	3.20257D 01	4.029180 01	4.69424D 01
-4.2	2.246486 00	6.954846 00	1.460696 01	2.42540D 01	3.43792D 01	4.361090 01	5.11356D 01
-4.3	2.276146 00	7.180976 00	1.531376 01	2.57498D 01	3.68788D 01	4.717250 01	5.56727D 01
-4.4	2.305396 00	7.410056 00	1.604326 01	2.73175D 01	3.95315D 01	5.099170 01	6.05787D 01
-4.5	2.334256 00	7.642046 00	1.679586 01	2.89592D 01	4.23447D 01	5.508420 01	6.58801D 01
-4.6	2.362716 00	7.877896 00	1.757176 01	3.06774D 01	4.53259D 01	5.946630 01	7.16052D 01
-4.7	2.390816 00	8.117576 00	1.837136 01	3.24744D 01	4.84828D 01	6.415520 01	7.77837D 01
-4.8	2.418556 00	8.360046 00	1.919476 01	3.43525D 01	5.18235D 01	6.916900 01	8.44471D 01
-4.9	2.445946 00	8.606276 00	2.004246 01	3.63141D 01	5.535616 01	7.452630 01	9.16289D 01
-5.0	2.472996 00	8.856226 00	2.091456 01	3.836186 01	5.908926 01	8.024690 01	9.936446 01
-5.1	2.499716 00	9.109856 00	2.181136 01	4.04978D 01	6.303146 01	8.635120 01	1.07691D 02
-5.2	2.526126 00	9.367156 00	2.273316 01	4.27248D 01	6.719186 01	9.286050 01	1.16648D 02
-5.3	2.552226 00	9.628076 00	2.368026 01	4.504536 01	7.157956 01	9.979710 01	1.26277D 02
-5.4	2.578036 00	9.892586 00	2.465286 01	4.746176 01	7.620416 01	1.071846 02	1.36623D 02
-5.5	2.603556 00	1.016376 01	2.565126 01	4.997676 01	8.107516 01	1.150466 02	1.47730D 02
-5.6	2.628796 00	1.043536 01	2.667566 01	5.259296 01	8.620286 01	1.234086 02	1.59648D 02
-5.7	2.653766 00	1.071946 01	2.772646 01	5.531276 01	9.159726 01	1.322966 02	1.72429D 02
-5.8	2.678486 00	1.090606 01	2.880366 01	5.813906 01	9.726886 01	1.417376 02	1.86126D 02
-5.9	2.702946 00	1.117516 01	2.990766 01	6.107446 01	1.032296 02	1.517596 02	2.00796D 02
-6.0	2.727156 00	1.144666 01	3.103876 01	6.412146 01	1.094876 02	1.623926 02	2.16496D 02
-6.1	2.751126 00	1.172056 01	3.219706 01	6.728306 01	1.160576 02	1.736676 02	2.33296D 02
-6.2	2.774876 00	1.199686 01	3.338296 01	7.056186 01	1.229486 02	1.856146 02	2.51254D 02
-6.3	2.798396 00	1.227556 01	3.459656 01	7.396056 01	1.301736 02	1.982676 02	2.70442D 02
-6.4	2.821696 00	1.255656 01	3.583816 01	7.748206 01	1.377446 02	2.116606 02	2.909326 02
-6.5	2.844776 00	1.284006 01	3.710796 01	8.112916 01	1.456746 02	2.258286 02	3.128006 02
-6.6	2.867666 00	1.312546 01	3.840616 01	8.490456 01	1.539746 02	2.408086 02	3.361256 02
-6.7	2.890346 00	1.341336 01	3.973306 01	8.881126 01	1.626596 02	2.566366 02	3.609906 02
-6.8	2.912826 00	1.370356 01	4.108886 01	9.285216 01	1.717416 02	2.733536 02	3.874826 02
-6.9	2.935126 00	1.399596 01	4.247386 01	9.703006 01	1.812346 02	2.909986 02	4.156916 02
-7.0	2.957236 00	1.429056 01	4.388816 01	1.013486 02	1.911526 02	3.096136 02	4.457146 02
-7.1	2.979166 00	1.458736 01	4.533206 01	1.058096 02	2.015086 02	3.292436 02	4.776486 02
-7.2	3.000926 00	1.488636 01	4.680566 01	1.104156 02	2.123186 02	3.499306 02	5.115976 02
-7.3	3.022506 00	1.518756 01	4.830936 01	1.151716 02	2.235966 02	3.717226 02	5.476716 02
-7.4	3.043926 00	1.549086 01	4.984326 01	1.200786 02	2.353576 02	3.946666 02	5.859856 02
-7.5	3.065176 00	1.579636 01	5.140766 01	1.251406 02	2.476176 02	4.188106 02	6.266446 02
-7.6	3.086276 00	1.610396 01	5.300256 01	1.303616 02	2.603916 02	4.442066 02	6.697846 02
-7.7	3.107206 00	1.641356 01	5.462846 01	1.357426 02	2.736956 02	4.709066 02	7.155286 02
-7.8	3.127996 00	1.672536 01	5.628536 01	1.412876 02	2.875456 02	4.989636 02	7.640106 02
-7.9	3.148636 00	1.703916 01	5.797356 01	1.470006 02	3.019586 02	5.284346 02	8.153686 02
-8.0	3.169126 00	1.735506 01	5.969326 01	1.528836 02	3.169506 02	5.593746 02	8.697466 02
-8.1	3.189486 00	1.767306 01	6.144466 01	1.589406 02	3.325406 02	5.918446 02	9.272946 02
-8.2	3.209696 00	1.799296 01	6.322796 01	1.651736 02	3.487446 02	6.259036 02	9.881686 02
-8.3	3.229776 00	1.831496 01	6.504326 01	1.715866 02	3.655816 02	6.616136 02	1.052536 03
-8.4	3.249716 00	1.863896 01	6.689096 01	1.781836 02	3.830676 02	6.990406 02	1.120556 03
-8.5	3.269536 00	1.896486 01	6.877116 01	1.849666 02	4.012236 02	7.382496 02	1.192406 03
-8.6	3.289226 00	1.929286 01	7.068396 01	1.919386 02	4.200676 02	7.793086 02	1.268266 03
-8.7	3.308786 00	1.962276 01	7.262976 01	1.991046 02	4.396176 02	8.222866 02	1.348326 03
-8.8	3.328226 00	1.995456 01	7.460856 01	2.064656 02	4.598946 02	8.672566 02	1.432786 03
-8.9	3.347546 00	2.028836 01	7.662076 01	2.140266 02	4.809176 02	9.142906 02	1.521846 03
-9.0	3.366756 00	2.062406 01	7.866636 01	2.217906 02	5.027066 02	9.634656 02	1.615716 03

(*continued*)

Cont. from Table I.

-9.1	3.385830 00	2.096170 01	8.074550 01	2.297610 02	5.252820 02	1.014860 03	1.714610 03
-9.2	3.404810 00	2.130120 01	8.285870 01	2.379410 02	5.486650 02	1.068550 03	1.818760 03
-9.3	3.423670 00	2.164260 01	8.500580 01	2.463340 02	5.728770 02	1.124620 03	1.928400 03
-9.4	3.442430 00	2.198590 01	8.718730 01	2.549430 02	5.979390 02	1.183150 03	2.043770 03
-9.5	3.461080 00	2.233110 01	8.940310 01	2.637720 02	6.238730 02	1.244230 03	2.165110 03
-9.6	3.479620 00	2.267810 01	9.165350 01	2.728250 02	6.507010 02	1.307960 03	2.292700 03
-9.7	3.498060 00	2.302700 01	9.393880 01	2.821040 02	6.784460 02	1.374410 03	2.426800 03
-9.8	3.516400 00	2.337770 01	9.625900 01	2.916140 02	7.071300 02	1.443680 03	2.567680 03
-9.9	3.534640 00	2.373030 01	9.861440 01	3.013570 02	7.367760 02	1.515860 03	2.715630 03
-10.0	3.552780 00	2.408470 01	1.010050 02	3.113380 02	7.674090 02	1.591060 03	2.870950 03

[1]P. Rhodes, Proc. Roy. Soc. A **204**, 396 (1950).
[2]A. Sommerfeld, Z. Phys. **47**, 1 (1928).
[3]L. Nordheim, "Muller Pomillets Lehrbuch der Physik" **4** (1934) Braunschweig, as quoted in Ref. 4.
[4]J. McDougall and E. C. Stoner, Philos. Trans. A **237**, 67 (1938).
[5]A. C. Beer, M. N. Chase, and P. F. Choquard, Helv. Phys. Acta **28**, 529 (1955).
[6]R. Latter, Phys. Rev. **99**, 1854 (1955).
[7]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables* (Dover, New York, 1972).
[8]A. Bañuelos, R. A. Depine, and R. C. Mancini (To be published in Comput. Phys. Commun.)

On nearest neighbor degeneracies of indistinguishable particles

Benjamin Kedem[a)]
Department of Mathematics, University of Maryland, College Park, Maryland and Department of Mathematics, Ben Gurion University of the Negev, Beer Sheva, Israel

(Received 6 June 1980; accepted for publication 25 September 1980)

Arrangement degeneracies suggested by sufficient statistics associated with binary stationary mth order Markov chains are discussed, and are shown to correspond and generalize some degeneracies arising when indistinguishable particles are placed on a one-dimensional lattice with n compartments. From these statistics it is possible to define an mth order unit. The arrangement degeneracy obtained from s 1's and $n-s$ 0's so that lower order units are placed in higher order is difficult. For this case only the third order arrangement degeneracy is obtained, the first and second orders being relatively simple. These results are applied in determining the asymptotic distributions of rare events.

PACS numbers: 02.50.Ga

1. INTRODUCTION

In statistical mechanical treatment of cooperative phenomena such as magnetic spin, binary alloys, elasticity, etc., the total energy of interaction E_i is given as a linear combination of potential energies associated with certain arrangments of indistinguishable particles on a one-dimensional lattice. For example, if only occupied nearest neighbors and next nearest neighbors of the types 11, 101, 111, are of interest then

$$E_i = n_{11}V_{11} + n_{101}V_{101} + n_{111}V_{111} \tag{1}$$

where the V's stand for potential energy and the n's refer to the frequency of occurrence of the neighbor type. Here 1 refers to an occupied site while 0 refers to a vacancy. Of interest then is knowledge of the arrangment degeneracy associated with the types of neighbors. That is, the number of binary sequences of size n which satisfy certain restrictions (numbers of types). Much attention to these combinatorial problems has been given by McQuistan in a series of articles in which he considered both simple and complex particles such as dumbbells, e.g., see Refs. 1–3. Parallel to the physical interpretation of arrangement degeneracies there is a purely statistical one in which arrangement degeneracies are used in approximating important distributions such as the distributions of crossings and upcrossings of a fixed level by a stationary process and the distribution of extremes in such processes. This has recently been dealt with in Ref. 4.

The purpose of this article is to show the connection between the purely statistical and statistical mechanical approaches in regard to arrangement degeneracies by examining *sufficient statistics* associated with mth order Markov chains. At the same time we shall extend some of McQuistan results by introducing higher order degeneracies whose usefulness will be demonstrated in finding the asymptotic distribution of "rare" events.

It should be made clear that we shall mainly deal with degeneracies associated with pairs in which the inner elements do not contain 1's, but in the last section we allude to more general degeneracies associated with more complicated pairs such as pairs of the type 0-1, 0-1-1, 1-0-0.

[a)]Research supported by AFOSR F49620-79-C-0095.

2. A CONNECTION BETWEEN NEAREST AND NEXT NEAREST NEIGHBOR DEGENERACY AND SUFFICIENCY

Let $\{X_t, t = 0, \pm 1, \cdots\}$ be a two state (0-1) stationary Markov chain and consider a binary time series from the chain, $X_1, \ldots, X_n$. The sufficient statistics associated with the chain are[4]

$$S = \sum_{i=1}^{n} X_i, \quad R_1 = \sum_{i=2}^{n} X_i X_{i-1}, \quad H = X_1 + X_n.$$

In order to find the joint distribution of S, R_1, H it is necessary to determine the number of binary sequences $M_n(s, r_1, h)$ for which $S = s, R_1 = r_1, H = h$. This number is easily found as follows. Form a succession of 0-cells (0-runs) and 1-cells (1-runs) in one of $\binom{2}{h}$ possible ways. We can now distribute the 1's and 0's in their respective cells so that none is empty. First place the s 1's in the $s - r_1$ 1-cells in

$$\binom{s-1}{s-r_1-1} = \binom{s-1}{r_1}$$

ways. Next place the $n - s$ 0's in the $s - r_1 - h + 1$ 0-cells in

$$\binom{n-s-1}{s-r_1-h}$$

ways. Therefore,

$$M_n(s, r_1, h) = \binom{2}{h}\binom{s-1}{r_1}\binom{n-s-1}{s-r_1-h}, \tag{2}$$

and the number of sequences for which only the first two conditions are satisfied is

$$M_n(s, r_1) = \sum_{h=0}^{2} M_n(s, r_1, h) = \binom{s-1}{r_1}\binom{n-s+1}{s-r_1}, \tag{3}$$

a result which was also obtained by McQuistan[2] for the degeneracy of nearest neighbor pairs.

If the chain $\{X_t\}$ is of second order then[4] the sufficient

 022-2488/81/030456-06$1.00

statistics are S, R_1, H together with $R_2 = \sum_{i=3}^{n} X_i X_{i-2}$, $C = \sum_{i=3}^{n} X_i X_{i-1} X_{i-2}$, $U = X_2 + X_{n-1}$, $V = X_1 X_2 + X_{n-1} X_n$. Again the joint distribution of these statistics requires the knowledge of the number of binary sequences $M_n(s, r_1, r_2, c, h, u, v)$ for which $S = s, \ldots, V = v$. This was found[5] to be

$$M_n(s, r_1, r_2, c, h, u, v) = \binom{2}{\max(h,u)} \binom{\max(h,u)}{v} \binom{r_1 - 1}{c} \times \binom{s - r_1 - h}{r_1 - c - v} \binom{s - r_1 - 1}{r_2 - c} \binom{n - 2s + r_1 + h - 2}{s - r_1 - r_2 + c - h - u + v}, \tag{4}$$

with the convention

$$\binom{-1}{-1} = 1$$

and where (h,u,v) takes value in $\{(0,0,0),(0,1,0),(0,2,0),(1,0,0),(2,0,0),(1,1,0)(1,1,1),(2,1,1),(1,2,1),(2,2,2)\}$. It follows that the number of binary sequences for which only the first four conditions are fixed is obtained by summing over (h,u,v). We have

$$\begin{aligned} M_n(s, r_1, r_2, c) &= \sum_{(h,u,v)} M_n(s, r_1, r_2, c, h, u, v) \\ &= \sum_{(0,0,0),\,(0,1,0),\,(0,2,0)} + \sum_{(1,0,0),\,(2,0,0),\,(1,1,0)} \\ &\quad + \sum_{(1,1,1),\,(2,1,1),\,(1,2,1)} + M_n(s, r_1, r_2, c, 2, 2, 2) \\ &= \binom{r_1 - 1}{c} \binom{s - r_1 - 1}{r_2 - c} \binom{s - r_1}{r_1 - c} \binom{n - 2s + r_1}{s - r_1 - r_2 + c} \\ &\quad + 2\binom{s - r_1 - 1}{r_1 - c} \binom{n - 2s + r_1}{s - r_1 - r_2 + c - 1} + \binom{n - 2s + r_1}{s - r_1 - r_2 + c - 2} \\ &\quad \times \binom{s - r_1 - 2}{r_1 - c} + 2\binom{s - r_1 - 1}{r_1 - c - 1} \binom{n - 2s + r_1}{s - r_1 - r_2 + c - 1} \\ &\quad + 2\binom{n - 2s + r_1}{s - r_1 - r_2 + c - 2} \binom{s - r_1 - 2}{r_1 - c - 1} \\ &\quad + \binom{n - 2s + r_1}{s - r_1 - r_2 + c - 2} \binom{s - r_1 - 2}{r_1 - c - 2} \\ &= \binom{r_1 - 1}{c} \binom{s - r_1 - 1}{r_2 - c} \binom{s - r_1}{r_1 - c} \left[\binom{n - 2s + r_1}{s - r_1 - r_2 + c} \right. \\ &\quad \left. + 2\binom{n - 2s + r_1}{s - r_1 - r_2 + c - 1} + \binom{n - 2s - r_1}{s - r_1 - r_2 + c - 2} \right] \\ &= \binom{r_1 - 1}{c} \binom{s - r_1}{r_1 - c} \binom{s - r_1 - 1}{r_2 - c} \binom{n - 2s + r_1 + 2}{s - r_1 - r_2 + c}. \end{aligned} \tag{5}$$

Upon noting that $r_1 = n_{11}, c = n_{111}, r_2 - c = n_{101}$, where $n_{ij\ldots k}$ refers to the frequency of $ij, \ldots, k$ in the binary sequence, we recognize the occupied next nearest neighbor degeneracy obtained by McQuistan[2]

$$M_n(s, n_{11}, n_{101}, n_{111}) = \binom{n_{11} - 1}{n_{111}} \binom{s - n_{11}}{n_{11} - n_{111}} \binom{s - n_{11} - 1}{n_{101}} \times \binom{n - 2s + n_{11} + 2}{s - n_{11} - n_{101}}. \tag{5'}$$

Obviously (5′) can be obtained from (5) by linearity. Here, and in what follows, we use the notation $M_n(c_1, \ldots, c_k)$ to denote the number of binary sequences for which $c_1, \ldots, c_k$ are fixed. It is readily seen that (5) reduces to (3) by summing over $r_2 - c$ and c.

3. HIGHER ORDER DEGENERACIES

In the previous section we illustrated via two examples that arrangement degeneracies may be viewed as special cases of counting problems associated with sufficient statistics in Markov chains. The next thing which comes to mind is the question of generalizations. It is quite clear now that if one wants to look into or define higher degeneracies one should examine the sufficient statistics and linear functions thereof of higher order two state Markov chains. By the very definition of sufficient statistics, it is intuitively clear that every conceivable arrangement degeneracy can be obtained by summing over M_n of a given order. We need not consider all the sufficient statistics associated with a given order but only those which define desired neighbor types. To make this point clear we shall defer the general counting problem to the next section while concentrating here on straightforward generalizations of (3) and (5). We shall first illustrate this claim by examining sufficient statistics associated with third order chains. In this connection the notion of a "unit" is useful

Definition: An mth order unit is a binary sequence which starts with a 1, ends with m separating 0's, (if needed to separate it from other units) and in which each 0-run, if not an end run, consists of at most $m - 1$ 0's. Thus a unit is a block made of 0-runs and 1-runs where the length of the 0-runs is restricted while the length of the 1-runs is unrestricted.

Definition: A free 0 is a 0 which does not belong to any unit. Observe that an end unit is a unit which does not need the separating 0's at the end in order to be recognized. The notion of a unit is useful as it determines the general *form* of a binary series which satisfies our predetermined conditions. For example, consider a first order chain and suppose it is desired to count the number of sequences for which S and R_1 are fixed at s and r_1. There are $s - r_1$ first order units which we permute with the free 0's in

$$\binom{s - r_1 + (n - s) - (s - r_1 - 1)}{s - r_1} = \binom{n - s + 1}{s - r_1}$$

ways. Next we place a 0 at the end of each first order unit in one way. This determines the form of the binary series. Finally distribute the s 1's in the $s-r_1$ units so that none is empty in

$$\binom{s-1}{r_1}$$

ways. This yields Eq. (3). Equation (5) can be obtained in the same way by first determining the positions of the second order units and then placing the first order units in the second order units. This is followed by the distribution of the s 1's. This procedure gives rise to an immediate extension of (3) and (5).

Consider the third order analog of (3) and (5). The statistics of interest are S,R_1,R_2,C, together with $K=\sum_{i=4}^{n}X_iX_{i-1}X_{i-2}X_{i-3}$ and n_{1001}, the frequency of 1001 in the sequence. These are sufficient statistics (not all of them!) associated with third order chains. We shall determine $M_n(s,r_1,r_2,c,k,n_{1001})$.

There are $(s-r_1)-(r_2-c)-n_{1001}$ third order units, $(s-r_1)-(r_2-c)$ second order units, and $s-r_1$ first order units. First permute the third order units with the free 0's in

$$\binom{n-3s+2r_1+r_2-c+3}{s-r_1-r_2+c-n_{1001}}$$

ways. Next place the second order units in the third order units leaving none empty in

$$\binom{s-r_1-r_2+c-1}{n_{1001}}$$

ways. Then place the first order units in the second order units in

$$\binom{s-r_1-1}{r_2-c}$$

ways and place the separating 0's in the respective units in one way. Put one 1 in each first order unit in one way. There remain r_1 1's. There are r_1-c first order units with two or more 1's. So choose r_1-c first order units from $s-r_1$ and put one 1 in each in

$$\binom{s-r_1}{r_1-c}$$

ways. There are $c-k$ first order units with three or more 1's. So select $c-k$ first order units from r_1-c and put 1 in each in

$$\binom{r_1-c}{c-k}$$

ways. Finally place the remaining k 1's in the $c-k$ first order units allowing "empty" units in

$$\binom{c-1}{k}$$

ways. Whence

$$M_n(s,r_1,r_2,c,k,n_{1001})=\binom{c-1}{k}\binom{r_1-c}{c-k}\binom{s-r_1}{r_1-c}\binom{s-r_1-1}{r_2-c}\times\binom{s-r_1-r_2+c-1}{n_{1001}}\binom{n-3s+2r_1+r_2-c+3}{s-r_1-r_2+c-n_{1001}},\quad(6)$$

and by summing over n_{1001} we immediately obtain

$$M_n(s,r_1,r_2,c,k)=\binom{c-1}{k}\binom{r_1-c}{c-k}\binom{s-r_1}{r_1-c}\binom{s-r_1-1}{r_2-c}\times\binom{n-2s+r_1+2}{s-r_1-r_2+c}$$

This last expression yields (5) upon summation over k. We can rewrite (6) in a form which resembles (5'). Let fz_3 denote the number of free 0's associated with a third order degeneracy. That is, the number of 0's which do not belong to any third order unit. Then $fz_3=(n-s)-3[(n-s)-n_{101}-n_{1001}-1]-2n_{1001}-n_{101}$, and (6) becomes

$$M_n(s,n_{11},n_{111},n_{1111},n_{101},n_{1001})=\binom{n_{111}-1}{n_{1111}}\binom{n_{11}-n_{111}}{n_{111}-n_{1111}}\times\binom{s-n_{11}}{n_{11}-n_{111}}\binom{s-n_{11}-1}{n_{101}}\binom{s-n_{11}-n_{101}-1}{n_{1001}}\times\binom{fz_3+s-n_{11}-n_{101}-n_{1001}}{s-n_{11}-n_{101}-n_{1001}}.\quad(6')$$

From (3),(5'),(6') we see that a pattern of arrangement degeneracies begins to emerge whereby the highest order units are permuted with the free 0's then lower order units are placed in higher order units, successively, then the separating 0's are placed in one way, this followed by the distributions of 1's so that $s,n_{11},n_{111},\ldots,n_{101},\ldots,n_{100\cdots1}$, are preserved. In order to arrive at the general result suggested by the above scheme we shall employ the notation $*(m)$ to mean a $*$-run of length m. Then we readily have

$$M_n(s,n_{11},n_{111},\ldots,n_{1(m+1)},n_{101},n_{1001},\ldots,n_{10(m-1)1})=\binom{n_{1(m)}-1}{n_{1(m+1)}}\binom{n_{1(m-1)}-n_{1(m)}}{n_{1(m)}-n_{1(m+1)}}\cdots\binom{n_{11}-n_{111}}{n_{111}-n_{1111}}\binom{s-n_{11}}{n_{11}-n_{111}}\times\binom{s-n_{11}-1}{n_{101}}\binom{s-n_{11}-n_{101}-1}{n_{1001}}\cdots\times\binom{s-n_{11}-n_{101}-\cdots-n_{10(m-2)1}-1}{n_{10(m-1)1}}\times\binom{fz_m+s-n_{11}-n_{101}-\cdots-n_{10(m-1)1}}{s-n_{11}-n_{101}-\cdots-n_{10(m-1)1}},\quad(7)$$

where fz_m stands for the number of free 0's and is given by

$$fz_m=(n-s)-m[(s-n_{11})-n_{101}-n_{1001}-\cdots-n_{10(m-1)1}-1]$$
$$-n_{101}-2n_{1001}-\cdots-(m-1)n_{10(m-1)1}. \quad (8)$$

Observe that the number of mth order units is equal to

$$(s-n_{11})-n_{101}-n_{1001}-\cdots-n_{10(m-1)1}, \quad (9)$$

since we essentially evaluate $\sum X_i(1-X_{i-1})\cdots(1-X_{i-m})$. From (7) we obtain (3),(5′),(6′) as special cases.

4. THE GENERAL ARRANGEMENT DEGENERACY OF MARKOV CHAINS AND ITS APPLICATION TO THE DISTRIBUTION OF RARE EVENTS

In the previous section we dealt with one way of extending the next nearest neighbor degeneracy. However, this is only a special case of the general arrangement degeneracy associated with an mth order Markov chain where the problem is to count the number of binary sequences for which all the sufficient statistics are fixed. This is a challenging problem for which no general solution exists as far as the present author knows. The main difficulty is the fact that not all the low order units can simply be placed in higher order units as in the previous section since some conditions are violated. Clearly, if such a result is available, numerous arrangement degeneracies can be deduced from it. This solution will not be attempted here. However, we shall give the solution for third order chains and this will give us a clue as to the general behavior of rare events in binary Markov chains.

The sufficient statistics associated with a stationary third order Markov chain, apart from ends statistics, are S,R_1,R_2,C,K as above together with

$$R_3=\sum_{i=4}^{n}X_iX_{i-3},\ R_{12}=\sum_{i=4}^{n}X_iX_{i-1}X_{i-3},\ R_{21}=\sum_{i=4}^{n}X_iX_{i-2}X_{i-3}.$$

We shall construct a sequence for which $S,R_1,\ldots,R_{21}$ are fixed. First observe that

$$R_3-R_{12}-R_{21}+K=n_{1001}. \quad (10)$$

Also define the statistics

$A_{11\text{-}11}\equiv$# of second order units which start and end with 11,

$A_{1\text{-}1}\equiv$# of second order units which start and end with 1,

$A_{11\text{-}1}\equiv$# of second order units which start with 11 and end with 1,

$A_{1\text{-}11}\equiv$# of second order units which start with 1 and end with 11,

$A_{11}\equiv$# of second order units which contain exactly two consecutive 1's,

$A_1\equiv$# of second order units which contain exactly one 1.

Clearly

$$\begin{aligned}&A_{11\text{-}11}+A_{11\text{-}1}+A_{1\text{-}11}+A_{1\text{-}1}+A_{11}+A_1=(S-R_1)-(R_2-C),\\&A_{11}+A_{11\text{-}11}+A_{11\text{-}1}=(R_1-C)-(R_{12}-K),\\&A_{11}+A_{11\text{-}11}+A_{1\text{-}11}=(R_1-C)-(R_{21}-K),\end{aligned} \quad (11)$$

which means that if A_1, A_{11}, and $A_{11\text{-}11}$ are known then so are $A_{1\text{-}11}$, $A_{11\text{-}1}$ and $A_{1\text{-}1}$. As it turns out our problem is simplified greatly if $A_1,A_{11},A_{11\text{-}11}$ are added to the other eight conditions. The reason for introducing the A's is that it is difficult to keep track of R_{12} and R_{21}, while from (11) it is seen that when the A's are fixed in addition to S,R_1,R_2,C,K, so are R_{12},R_{21}. When the arrangement degeneracy is obtained the A's can be removed by summation.

Recall that there are

$$(s-r_1)-(r_2-c)-(r_3-r_{12}-r_{21}+k)$$

third order units and

$$(n-s)-3[(s-r_1)-(r_2-c)-(r_3-r_{12}-r_{21}+k)-1]$$
$$-2(r_3-r_{12}-r_{21}+k)-(r_2-c)$$

free 0's which we permute in

$$\begin{pmatrix}(n-s)-2(s-r_1)+(r_2-c)+3\\(s-r_1)-(r_2-c)-(r_3-r_{12}-r_{21}+k)\end{pmatrix}$$

ways. There are $(s-r_1)-(r_2-c)$ second order units which we place in the third order units in

$$\begin{pmatrix}(s-r_1)-(r_2-c)-1\\r_3-r_{12}-r_{21}+k\end{pmatrix}$$

ways. Place the separating 0's in one way. Now, we *cannot* place the first order units in the second order units as we did in the previous section as R_{12},R_{21} change. This is precisely why we need the A's. So according to $A_{11},A_1,A_{11\text{-}1},A_{1\text{-}11},A_{11\text{-}11},A_{1\text{-}1}$, assign "types" to the second order units. That is let $A_{11\text{-}11}$ second order units start and end with 11, $A_{11\text{-}1}$ second order units start with 11 and end with 1, etc. This can be done in

$$\begin{pmatrix}(s-r_1)-(r_2-c)\\A_1,A_{11},A_{11\text{-}11},A_{11\text{-}1},A_{1\text{-}11},A_{1\text{-}1}\end{pmatrix}$$

ways. There are now r_2-c 0's left which we distribute in $A_{11\text{-}11}+A_{11\text{-}1}+A_{1\text{-}11}+A_{1\text{-}1}$ second order units "non-empty" in

$$\begin{pmatrix}r_2-c-1\\A_{11\text{-}11}+A_{11\text{-}1}+A_{1\text{-}11}+A_{1\text{-}1}-1\end{pmatrix}$$
$$=\begin{pmatrix}r_2-c-1\\(s-r_1)-(r_2-c)-(A_{11}+A_1)-1\end{pmatrix}$$

ways. This takes care of the 0's and the first order units as well! It remains now to distribute the 1's. According to the type assignment place 11 and 1 in the second order units as needed in one way, and then put 1 in every empty first order unit (1-cell or 1-run) in one way. There are r_1-c first order units with two or more 1's. But we have already $A_{11\text{-}1}+A_{1\text{-}11}+2A_{11\text{-}11}+A_{11}$ first order units with two 1's. So select $(r_1-c)-(A_{11\text{-}1}+A_{1\text{-}11}+2A_{11\text{-}11}+A_{11})$ first order units from $(s-r_1)-(A_{11\text{-}1}+A_{1\text{-}11}+2A_{11\text{-}11}+A_{11})-(A_{11\text{-}1}+A_{1\text{-}11}+2A_{1\text{-}1}+A_1)$ first order units in

$$\begin{pmatrix}(s-r_1)-(A_{11\text{-}1}+A_{1\text{-}11}+2A_{11\text{-}11}+A_{11})-(A_{11\text{-}1}-A_{1\text{-}11}+2A_{1\text{-}1}+A_1)\\(r_1-c)-(A_{11\text{-}1}+A_{1\text{-}11}+2A_{11\text{-}11}+A_{11})\end{pmatrix}$$

ways and put one 1 in each. We now have r_1-c first order units with exactly two 1's, since previously no

first order unit was empty. There are $c-k$ first order units with three or more 1's. From the r_1-c first order units which contain exactly two 1's select $c-k$ units and put one 1 in each in

$$\binom{r_1-c}{c-k}$$

ways. We now have $c-k$ first order units with exactly three 1's. Finally, there remain k 1's which we place in these $c-k$ first order units allowing "empty" units in

$$\binom{c-k+k-1}{c-k-1}=\binom{c-1}{k}$$

ways. It follows that

$$\begin{aligned}
&M_n(s,r_1,r_2,c,r_3,r_{12},r_{21},k,A_{11},A_1,A_{11-11})\\
&=\binom{c-1}{k}\binom{r_1-c}{c-k}\binom{(s-r_1)-2(A_{11-11}+A_{11-1}+A_{1-11}+A_{1-1})-(A_{11}+A_1)}{(r_1-c)-(A_{11-1}+A_{1-11}+2A_{11-11}+A_{11}}\binom{r_2-c-1}{(s-r_1)-(r_2-c)-(A_{11}+A_1)-1}\\
&\times\binom{(s-r_1)-(r_2-c)}{A_1,A_{11},A_{11-11},r_1-c-r_{12}+k-A_{11}-A_{11-11},r_1-c-r_{21}+k-A_{11}-A_{11-11},A_{1-1}}\\
&\times\binom{(s-r_1)-(r_2-c)-1}{r_3-r_{12}-r_{21}+k}\binom{(n-s)-2(s-r_1)+(r_2-c)+3}{(s-r_1)-(r_2-c)-(r_3-r_{12}-r_{21}+k)},
\end{aligned}\tag{12}$$

where A_{1-1} is obtained from (11) in terms of A_{11-11},A_{11}, $A_1,s,r_1,r_2,c,r_{12},r_{21},k$, and is too long to write. Thus the desired arrangement degeneracy is obtained by summing over A_{11},A_{11-11},A_1, adhering to the convention

$$\binom{-1}{-1}=1.$$

We have

$$\begin{aligned}
&M_n(s,r_1,r_2,c,r_3,r_{12},r_{21},k)\\
&=\binom{c-1}{k}\binom{r_1-c}{c-k}\binom{(s-r_1)-(r_2-c)-1}{r_3-r_{12}-r_{21}+k}\\
&\times\binom{(n-s)-2(s-r_1)+(r_2-c)+3}{(s-r_1)-(r_2-c)-(r_3-r_{12}-r_{21}+k)}\\
&\times\sum_{A_{11},A_1,A_{11-11}}B_n(s,r_1,r_2,c,r_3,r_{12},r_{21},k,A_{11},A_1,A_{11-11}),
\end{aligned}\tag{13}$$

where B_n is equal to the product of all the coefficients in (12) which involve A_{11},A_1,A_{11-11}. Note that

$$\sum_{A_{11},A_1,A_{11-11}}B_n(s,0,0,\ldots,0,A_{11},A_1,A_{11-11})=1,\tag{14}$$

since $A_1=s$ and the rest of the A's vanish.

Now the joint distribution of a binary time series from a third order stationary Markov chain is given by

$$\begin{aligned}
&p(x_1,x_2,\ldots,x_n)=(\text{powers in } x_1,x_2,x_3,x_{n-2},x_{n-1},x_n)\\
&\times p_{1011}^{r_{21}-k}p_{1101}^{r_{12}-k}p_{0101}^{r_2-c-r_{12}+k}p_{1010}^{r_2-c-r_{21}+k}p_{1001}^{r_3-r_{21}-r_{12}+k}p_{1111}^{k}p_{1110}^{c-k}p_{0111}^{c-k}\\
&\times p_{1100}^{r_1-c-r_{12}+k}p_{0011}^{r_1-c-r_{21}+k}p_{0110}^{r_1-2c+k}[p_{1000}p_{0001}]^{s-r_1-r_2+c-r_3+r_{12}+r_{21}-k}\\
&\times p_{0100}^{s-2r_1-r_2+2c+r_{12}-k}p_{0010}^{s-2r_1-r_2+2c+r_{21}-k}\\
&\times p_{0000}^{n-4s+3r_1+2(r_2-c)+r_3-r_{12}-r_{21}+k-3},
\end{aligned}\tag{15}$$

where $p_{x_ix_{i-1}x_{i-2}x_{i-3}}=Pr(X_i=x_i\,|\,X_{i-1}=x_{i-1},X_{i-2}=x_{i-2},X_{i-3}=x_{i-3})$, from which it follows that the joint distribution of $S,R_1,R_2,C,R_3,R_{12},R_{21},K,X_1,X_2,X_3,X_{n-2},X_{n-1},X_n$ is the product of (15) and the arrangement degeneracy of these statistics. But when $X_1,X_2,X_3,X_{n-2},X_{n-1},X_n$ are equal to 0 this arrangement degeneracy is asymptotically, as $n\to\infty$, the same as (13). Therefore from (13), (14), and (15)

$$\begin{aligned}
&Pr(S=s,R_1=0,R_2=0,\ldots,K=0,X_1=X_2=X_3=X_{n-2}=X_{n-1}=X_n=0)\\
&\sim\binom{n-3s+3}{s}p_{1000}^{s}p_{0001}^{s}p_{0100}^{s}p_{0010}^{s}p_{0000}^{n-4s-3}.
\end{aligned}\tag{16}$$

Assume that as $n\to\infty$

(i) The 1's become rare separated by long 0-runs so that the event $\{R_1=0,R_2=0,\ldots,K=0,X_1=X_2=X_3=X_{n-2}=X_{n-1}=X_n=0\}$ becomes a sure event;

(ii) $p_{1000}\to 0$ such that $np_{1000}=\alpha$, fixed;

(iii) $p_{0001}\sim p_{0100}\sim p_{0010}\sim p_{0000}=1-p_{1000}$.

Then from (16), as n becomes large,

$$\begin{aligned}
Pr(S=s)&\sim\frac{(n-3s+3)!}{(n-4s+3)!s!}\left(\frac{\alpha}{n}\right)^s\left(1-\frac{\alpha}{n}\right)^{n-s-3}\\
&=\frac{\alpha^s}{s!}\left(1-\frac{3s-3}{n}\right)\cdots\left(1-\frac{4s-4}{n}\right)\left(1-\frac{\alpha}{n}\right)^{-s-3}\left(1-\frac{\alpha}{n}\right)^{n}\\
&\to\frac{e^{-\alpha}\alpha^s}{s!},
\end{aligned}\tag{17}$$

a result which is well expected.[4]

This procedure for finding the asymptotic distribution of S as $n\to\infty$ can be easily extended to the mth order case. Arguing as above we have from (7) (as the last binomial coefficient in the most general case is given by the last coefficient in (7) which stands for the number of permutations of the mth order units with the free 0's)

$$\begin{aligned}
Pr(S=s)&\sim\binom{n-ms+m}{s}p_{10(m)}^{s}p_{0(m+1)}^{n-s-m}\\
&\sim\frac{\alpha^s}{s!}\frac{(n-ms)\cdots(n-(m+1)s+1)}{n^s}\left(1-\frac{\alpha}{n}\right)^n\to\frac{\alpha^s}{s!}e^{-\alpha},
\end{aligned}\tag{18}$$

where $n \to \infty$ and $p_{100\cdots0} \to 0$ such that $np_{100\cdots0} = \alpha$.

More results of this nature can be obtained once the arrangement degeneracy of a specific order is known. For example, from (12) it should not be too difficult to show that under some conditions similar to (i)–(iii) $S - R_1$ is also asymptotically Poisson. In fact Poisson with parameter $\alpha(1-\lambda)$, where $\lambda \in (0,1)$ is a measure of the density, or clutering tendencey, of rare events. For an interpretation of this fact see Ref. 4.

ACKNOWLEDGMENTS

The author wishes to thank Don Zagier for his useful remarks and interest in this problem.

[1]R. B. McQuistan, J. Math. Phys. **15**, 1192 (1974).
[2]R. B. McQuistan, J. Math. Phys. **15**, 1845 (1974).
[3]R. B. McQuistan, J. Math. Phys. **18**, 1191 (1977).
[4]B. Kedem, *Binary Time Series* (Marcel-Dekker, New York, 1980).
[5]B. Kedem, Discrete Math. **18**, 285 (1977).

Moment problem approximants to the Chandrasekhar H-equation[a)]

D. Masson
Department of Mathematics, University of Toronto, Toronto, M5S 1A1 Canada

(Received 3 September 1980; accepted for publication 7 October 1980)

Approximate solutions to the Chandrasekhar H-equation are obtained by considering a truncated moment problem. Convergence to the physical solution is proved and a numerical example is outlined.

PACS numbers: 02.60.Cb, 95.30.Jx

1. INTRODUCTION

The H-equation given by

$$H(z)=1+zH(z)\int_0^1 \frac{H(x)d\sigma(x)}{x+z} \tag{1}$$

and its operator-valued analogs play an important role in transport theory.[1-3] Iterative methods for its solution have been re-examined in the recent literature.[3,4] A typical result is that if $d\sigma/dx$ is a positive element of $L_1(0,1)$ with norm less than $\frac{1}{2}$ then the iterative scheme given by

$$H_{n+1}(z)=1+zH_{n+1}(z)\int_0^1 \frac{H_n(x)\,d\sigma(x)}{x+z} \tag{2}$$

converges monotonically and uniformly on $(0,1)$ to the "physical" solution $H(z)$. $H(z)$ is positive, bounded, continuous, and monotonic on $(0,\infty)$ and subject to the constraint $1/H(-\nu_0)=0, \nu_0>0$. This constraint is equivalent to the requirement $\Lambda(\nu_0)=0$, where $\Lambda(z)$ is the dispersive function given by

$$\Lambda(z)=1-2z^2\int_0^1 \frac{d\sigma(x)}{z^2-x^2}, \tag{3}$$

since one has the identity

$$(H(z)H(-z))^{-1}=\Lambda(z). \tag{4}$$

Here we examine approximants to the physical solution of (1) which are associated with the power-moment problem. It is not necessary to know σ. Only its moments are required. The approximants are introduced in Sec. 2. In Sec. 3 we examine their convergence. An explicit numerical example is given in Sec. 4.

2. THE APPROXIMANTS

A. Nonconservative case $c_0 < 1/2$

Let σ be a measure with known moments

$$c_n=\int_0^1 x^n\,d\sigma(x), \quad n=0,1,2\cdots \tag{5}$$

and with $c_0<1/2$.

Let σ_N be the measure, associated with the truncated problem of order $2N-1$, given by[5]

$$\frac{d\sigma_N}{dx}(x)=\sum_{i=1}^{N} a_{Ni}\delta(x-x_{Ni}), \tag{6}$$

with

$$\int_0^1 x^n\,d\sigma_N(x)=\int_0^1 x^n\,d\sigma(x), \quad n=0,1,\ldots,2N-1, \tag{7}$$

where the x_{Ni} are the poles and the a_{Ni} are the residues of the $[N-1/N]$ Padé approximant[6] to the function $\int_0^1 d\sigma(x)/(z-x)$.

Let $H_N(z)$ be the physical solution to the H-equation with the approximate measure σ_N. That is,

$$H_N(z)=1+zH_N(z)\int_0^1 \frac{H_N(x)\,d\sigma_N(x)}{z+x}. \tag{8}$$

This may be explicitly solved[1] to obtain

$$H_N(z)=(1-2c_0)^{-1/2}\prod_{i=1}^{N}\left(\frac{z+x_{Ni}}{z+\nu_{Ni}}\right), \tag{9}$$

where the ν_{Ni} are the positive zeros (intertwining the positive poles x_{Ni}) of the approximate dispersive function

$$\Lambda_N(z)=1-2z^2\int_0^1 \frac{d\sigma_N(x)}{z^2-x^2}. \tag{10}$$

Equation (9) follows easily from (8) and the identity $(H_N(z)H_N(-z))^{-1}=\Lambda_N(z)$ which, in particular, implies that

$$\lim_{|z|\to\infty}(H_N(z))^{-1}=\lim_{|z|\to\infty}(H(z))^{-1}=(1-2c_0)^{1/2}. \tag{11}$$

This correct asymptotic behavior (together with higher order identities in z^{-1}) is an attractive feature of the approximants associated with the moment problem.

B. Conservative case $c_0 = 1/2$

In the conservative case $c_0=1/2$ one has $\lim_{|z|\to\infty}H(z)=\infty$ and one needs a modification. This is easily obtained by passing to the limit $c_0=1/2$ from case A. From $H_N(0)=1$ one has that

$$\prod_{i=1}^{N}\left(\frac{x_{Ni}}{\nu_{Ni}}\right)=(1-2c_0)^{1/2}. \tag{12}$$

The second moment in (10) yields (the term in z^{-2})

$$\sum_{i=1}^{N}\nu_{Ni}^2-x_{Ni}^2=2c_2/(1-2c_0), \quad N\geq 2. \tag{13}$$

It follows that as $c_0\to 1/2-$ the largest zero ν_{NN} becomes infinite with the behavior

$$\nu_{NN}\approx(2c_2)^{1/2}/(1-2c_0)^{1/2}. \tag{14}$$

Taking the $c_0\to 1/2-$ limit of (9) thus yields for $N\geq 2$

[a)]Supported in part by the National Sciences and Engineering Research Council of Canada.

 002-2488/81/030462-03$1.00

$$H_N(z)=(2c_2)^{-1/2}(z+x_{NN})\prod_{i=1}^{N-1}\left(\frac{z+x_{Ni}}{z+\nu_{Ni}}\right). \tag{15}$$

3. CONVERGENCE

We are able to prove that the sequence of approximants $H_N(z)$ converges uniformly on $[0,\infty)$ to the physical solution $H(z)$ if c_0 is sufficiently small. The sufficiently small condition of our proof turns out to be $c_0<\frac{4}{9}$. A cleverer proof could presumably improve this to $c_0\leq\frac{1}{2}$.

A basic estimate is

$$1<H(z)<d \tag{16a}$$

and

$$1<H_N(z)<d, \tag{16b}$$

where $d=(1-2c_0)^{-1/2}$ and $z\in(0,\infty)$. For small z the upper bound may be improved by using

$$H(z)<1+zd\int_0^1\frac{d\sigma(x)}{x+z} \tag{17a}$$

and

$$H_N(z)<1+zd\int_0^1\frac{d\sigma_N(x)}{x+z}. \tag{17b}$$

From (16a), (16b) and (17a), (17b) one has

$$-zd\int_0^1\frac{d\sigma(x)}{x+z}<H_N(z)-H(z)<zd\int_0^1\frac{d\sigma_N(x)}{x+z} \tag{18}$$

for $z\in(0,\infty)$. Using the estimate[6]

$$\int_0^1\frac{d\sigma_N(x)}{x+z}\leq\int_0^1\frac{d\sigma(x)}{x+z},\quad z\in(0,\infty), \tag{19}$$

one has

$$|H_N(z)-H(z)|<zd\int_0^1\frac{d\sigma(x)}{x+z},\quad z\in(0,\infty). \tag{20}$$

If σ is continuous at $x=0$ it follows that

$$|H_N(z)-H(z)|<\epsilon,\quad 0\leq z<\delta(\epsilon), \tag{21}$$

with δ independent of N. It remains to obtain estimates for $z\geq\delta$.

From (1) and (8) one obtains

$$H_N(z)-H(z)=zH_N(z)H(z)\left(\int_0^1\frac{H_N(x)\,d\sigma_N(x)}{x+z}-\int_0^1\frac{H(x)\,d\sigma(x)}{x+z}\right). \tag{22}$$

Thus

$$|H_N(z)-H(z)|\leq H_N(z)d\left(\int_0^1\frac{z|H_N(x)-H(x)|\,d\sigma_N(x)}{x+z}+\left|\int_0^1\frac{zH(x)[d\sigma_N(x)-d\sigma(x)]}{x+z}\right|\right). \tag{23}$$

Now $zH(x)/(x+z)$ is continuous in x if $z\geq\delta>0$. It follows from Theorem 64.1 in Ref. 7 [provided that σ is continuous at $x=0,1$ and a dense set of points in $(0,1)$] that

$$\lim_{N\to\infty}\int_0^1\frac{zH(x)\,d\sigma_N(x)}{x+z}=\int_0^1\frac{zH(x)\,d\sigma(x)}{x+z}. \tag{24}$$

Thus

$$\left|z\int_0^1\frac{H(x)[d\sigma_N(x)-d\sigma(x)]}{x+z}\right|<\epsilon_1,\quad N>M_1(z,\epsilon_1). \tag{25}$$

Let $N_1=\sup_{z\in[\delta,\infty)}M_1(z,\epsilon_1)$. Then

$$|H_N(z)-H(z)|\leq H_N(z)d\int_0^1\frac{z|H_N(x)-H(x)|}{x+z}\,d\sigma_N(x)+\epsilon_1d^2,\quad N>N_1(\epsilon_1),\quad z\geq\delta. \tag{26}$$

Similarly one may start from an equivalent form of (22)

$$H_N(z)-H(z)=-H_N(z)H(z)\int_0^1\frac{x[H_N(x)\,d\sigma_N(x)-H(x)\,d\sigma(x)]}{x+z} \tag{27}$$

to obtain

$$|H_N(z)-H(z)|<dH_N(z)\int_0^1\frac{x|H_N(x)-H(x)|\,d\sigma_N(x)}{x+z}+\epsilon_2d^2,\quad z\in[0,\infty),\quad N>N_2, \tag{28}$$

where

$$\left|\int_0^1\frac{xH(x)[d\sigma_N(x)-d\sigma(x)]}{x+z}\right|<\epsilon_2\quad\text{for } N>M_2(z,\epsilon_2)$$

and $N_2=\sup_{z\in[0,\infty)}M_2(z,\epsilon_2)$. Adding together inequalities (21), (26), and (28) one obtains

$$2|H_N(z)-H(z)|<dH_N(z)\int_0^1|H_N(x)-H(x)|\,d\sigma_N(x)+\epsilon+d^2(\epsilon_1+\epsilon_2) \tag{29}$$

for $z\in[0,\infty)$ and N sufficiently large.

Multiplying (29) by $d\sigma_N(z)$, integrating, and solving algebraically for $\int_0^1|H_N(x)-H(x)|\,d\sigma_N(x)$ yields the estimate

$$\int_0^1|H_N(x)-H(x)|\,d\sigma_N(x)<c_0[\epsilon+d^2(\epsilon_1+\epsilon_2)]/\left(2-d\int_0^1H_N(x)\,d\sigma_N(x)\right) \tag{30}$$

provided that

$$2-d\int_0^1H_N(x)\,d\sigma_N(x)>0, \tag{31}$$

where $\int_0^1H_N(x)\,d\sigma_N(x)=1-1/d$ and $d=(1-2c_0)^{-1/2}$. Thus (31) is the condition that $3-(1-2c_0)^{-1/2}>0$ or $c_0<\frac{4}{9}$. Thus if $c_0<\frac{4}{9}$ and $\sigma(x)$ is continuous at $x=0,1$ and a dense set of points in $(0,1)$ then $\int_0^1|H_N(x)-H(x)|\,d\sigma_N(x)$ is arbitrarily small for N sufficiently large. Estimates (29) and (16) then yield $\lim_{N\to\infty}H_N(z)=H(z)$ uniformly for $z\in[0,\infty)$.

4. NUMERICAL EXAMPLE

The simple case $\sigma(x)=cx$ has been extensively tabulated.[1,2] Here we compare the exact results for $c=\frac{1}{4}$ and $c=\frac{1}{2}$ with the lowest-order approximants $N=1$, 2, and 3, which are easily obtained by hand calculation.

Solving Eq. (7) for the a_{Ni} and x_{Ni} in (6) yields

$$a_{11}=c,\quad x_{11}=1/2,$$

$$a_{21}=c/2,\quad x_{21}=1/2-1/2\sqrt{3}=0.211325,$$

TABLE I. ($c=1/4$) approximate solutions $H_N(z)$ and exact solution $H(z)$ to the Chandrasekhar H-equation for $\sigma(x)=x/4$.

z	$H_1(z)$	$H_2(z)$	$H(z)$
0.2	1.0914	1.1096	1.1135
0.4	1.1496	1.1666	1.1680
0.6	1.1901	1.2039	1.2044
0.8	1.2199	1.2308	1.2309
1.0	1.2426	1.2513	1.2513

$$a_{22}=c/2\,,\quad x_{22}=1/2+1/2\sqrt{3}=0.788675$$

and

$$a_{31}=5c/18\,,\quad x_{31}=(1-\sqrt{3/5})/2=0.112702\,,$$

$$a_{32}=4c/9\,,\quad x_{32}=1/2\,,$$

$$a_{33}=5c/18\,,\quad x_{33}=(1+\sqrt{3/5})/2=0.887298\,.$$

The approximate dispersive function Λ_N given by Eq. (10) can then be calculated.

$$\Lambda_1(z)=(z^2(1-2c)-1/4)/(z^2-1/4)\,,$$

$$\Lambda_2(z)=\frac{z^4(1-2c)-2z^2(1-c)/3+1/36}{z^4-2z^2/3+1/36}\,,$$

$$\Lambda_3(z)$$

$$=\frac{z^6(1-2c)-z^4(63-43c)/60+z^2(21-12c)/100-1/400}{z^6-21z^4/20+21z^2/100-1/400}\,.$$

From this the positive zeros ν_{Nl}, $N=1,2$, can then be obtained.

$$\nu_{11}(c)=\tfrac{1}{2}(1-2c)^{1/2}\,,$$

$$\nu_{\substack{21\\(2)}}(c)=[(1-c_{(+)}((1-c)^2-(1-2c)/4)^{1/2}/3(1-2c)]^{1/2}$$

Equation (9) then yields H_N, $N=1,2$,

$$H_1(z)=(z+\tfrac{1}{2})/(z(1-2c)^{1/2}+\tfrac{1}{2})\,.$$

$$H_2(z)=\frac{z^2+z+1/6}{(1-2c)^{1/2}(z^2+(\nu_{21}(c)+\nu_{22}(c))z+1/6(1-2c)^{1/2})}\,.$$

For $c=\frac{1}{4}$ one thus has

$$\nu_{11}(1/4)=1/\sqrt{2}=0.707107\,,$$

TABLE II. ($c=1/2$) approximate solutions $H_N(z)$ and exact solution $H(z)$ to the Chandrasekhar H-equation for $\sigma(x)=x/2$.

z	$H_1(z)$	$H_2(z)$	$H_3(z)$	$H(z)$
0.2	1.4000	1.4414	1.4490	1.4503
0.4	1.8000	1.8276	1.8294	1.8293
0.6	2.2000	2.1959	2.1946	2.1941
0.8	2.6000	2.5562	2.5532	2.5527
1.0	3.0000	2.9121	2.9082	2.9078

$$\nu_{21}(1/4)=(1/2-\sqrt{7}/6)^{1/2}=0.242984\,,$$

$$\nu_{22}(1/4)=(1/2+\sqrt{7}/6)^{1/2}=0.970030\,,$$

and the values in Table I. For the limiting case $c=\frac{1}{2}$ one obtains $[\nu_{NN}(1/2)=\infty]$

$$\nu_{21}(1/2)=1/2\sqrt{3}=0.288675\,,$$

$$\nu_{31}(1/2)=\tfrac{1}{2}[(9-\sqrt{69})/10]^{1/2}=0.131660\,,$$

$$\nu_{32}(1/2)=\tfrac{1}{2}[(9+\sqrt{69})/10]^{1/2}=0.657773\,,$$

$$H_1(z)=1+2z\,,$$

$$H_2(z)=\frac{z^2+z+1/6}{z/\sqrt{3}+1/6}\,,$$

$$H_3(z)=\frac{\sqrt{3}(z^3+3z^2/2+3z/5+1/20)}{z^2+[(9-\sqrt{69})^{1/2}+(9+\sqrt{69})^{1/2}]z/2\sqrt{10}+\sqrt{3/20}}\,,$$

and the values in Table II.

[1] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[2] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, Mass., 1967).

[3] C. T. Kelley, J. Math. Phys. 21, 408, 1980.

[4] R. L. Bowden, J. Math. Phys. 20, 608, 1979.

[5] N. I. Akhiezer, *The Classical Moment Problem* (Oliver and Boyd, Edinburgh, 1965).

[6] G. A. Baker, Jr., "The Padé Approximant Method and Some Related Generalizations," Chap. 1 in *The Padé Approximant in Theoretical Physics*, edited by G. A. Baker, Jr. and J. L. Gammel (Academic, New York, 1970).

[7] H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, Princeton, N. J., 1948).

An exact invariant for a class of time-dependent anharmonic oscillators with cubic anharmonicity

P. G. L. Leach
Department of Applied Mathematics, La Trobe University, Bundoora, 3083, Australia

(Received 23 October 1979; accepted for publication 29 February 1980)

An exact invariant is constructed for a class of time-dependent anharmonic oscillators using the method of the Lie theory of extended groups. The presence of the anharmonic term imposes a constraint on the nature of the time dependence. For a sub-class it is possible to obtain an energy-like integral and a condition under which the motion is bounded.

PACS numbers: 03.20. + i

1. INTRODUCTION

In the investigation of the behavior of plasma one of the models which was early adopted was that of the motion of a charged particle in an axially symmetric field. An advantage of such a model was that, under suitable approximations, the radial equation reduced to that of the simple harmonic oscillator. One could not ask for a better model as far as the resulting mathematics was concerned. However, the Zeta machine and its like did not work and a more refined model was needed. This lead to the equation of the time-dependent harmonic oscillator to take into account time-varying fields. This oscillator system had attracted earlier attention, primarily as an approximation to the lengthening pendulum. Unfortunately, unlike the time-independent oscillator, there was no known exact invariant for the time-dependent oscillator.

That deficiency was overcome by Lewis[1] using a method based on Kruskal's scheme.[2] The Lewis invariant has attracted a considerable amount of attention from a variety of viewpoints.[3–5] We were able to offer a simple derivation and interpretation of the invariant[6] and to provide invariants for similar linear systems.[7] Indeed each member of the whole class of quadratic Hamiltonians was shown to be equivalent to any other member to within a (time-dependent) linear canonical transformation. The application of these results to quantum mechanics was begun by Lewis and Riesenfeld[8] and extended by us.[9] Naturally for quantum mechanics the class of permissible linear transformations is restricted, but, for a time-dependent quadratic Hamiltonian of constant signature, there would appear to be no difficulty. In fact the quantum mechanical results for time-dependent linear transformations closely parallel those for the time-independent case as reported by Wolf and others.[10]

We have not heard of the Zeta machine for many years, but it appears as if its successors require an even better model than that of the time-dependent harmonic oscillator. It was suggested[11] that a time-dependent anharmonic oscillator with cubic anharmonicity in the Hamiltonian would be suitable as a starting point and that an invariant for such a system could be informative. Assuming that a nonlinear system required nonlinear transformations, we proposed a scheme for constructing an invariant related to the Hamiltonian by means of such a transformation.[12] In general it was anticipated that the transformation would be an infinite series and probably divergent as well. However, in celestial mechanics truncation of similar series has been used with some success, for instance in the work of Gustavson[13] on the notorious Hénon–Heiles problem.

To provide the solution to a problem in the form of an infinite series is sometimes acceptable (as in the case of the ordinary oscillator), but generally speaking it may be regarded as less than satisfactory. The whole advantage of the Lewis invariant is that it is concise, easy to work and has a precise "physical" interpretation. We became convinced that useful transformations must nearly always be linear although the opposite viewpoint has recently been advanced by Maharatna, Dutt, and Chattarji.[14] It is our opinion that the results obtained here support our viewpoint.

It is evident that in general a time-dependent problem will not possess an invariant. This is not surprising. The point is to be developed elsewhere. It will be seen in the case of the problem discussed here that the determination and interpretation of such invariants rely on point transformations of the type $t \rightarrow T$, $q \rightarrow Q$. In the Hamiltonian context this means a linear canonical transformation coupled with a change of time scale. That such should be the case is fortunate for the results may readily be extended to quantum mechanics. Indeed the linear transformation belongs to the class of transformations for which the Schrödinger wavefunctions are related by means of a geometric transform rather than the more general integral transform.[10a].

Apart from the context of Hamilton–Jacobi theory, time-dependent transformations have not much been used until recently. Nevertheless we do express some surprise that the methods employed here have not been adopted generally. Specifically, the theories of Lie and Noether have been around for a considerable time and yet, to our knowledge, have only been applied, in a context similar to the present one, in recent years. Many of the ideas employed in this paper have been developed in earlier papers to which reference is made when appropriate.

2. THE PROBLEM

As reported above, it has been suggested that a better model for the motion of a charged particle in an axial field should include allowance for anharmonicity. Let us be more generous and allow velocity dependent damping and a coordinate free forcing term. Whether the addition of such terms is helpful to the model is unknown to us and we leave that matter to the physicists. All we wish to do here is to

 022-2488/81/030465-06$1.00

provide the maximum possible flexibility for the model.

The Newtonian equation of motion for such a particle may be written as

$$\ddot{q} + a(t)\dot{q} + b(t)q + c(t)q^2 + d(t) = 0. \tag{2.1}$$

The time-dependent parameters a, b, c, and d are not specified as to properties, but are assumed to be as good as the occasion requires. Following Caldriola and Kanai,[15] we use an integrating factor to construct a Lagrangian which is

$$L = \tfrac{1}{2}\dot{q}^2 A - (\tfrac{1}{2}bq^2 + \tfrac{1}{3}cq^3 + dq)A, \tag{2.2}$$

where

$$A = A(t) = \exp\int_{t_0}^{t} a(t')\,dt'. \tag{2.3}$$

The conjugate momentum is

$$p = qA \tag{2.4}$$

and the Hamiltonian is

$$H = \tfrac{1}{2}p^2 A^{-1} + (\tfrac{1}{2}bq^2 + \tfrac{1}{3}cq^3 + dq)A. \tag{2.5}$$

Under the change of time scale given by

$$T = \int_{t_0}^{t} A^{-1}(t')\,dt', \tag{2.6}$$

the problem may be discussed in terms of the equivalent system with Hamiltonian

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}bq^2 + \tfrac{1}{3}cq^3 + dq. \tag{2.7}$$

Note that under the succession of transformations used here, it is the form of the expression for the Hamiltonian rather than the precise relationship of one symbol to the corresponding proceeding one which is of interest. The alternative normal procedure would require a new set of symbols for each equation and naturally would be followed in a practical application. Under the translation

$$q \to q + r(t), \quad p \to p + s(t), \tag{2.8}$$

we obtain a new Hamiltonian of form

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}bq^2 + \tfrac{1}{3}cq^3 + g(t), \tag{2.9}$$

provided, in terms of the coefficients of Eq. (2.7),

$$\ddot{r} + br - cr^2 - d = 0, \quad s = \dot{r}. \tag{2.10}$$

As far as Eq. (2.9) is concerned, we may ignore the garbage term $g(t)$ as it plays no role in the equations of motion. It should be noted that in quantum mechanics $g(t)$ would appear in the phase.

Apart from the cubic term, Eq. (2.9) is now in time-dependent oscillator form and so the appropriate transformation is (cf. Ref. 6)

$$Q = \rho^{-1}q, \quad P = \rho p - \dot{\rho}q, \tag{2.11}$$

where $\rho(t)$ satisfies the auxiliary equation

$$\ddot{\rho} + b\rho = \rho^{-3}. \tag{2.12}$$

The Hamiltonian now takes the form

$$H = \rho^{-2}\{\tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{3}cq^3\}, \tag{2.13}$$

which, under the change of time scale

$$T = \int_{t_0}^{t} \rho^{-2}(t')\,dt', \tag{2.14}$$

becomes

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{3}Bq^3. \tag{2.15}$$

Thus we see that all Newtonian equations of motion of the form of Eq. (2.1) may be discussed in terms of the Newtonian equation

$$\ddot{q} + q + B(t)q^2 = 0, \tag{2.16}$$

provided the original linear coefficient is sufficient to warrant the final positive sign. In terms of the physical model this restriction is reasonable. We are looking at a physical situation which does involve attraction to the first order. Otherwise kinking is inevitable.

3. CHOICE OF APPROACH

The initial problem has been reduced to a discussion of the three (equivalent) alternative forms, respectively, the Hamiltonian, the Langrangian, and the Newtonian

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}q^2 + \tfrac{1}{3}Bq^3, \tag{3.1}$$

$$L = \tfrac{1}{2}\dot{q}^2 - \tfrac{1}{2}q^2 - \tfrac{1}{3}Bq^3, \tag{3.2}$$

$$N = \ddot{q} + q + Bq^2 = 0. \tag{3.3}$$

Considering that we have possible quantum mechanical applications in mind, we must pose the question: Upon which of Eqs. (3.1), (3.2), or (3.3) do we base our analysis? The Hamiltonian is most closely related to quantum mechanics. However, our experience of applying transformations to nonlinear Hamiltonians has not been happy.[12,16] The Lagrangian is not far removed from the Hamiltonian and is susceptible to treatment by the generalized Noether's theorem. Furthermore any invariant found in this way has a corresponding Hamiltonian invariant.[17] Also the invariants are easily found using Noether's theorem as there is an explicit formula. On the other hand, the Newtonian equation of motion is the most generous when it comes to providing invariants derived by the method of the Lie theory of extended groups.[18]

The approach chosen here is that of the Lie theory of extended groups. Those who are familiar with the method will know that the price of generality is a more awkward determination of invariants *vis à vis* the Noether method. This drawback may be allieviated by the use of point transformations which we have discussed elsewhere.[19] This procedure is particularly effective when there is only one possible invariant. Essentially a new time and coordinate system is found in which the invariant is a function of coordinate and velocity only. The transformation is one of time scale and linear in the coordinate. This relates very well to the use of change of time scale and linear canonical transformation in the Hamiltonian formalism. It will be particularly attractive if the invariant (in the new coordinates) is energy-like since this will have useful quantum mechanical applications. If it is not, then at least we will have an invariant.

4. THE METHOD OF THE LIE THEORY OF EXTENDED GROUPS

Although the Lie theory of extended groups has enjoyed a rightful resurgence of attention recently,[20] it may not be familiar to all readers. Accordingly we provide a brief

resumé relevant to Newtonian equations of motion. Suppose there exists a transformation with generator

$$G = \xi(q,t)\frac{\partial}{\partial t} + \eta(q,t)\frac{\partial}{\partial q}. \tag{4.1}$$

(We restrict ourselves to the one-dimensional problem.) If under the transformation generated by this operator a Newtonian equation of motion is to remain invariant, then, writing the equation as

$$N(\ddot{q}, \dot{q}, q, t) = 0, \tag{4.2}$$

we require

$$G^{(2)}N(\ddot{q}, \dot{q}, q, t) = 0, \tag{4.3}$$

whenever Eq. (4.2) is satisfied. The second extention of G, denoted by $G^{(2)}$ is given by

$$G^{(2)} = G + \eta^{(1)}\frac{\partial}{\partial \dot{q}} + \eta^{(2)}\frac{\partial}{\partial \ddot{q}}, \tag{4.4}$$

$$\eta^{(1)} = \dot{\eta} - \dot{\xi}\dot{q}, \quad \eta^{(2)} = \ddot{\eta} - \ddot{\xi}\dot{q} - 2\dot{\xi}\ddot{q}. \tag{4.5}$$

Presuming that Eq. (4.3) has non-trivial solutions for ξ and η, a constant of the motion may be found by imposing the double requirement that

$$G^{(1)}\Phi(\dot{q},q,t) = 0, \tag{4.6}$$

$$D\Phi(\dot{q},q,t) = 0, \tag{4.7}$$

where $D \equiv d/dt$.

In the analysis of certain linear systems[21] it became obvious that the task of finding the G's, let alone solving Eqs. (4.6) and (4.7), was complicated. It had been observed that linear systems of the same dimension had the same symmetry group, Sl($n+2$,R).[22] There was also the Hamiltonian result about the equivalence of linear systems under linear canonical transformations. This suggested that a point transformation of the type $t \to T$, $q \to Q$ would simplify matters. In quantum mechanics a time-independent energy-type invariant is desirable[23] and in classical mechanics it is most useful. The generator for such a constant is

$$G = \frac{\partial}{\partial t}, \tag{4.8}$$

provided of course that the appropriate space–time frame of reference is used. Supposed that in solving Eq. (4.3) a generator of the form

$$G = f(t)\frac{\partial}{\partial t} + \{g(t)q + h(t)\}\frac{\partial}{\partial q} \tag{4.9}$$

is obtained. All one needs to do is to change to new space and time variables, Q and T, by means of transformation linear in Q such that now

$$G = \partial/\partial T \tag{4.10}$$

and an invariant independent of the new time may be obtained. As a function of Q and $\dot{Q}$ it may be an energy type integral, in which case it has suitable quantum mechanical features. If the invariant is not of that form, the quantum mechanical applications are not so obvious, but at least an invariant does exist. We note that it has been found that, in the case for which there are several generators, a common transformation reduces them to a commonly simpler form. This is the case with linear systems.[24] For nonlinear systems the existence of several generators does not usually occur. Indeed it should be mentioned in passing that even in the instance of integrable nonlinear systems there may only be one generator.[25] This seems strange as an integrable system has two constants of integration for a one-dimensional system. It is hoped to discuss this point at another time.

To conclude this section we make a brief summary of the generalized Noether's theorem. Suppose a transformation with generator G leaves the action integral invariant, with G as defined as in Eq. (4.1). Then there exists an invariant given by

$$\Phi(\dot{q},q,t) = -\{\xi L + (\eta - \xi\dot{q})(\partial L/\partial\dot{q}) + f(q,t)\}, \tag{4.11}$$

where $f(q,t)$ is determined along with ξ and η.

5. A USEFUL RESULT

As a result, which is useful in this work and which does not appear to have been stated before, concerns the form of the generator G for the class of Newtonian equations of motion

$$N(\ddot{q},q,t) \equiv \ddot{q} + g(q,t) = 0. \tag{5.1}$$

Under the requirement that

$$G^{(2)}N(\ddot{q},q,t) = 0, \tag{5.2}$$

we have

$$\xi\frac{\partial g}{\partial t} + \eta\frac{\partial g}{\partial q} + (\ddot{\eta} - \ddot{\xi}\dot{q} - 2\dot{\xi}\ddot{q}) = 0. \tag{5.3}$$

Remembering that both ξ and η are functions of q and t only, we separate coefficients of powers of $\dot{q}$ to obtain

$$\frac{\partial^2\xi}{\partial q^2} = 0, \tag{5.4}$$

$$\frac{\partial^2\eta}{\partial q^2} - 2\frac{\partial^2\xi}{\partial q\partial t} = 0, \tag{5.5}$$

$$2\frac{\partial^2\eta}{\partial q\partial t} - \frac{\partial^2\xi}{\partial t^2} + 3g\frac{\partial\xi}{\partial q} = 0, \tag{5.6}$$

$$\frac{\partial^2\eta}{\partial t^2} - g\frac{\partial\eta}{\partial q} + 2g\frac{\partial\xi}{\partial t} + \xi\frac{\partial g}{\partial t} + \eta\frac{\partial g}{\partial q} = 0. \tag{5.7}$$

From (5.4)

$$\xi = a(t) + b(t)q. \tag{5.8}$$

From (5.5)

$$\eta = \dot{b}(t)q^2 + c(t)q + d(t). \tag{5.9}$$

Thus ξ is at most linear in q and η at most quadratic in q. We shall use these forms in the development below.

A similar result may be obtained for Noether's theorem. The Lagrangian corresponding to the Newtonian equation (5.1) is

$$L = \tfrac{1}{2}\dot{q}^2 - F(q,t), \quad F(q,t) = \int^q g(q',t)\,dq'. \tag{5.10}$$

The Noether invariant (4.11) has the form

$$\Phi(\dot{q},q,t) = \xi\{\tfrac{1}{2}\dot{q}^2 + F(q,t)\} - \eta\dot{q} - f(q,t). \tag{5.11}$$

Taking the total time derivative of (5.11) and equating pow-

ers of $\dot{q}$ to zero we obtain

$$\frac{\partial \xi}{\partial q} = 0, \tag{5.12}$$

$$\frac{1}{2}\frac{\partial \xi}{\partial t} - \frac{\partial \eta}{\partial q} = 0, \tag{5.13}$$

$$F\frac{\partial \xi}{\partial q} - \frac{\partial \eta}{\partial t} - \frac{\partial f}{\partial q} = 0, \tag{5.14}$$

$$F\frac{\partial \xi}{\partial t} + \xi\frac{\partial F}{\partial t} + \eta g - \frac{\partial f}{\partial t} = 0. \tag{5.15}$$

From Eq. (5.12) we have

$$\xi(q,t) = a(t), \tag{5.16}$$

and using this in Eq. (5.13),

$$\eta(q,t) = \tfrac{1}{2}\dot{a}(t)\,q + b(t). \tag{5.17}$$

From Eq. (5.14) we find that $f(q,t)$ has the form

$$f(q,t) = -\tfrac{1}{4}\ddot{a}(t)\,q^2 - \dot{b}(t)\,q. \tag{5.18}$$

The arbitrary function of time from the integration of Eq. (5.14) is neglected. The possible expressions for $a(t)$ and $b(t)$ are then found from Eq. (5.15). At this point we simply note that the Noether's theorem generators are simpler in general form to those obtained using the Lie theory. A comparison of the expressions for ξ and η might tempt some to identify $c(t)$ in (5.9) with $\frac{1}{2}\dot{a}(t)$. As will be seen below, this is not the case.

6. APPLICATION OF THE LIE THEORY TO THE ANHARMONIC OSCILLATOR

The reduced problem is the existence of the invariants for the Newtonian equation of motion

$$\ddot{q} + q + B(t)q^2 = 0. \tag{6.1}$$

From Sec. 5 we know that the generator is of the form

$$G(q,t) = (a + bq)\partial/\partial t + (\dot{b}q^2 + cq + d)\partial/\partial q, \tag{6.2}$$

only $\eta^{(2)}$ is required and it is

$$\begin{aligned}\eta^{(2)} &= \ddot{\eta} - \ddot{\xi}\dot{q} - 2\dot{\xi}\ddot{q} \\ &= \ddot{b}q^2 + 3\dot{b}q\dot{q} - 3b\dot{q}\ddot{q} \\ &\quad + \ddot{c}q + 2\dot{c}\dot{q} + c\ddot{q} + \ddot{d} - \dot{q}\ddot{a} - 2\ddot{q}\dot{a}.\end{aligned} \tag{6.3}$$

Substituting into

$$G^{(2)}(\ddot{q} + q + Bq^2) = 0, \tag{6.4}$$

and equating powers of $\dot{q}$ to zero, we find that

$$b = 0, \tag{6.5}$$

$$\ddot{d} + d = 0, \tag{6.6}$$

$$\ddot{c} + 2\dot{a} + 2Bd = 0, \tag{6.7}$$

$$a\dot{B} + Bc + 2\dot{a}B = 0, \tag{6.8}$$

$$2\dot{c} - \ddot{a} = 0. \tag{6.9}$$

From Eqs. (6.6–9) it follows that

$$d = D\sin t + E\cos t, \tag{6.10}$$

$$2c = \dot{a} + \alpha, \tag{6.11}$$

$$B = Ka^{-5/2}\exp\left\{-\tfrac{1}{2}\alpha\int^t dt'/a(t')\right\}, \tag{6.12}$$

where α, K, D, and E are constants and $a(t)$ satisfies the third order equation

$$\dddot{a} + 4\dot{a} + 2Bd + 0. \tag{6.13}$$

Thus the generator of a one parameter symmetry group for Eq. (6.1) is

$$G(q,t) = a\partial/\partial t + \{\tfrac{1}{2}(\dot{a} + \alpha)q + d\}\partial/\partial q, \tag{6.14}$$

provided $B(t)$ takes the form specified in Eq. (6.12). In view of the result for Noether's theorem, a generator for a transformation which preserves the action is

$$G(q,t) = a\partial/\partial t + (\tfrac{1}{2}\dot{a}q + d)\partial/\partial q, \tag{6.15}$$

i.e., α must be taken as zero.

From Eq. (6.13) we see that for $d \neq 0$, there is a considerable increase in the complexity of the differential equation defining a. From Eq. (6.12) the range of functions $B(t)$ is increased for $\alpha \neq 0$. However, this is at the price of removing the resulting invariant from the Noether class. As far as ease of manipulation is concerned, clearly the case $d \equiv 0$, $\alpha \equiv 0$ is the simplest. As to whether the non-Neotherian case $d \equiv 0, \alpha \neq 0$ is simpler than $d \neq 0$, $\alpha = 0$, it is not so easy to judge except when it comes to obtaining explicit expressions. The general case $d \neq 0$, $\alpha \neq 0$ is clearly the most complex. Because the case $\alpha \neq 0$ is non-Noetherian we would expect some qualitative differences in the invariant which will probably have direct bearing on possible quantum mechanical applications. For the moment we shall examine the general case and then discuss the particular cases in turn.

7. DEFINING EQUATION FOR THE INVARIANT

In Sec. 4 we suggested that the process of finding the invariant would be simplified by a transformation to a new space–time coordinate system. We define the transformation as

$$T = f(t), \quad Q = g(t)q + h(t). \tag{7.1}$$

By imposing the requirement that $G(q,t)$ as given in Eq. (6.15) take the form

$$G(Q,T) = \partial/\partial T, \tag{7.2}$$

we see that

$$a\dot{f} = 1, \quad a\dot{g} + \tfrac{1}{2}(\dot{a} + \alpha)g = 0, \quad a\dot{h} + dg = 0, \tag{7.3}$$

so that the parameters of the transformation are given by

$$f(t) = \int^t a^{-1}(t')\,dt', \tag{7.4}$$

$$g(t) = a^{-1/2}(t)\exp(-\tfrac{1}{2}\alpha f(t)), \tag{7.5}$$

$$h(t) = -\int^t d(t')\,a^{-3/2}(t')\exp(-\tfrac{1}{2}\alpha f(t'))\,dt'. \tag{7.6}$$

Applying the transformation to the Newtonian equation of motion (6.1), it now takes the time-independent form

$$Q'' + \alpha Q' + KQ^2 + MQ + N = 0, \tag{7.7}$$

where the constants M and N are given by

$$M = \tfrac{1}{2}\{a\ddot{a} - \tfrac{1}{2}(\dot{a} + \alpha)^2 + \alpha(\dot{a} + \alpha) + 2a^2 - 4Kh\}, \tag{7.8}$$

$$N = h(M + Kh) + g\{\tfrac{1}{2}(\dot{a} - \alpha)d - a\dot{d}\}, \tag{7.9}$$

and the prime represents differentiation with respect to T. The constancy of the expressions on the right-hand sides of

Eqs. (7.9) and (7.10) may be confirmed by direct differentiation and the use of the differential equations for the various parameters.

In the normal way of finding the invariant corresponding to an operator G, we would solve the Eqs. (4.6) and (4.7). However, as Eq. (7.8) is independent of T and the invariant is a function of Q and Q', we rewrite it as

$$Q'\frac{dQ'}{dQ}+\alpha Q'+KQ^2+MQ+N=0. \tag{7.10}$$

The invariant is then obtained by quadrature.

For the case $\alpha=0$, we immediately obtain the energy-like integral

$$\Phi(Q',Q)=\tfrac{1}{2}Q'^2+\tfrac{1}{3}KQ^3+\tfrac{1}{2}MQ^2+NQ. \tag{7.11}$$

This is a Noetherian invariant and so has a corresponding Hamiltonian form. We now examine the problem from the Hamiltonian viewpoint.

8. HAMILTONIAN VIEWPOINT

The Hamiltonian corresponding to the original Newtonian is

$$H=\tfrac{1}{2}p^2+\tfrac{1}{2}q^2+\tfrac{1}{3}Bq^3. \tag{8.1}$$

The transformation which reduces the Newtonian equation to a time-independent form involved both a change of time scale and a linear transformation of the coordinate. In the Hamiltonian context such a transformation is accomplished in two stages. The first involves a linear canonical transformation which removes the time dependence to a multiplicative factor. The second is a change of time scale so that the Hamiltonian is now the invariant. Thus under the transformations we expect the Hamiltonian to become

$$\bar{H}(Q,P,T)=\tfrac{1}{2}P^2+\tfrac{1}{3}KQ^3+\tfrac{1}{2}MQ^2+NQ, \tag{8.2}$$

which is the Hamiltonian version of Eq. (7.12). We now verify this result.

The form of the linear transformation may be inferred from Eqs. (7.1) and (7.3) (with $\alpha=0$) and the fact, implicit in Eq. (8.2), that $Q'=P$. Thus

$$Q=qg+h,\quad P=a\dot{g}q+agp-d\dot{g}, \tag{8.3}$$

where

$$g=a^{-1/2},\quad h=\int dg/a. \tag{8.4}$$

The type two generating function is

$$F_2(q,P,t)=gPq+\dot{a}q^2/4a+dq/a+hP. \tag{8.5}$$

Then

$$\begin{aligned}H'(Q,P,t)&=H(q,p,t)+\partial F_2(q,P,t)/\partial t\\&=a^{-1}\{\tfrac{1}{2}P^2+\tfrac{1}{3}KQ^3+\tfrac{1}{2}MQ^2+NQ\}.\end{aligned} \tag{8.6}$$

The change of time scale, $T=\int a^{-1}(t')\,dt'$, yields the Hamiltonian (8.2).

We emphasize that an invariant of the form given in Eq. (8.2) exists only in the case $\alpha=0$, i.e., for functions $B(t)$ in Eq. (8.1) given by

$$B(t)=Ka^{-5/2}, \tag{8.7}$$

where a is a solution of the nonlinear equation

$$\ddot{a}+4\dot{a}+2Ka^{-5/2}(D\sin t+E\cos t)=0. \tag{8.8}$$

In view of the form of Eq. (8.8) it is rather fortunate that the foregoing analysis may be performed without an explicit knowledg of $a(t)$. In terms of the original coordinates the invariant is

$$\begin{aligned}I(q,\dot{q},t)&=\tfrac{1}{2}a\dot{q}^2-\tfrac{1}{2}\dot{a}q\dot{q}+\tfrac{1}{3}Bq^3/a+(\tfrac{1}{2}a+\ddot{a}/a)q^2-d\dot{q}\\&\quad+\{gh(a\ddot{a}-\tfrac{1}{2}\dot{a}^2+2a^2-2Kh)+(a\dot{d}-\dot{a}d)/a\}q\\&\quad+d^2/2a+4Kh^3/3+3Mh^2/2+gh(\dot{a}d/2-a\dot{d}).\end{aligned} \tag{8.9}$$

For the simpler case when d is zero, this reduces to

$$I(q,\dot{q},t)=a\{\tfrac{1}{2}\dot{q}^2+\tfrac{1}{2}q^2+\tfrac{1}{3}Bq^3\}-\tfrac{1}{2}\dot{a}q\dot{q}+\ddot{a}q^2/a, \tag{8.10}$$

which is not overly complicated.

9. DISCUSSION

We have seen that an invariant may be obtained for certain time-dependent anharmonic systems. From the invariant we may obtain information regarding the boundedness of the motion of the particle. To illustrate this, we consider the simplest case $\alpha\equiv0$, $d\equiv0$. Then

$$a(t)=A+B\sin 2t+C\cos 2t, \tag{9.1}$$

$$B(t)=K(A+B\sin 2t+C\cos 2t)^{-5/2}, \tag{9.2}$$

$$M=A^2-B^2-C^2. \tag{9.3}$$

By requiring that $B(t)$ be finite and real we have

$$A>(B^2+C^2)^{1/2},\quad M>0.$$

In this particular instance, the transformation from (q,p) to (Q,P) is

$$Q=gq,\quad P=a\dot{g}q+agp, \tag{9.4}$$

so that, when $p=0$, the invariant is

$$I=\tfrac{1}{3}KQ^3+\tfrac{1}{2}(M+\tfrac{1}{4}\dot{a}^2)Q^2. \tag{9.5}$$

The values of I for which this cubic has three real distinct roots may be obtained from the discriminant of Cardan's formula.[26] We find that

$$I<(M+\tfrac{1}{4}\dot{a}^2)^3/6K^2 \tag{9.6}$$

and, since $\frac{1}{2}\dot{a}(=B\cos 2t-C\sin 2t)$ may be zero, the motion will be bounded for

$$I<M^3/6K^2. \tag{9.7}$$

A similar analysis may be performed for the case $d\not\equiv0$, $\alpha\equiv0$, but the result is more complicated due to the extra terms. When $\alpha\not\equiv0$, the stability of the motion is more difficult to determine because of the nature of the differential equation (7.11) determining the invariant. However, for particular cases it would be amenable to numerical treatment.

The value of the results obtained here depends upon the type of time variation of the field found in the experimental situation. If this time variation is one of the types allowed by the theory for an exact invariant, it would be most gratifying. If it is not, it may be possible to use the results obtained here as an approximation to find bounds within which the motion will remain.

ACKNOWLEDGMENTS

We wish to thank Ralph Lewis of Los Alamos for originally suggesting this problem, Willy Sarlet of Ghent for a preprint of a paper on invariants for certain nonlinear systems,[27] Geoffry Prince and Professor C.J. Eliezer of La Trobe for stimulating conversations on the Lie method.

[1]H.R. Lewis, Jr., Phys. Rev. Lett. **18**, 510 (1967); H.R. Lewis, Jr., Phys. Rev. **172**, 1313 (1968); H.R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[2]M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[3]C.J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).
[4]N.J. Günther and P.G.L. Leach, J. Math. Phys. **18**, 572 (1977).
[5]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[6]P.G.L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[7]P.G.L. Leach, J. Math. Phys. **18**, 1608 (1977); P.G.L. Leach, J. Math. Phys. **18**, 1902 (1977).
[8]H.R. Lewis Jr. and W.B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969)
[9]See Ref. 7b and also P.G.L. Leach, J. Math. Phys. **19**, 446 (1978).
[10]The clearest discussion is in K.B. Wolf, *Integral Transforms in Science and Engineering* (Plenum, New York, 1979), pp. 381ff. See also K.B. Wolf, J. Math. Phys. **17**, 601 (1976); M.H. Boon and T.H. Seligman, J. Math. Phys. **14**, 1224 (1973).
[11]H.R. Lewis, private communication, October 18, 1977.
[12]P.G.L. Leach, J. Math. Phys. **20**, 96 (1979).
[13]F.G. Gustavson, Astron. J. **71**, 670 (1966).
[14]A. Maharatna, R. Dutt, and D. Chattarji, J. Math. Phys. (to appear).
[15]P. Caldriola, Rend. R. Accad. Ital., serie VII, **11**, 896 (1941); E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[16]P.G.L. Leach, J. Math. Phys. **20**, 86 (1979).
[17]M. Lutzky, J. Phys. A **11**, 249 (1978).
[18]C.J. Eliezer, Hadronic J. (to appear); G.E. Prince and C.J. Eliezer, J. Phys. A (to appear); G.E. Prince (submitted to J. Phys. A).
[19]P.G.L. Leach, *Symmetries and Transformations for One-Dimensional Classical Dynamical Systems* (Research Report AM-79:05, Department of Applied Mathematics, La Trobe University).
[20]R.L. Anderson and S.M. Davison, J.Math. Anal. Appl. **48**, 301 (1974); C.E. Wulfman and B.J. Wybourne, J. Phys. A **9**, 507 (1976).
[21]P.G.L. Leach, J. Math. Phys. (to appear); P.G.L. Leach, Sl (3,R) *and the Repulsive oscillator* (Research Report, May 1979, Department of Applied Mathematics, La Trobe University). See also Refs. 18b and 20a.
[22]P.G.L. Leach, J. Aust. Math. Soc. B (to appear).
[23]Cf. P.A.M. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, England, 1930), p. 104.
[24]See Refs. 22 and 18a.
[25]For example the Emden equation for a selected value of the "energy" constant.
[26]A. Mostowski and M. Stark, *Introduction to Higher Algebra* (Pergamon, London, 1964), pp. 291–4.
[27]Willy Sarlet and Leon Y. Bahar, *A Direct Construction of First Integrals for Certain Non-Linear Dynamical Systems* (preprint, September 1979, Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 279-S9, B-9000 Gent, Belgium).

Scattering of a scalar wave from a slightly random surface

J. Nakayama, H. Ogura, and M. Sakata
Department of Electronics, Kyoto Institute of Technology, Matsugasaki, Kyoto 606, Japan

(Received 22 April 1980; accepted for publication 12 September 1980)

Scalar wave scattering from a slightly random surface is analyzed by a probabilistic method. We make use of the homogeneity of an infinite random surface, that is, the shift invariance property of the strictly homogeneous random field. By the group-theoretic consideration of such a shift invariance property, the wave solution proves to be a homogeneous random field multiplied by an exponential function. Then such a homogeneous random field is approximately solved for a slightly random surface to yield a wave solution involving multiple scattering. Several statistical properties of the scattering are calculated and shown in the figures. The accuracy of the approximate solution is examined in terms of the error of the boundary-value equation.

PACS numbers: 03.40.Kf

1. INTRODUCTION

This paper describes a new formulation of the random surface scattering that is a mathematical boundary-value problem of the wave equation (see Fig. 1). Our formulation is different from the small perturbation method,[1,2] the Kirchhoff approximation,[3] and the diagrammatic approach[4,5] in the multiple scattering theory but is analogous in the fundamental idea to the diffraction theory of periodic gratings.[6–8] For a plane wave incident on a periodic surface, the wave solution has the well-known Floquet form, which is a z dependent periodic function of $\mathbf{r}$, having the same period as the surface does, multiplied by an exponential phase factor. On the basis of the periodicity, such a z dependent function is commonly represented by a Fourier series with Fourier coefficients, so that the problem is reduced to finding such Fourier coefficients by solving the boundary condition on the surface. Starting with the periodicity, one can easily obtain the optical theorem (the conservation law of power flux) also.

The formulation proposed here is a stochastic version of such formulation for a periodic case, where we make use of the stochastic homogeneity and the ergodic property instead of the periodicity. In the next section, we assume that the infinite surface is described by a strictly homogeneous random field of $\mathbf{r}$. By use of the group-theoretic consideration of the shift invariance property concerning such a random field,[9,10] we then show that, for a plane wave incidence, the stochastic wave solution can be written by a z dependent homogeneous random field of $\mathbf{r}$ multiplied by a exponential phase factor. Such a form of solution is a stochastic analogue to the Floquet form for a periodic surface scattering and is generally applicable to the scattering from a homogeneous random surface.

Since the exponential phase factor is uniquely given by the phase factor of the incident plane wave, the problem is reduced to finding such a homogeneous random field as a functional of the random surface. Unlike the periodic surface case, the Fourier series does not work well for representing such a homogeneous random field. Therefore we employ Wiener's nonlinear theory[11–14] of the Brownian motion process in the probability theory in Sec. 3; formulas concerning the nonlinear theory are briefly described in the Appendix. Unlike the cluster expansion[5] for characteristic functions that are deterministic, the nonlinear theory makes it possible to represent a stochastic functional in terms of orthogonal functionals that are random. Assuming that the surface is a Gaussian random field generated by the complex Gaussian random measure, which is defined in the Appendix, we expand the z dependent homogeneous random field in terms of the orthogonal functionals associated with the complex Gaussian random measure. Then the boundary value problem is transformed into a set of equations for deterministic functions that are coefficients of such functional expansion. The equations are solved with a good approximation for a slightly random surface to yield a stochastic wave solution involving the effect of multiple scatterings. The accuracy of the approximate solution is then examined in terms of the mean-square error with respect to the boundary condition.

Since the wave solution is represented as a stochastic functional, any statistical quantities can be obtained by taking averages of desired quantities. Moreover, we can apply the ergodic theorem to several quantities of the wave field, because the solution is written in terms of a homogeneous random field. Applying the ergodic theorem to the power flux of the wave field, we systematically obtain the optical theorem in Sec. 4, which gives the power relation between the coherent scattering and the incoherent scattering. We calculate several quantities of the scattering such as the complex amplitude of the coherent wave, the angular distribution of the incoherent scattering, the powerflow of the surface wave, and the optical theorem, which are shown in the figures.

2. FORMULATION OF THE PROBLEM

Consider scalar wave scattering from an infinite random surface as shown in Fig. 1. Let us denote by $\mathbf{r} = (x,y)$ a vector in the two-dimensional plane $R^2 = \{-\infty < x,y < \infty\}$ and by ω a probability parameter describing a sample point in the sample space Ω_p. We assume that the random surface is described by a real, strictly homogeneous and isotropic random field in the form[9–12]

$$z = f(T^{\mathbf{r}}\omega), \tag{1}$$

where $T^{\mathbf{r}}$ is a measure-preserving transformation in Ω_p tak-

 022-2488/81/030471-07$1.00

ing ω into $T^{\mathbf{r}}\omega$ which enjoys the one-parameter group property;

$$T^{\mathbf{0}} = 1 \text{ (identity)}, \quad T^{\mathbf{a}}T^{\mathbf{b}} = T^{\mathbf{a}+\mathbf{b}}, \tag{2}$$

a and **b** being vectors in R^2. We further assume $T^{\mathbf{r}}$ to be metrically transitive[11-14] and that

$$\langle f(T^{\mathbf{r}}\omega)\rangle = 0, \quad \langle f(T^{\mathbf{r}}\omega)^2\rangle = \sigma^2, \tag{3}$$

where $\langle\ \rangle$ denotes the averaging over the sample space, and σ^2 the intensity of fluctuation.

The wave field $\psi(z,\mathbf{r},\omega)$ satisfies the three-dimensional wave equation

$$[\nabla^2 + k^2]\psi(z,\mathbf{r},\omega) = 0, \tag{4}$$

in the free space $z > f(T^{\mathbf{r}}\omega)$ and fulfills the Dirichlet boundary condition

$$\psi(z,\mathbf{r},\omega) = 0 \tag{5}$$

on the surface (1). Moreover, we assume that the wave field satisfies the radiation condition. For sufficiently small σ^2, (5) is well approximated by the effective boundary condition[2,4]

$$\psi(0,\mathbf{r},\omega) + f(T^{\mathbf{r}}\omega)\frac{\partial}{\partial z}\psi(0,\mathbf{r},\omega) = 0. \tag{6}$$

We will use (6) as an appropriate model of the random boundary condition in what follows; however, our formulation can be applied to a more general boundary condition without any difficulty.

To look for a form of the wave solution, we introduce a translation operator D under with $f(T^{\mathbf{r}}\omega)$ is invariant.[9,10] We define the operator D, which acts on a functional $\psi(z,\mathbf{r},\omega)$ of $f(T^{\mathbf{r}}\omega)$, by the relations

$$D^{\mathbf{a}}\psi(z,\mathbf{r},\omega) = \psi(z,\mathbf{r}+\mathbf{a},T^{-\mathbf{a}}\omega); \tag{7}$$

$$D^{\mathbf{0}} = 1, \quad D^{\mathbf{a}}D^{\mathbf{b}} = D^{\mathbf{a}+\mathbf{b}}. \tag{8}$$

($D^{\mathbf{a}}$ translates the variables **r** and ω but does not affect the variable z.)

If ψ is a solution satisfying (4) and (6), and the radiation condition, $D^{\mathbf{a}}\psi$ is also a solution satisfying the same conditions, because $D^{\mathbf{a}}$ commutes with $\nabla^2, f(T^{\mathbf{r}}\omega)$ and the radiation condition. Since the wave solution is to be uniquely determined except for a constant factor, ψ and $D^{\mathbf{a}}\psi$ are in a linear relation

$$D^{\mathbf{a}}\psi(z,\mathbf{r},\omega) = C(\mathbf{a},\omega)\psi(z,\mathbf{r},\omega). \tag{9}$$

Hence, we regard $\psi(z,\mathbf{r},\omega)$ as the eigenfunction of $D^{\mathbf{a}}$ and $C(\mathbf{a},\omega)$ as the eigenvalue[9,10] If we assume that ψ has the same translation property under $D^{\mathbf{a}}$ as the incident plane wave

$$\psi_i(z,\mathbf{r}) = -e^{i(\mathbf{K}_0\mathbf{r} - K_z z)}; \quad K_z = (k^2 - \mathbf{K}_0^2)^{1/2} > 0, \tag{10}$$

does, then we can easily verify that the wave solution should have the form

$$\psi(z,\mathbf{r},\omega) = e^{i\mathbf{K}_0\mathbf{r}}[-e^{-iK_z z} + e^{iK_z z} + U(z,T^{\mathbf{r}}\omega)]. \tag{11}$$

Here, the first factor in the right-hand side is the eigenvalue of $D^{\mathbf{r}}$, physically representing the phase factor, and the second factor is a z-dependent homogeneous random field of **r**, which is invariant under $D^{\mathbf{a}}$. The first term in the brackets indicates the incident plane wave having the unit amplitude, the second term the specularly reflected wave for completely smooth surface, and the third the scattered wave due to surface roughness. Note that (11) is analogous to the Floquet solution for the periodic surface scattering,[6-8] where the corresponding U is a z-dependent periodic function of **r** having the same period as the surface has.

We have seen that the problem is reduced to finding the z-dependent homogeneous random field U. In the next section, we will solve for such a random field by use of Wiener's nonlinear theory of the Brownian-motion process when the random surface is a Gaussian random field.

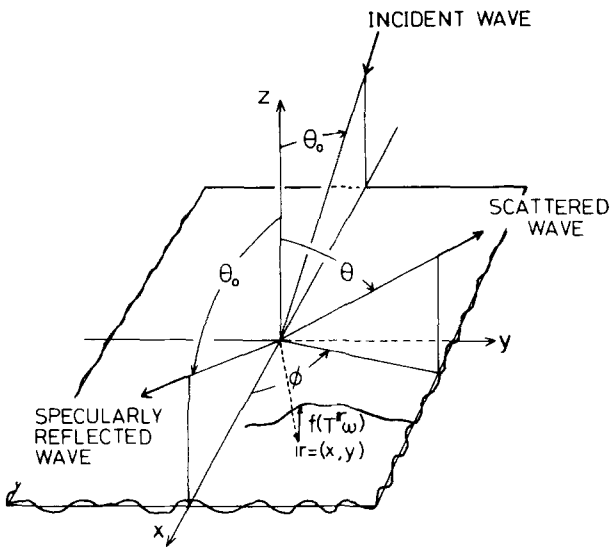


FIG. 1. Scattering of a scalar plane-wave from a random rough surface described by $z = f(T^{\mathbf{r}}\omega)$. θ_0: the angle of incidence, (θ,ϕ):a scattering angle.

3. APPROXIMATE METHOD FOR A GAUSSIAN SURFACE

A. Representation of the wave field

For concrete analysis we assume that $f(T^{\mathbf{r}}\omega)$ is a Gaussian random field generated by the complex Gaussian random measure (A1) and is represented as the Wiener integral (see A5)

$$f(T^{\mathbf{r}}\omega) = \int_{R^2} e^{i\boldsymbol{\lambda}\mathbf{r}} F(\boldsymbol{\lambda})\, dB(\boldsymbol{\lambda},\omega). \tag{12}$$

(Hereafter, we often drop ω for simplification.) In order to make $f(T^{\mathbf{r}}\omega)$ real and isotropic, we impose the conditions

$$F^*(\boldsymbol{\lambda}) = F(-\boldsymbol{\lambda}),$$
$$|F(\boldsymbol{\lambda})|^2 = |F(\Lambda)|^2; \quad \Lambda = |\boldsymbol{\lambda}| = (\lambda_x^2 + \lambda_y^2)^{1/2}. \tag{13}$$

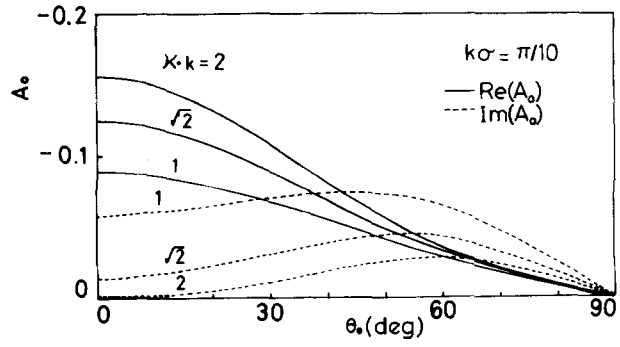


FIG. 2. A_0 vs. θ_0. [The amplitude of the coherent wave is $(1 + A_0)$ by (16).]

Then the correlation function (A6) of $f(T^r\omega)$ is reduced to

$$R_f(|\mathbf{a}|) = 2\pi \int_0^\infty J_0(\Lambda\,|\mathbf{a}|)|\,F(\Lambda\,)|^2 \Lambda\, d\Lambda,$$

where J_0 is the Bessel function of the first kind.

A solution U to (6) is a functional of $f(T^r\omega)$; however, it is regarded as a functional of $dB(\lambda)$ by (12), and hence it can be developed for any fixed z into the orthogonal functionals associated with $dB(\lambda)$ as (A13). In order to make the scattered wave $e^{i\mathbf{K}_0\mathbf{r}}U(z,T^r\omega)$ satisfy the wave equation (4), we rewrite (A13) as

$$\begin{aligned} U(z,T^r\omega) = A_0 e^{iK_z z} &+ \int_{R^2} A_1(\lambda) e^{i\lambda r} e^{ik_z(\lambda)z}\, dB(\lambda) \\ &+ \int_{R^2\times R^2} A_2(\lambda_1,\lambda_2) e^{i(\lambda_1+\lambda_2)r}\, e^{ik_z(\lambda_1+\lambda_2)z} \\ &\times h^{(2)}[dB(\lambda_1), dB(\lambda_2)] + \cdots, \end{aligned} \tag{14}$$

where $h^{(n)}$'s are the Wiener–Hermite differentials defined in the Appendix, and A_0 and A_n $(n\geqslant 1)$ are a constant and deterministic functions symmetric with respect to their arguments, respectively, and $k_z(\lambda)$ is a positive real or positive imaginary function given by

$$\begin{aligned} k_z(\lambda) &= (k^2 - (\mathbf{K}_0+\lambda)^2)^{1/2} > 0; \quad k^2 > (\mathbf{K}_0+\lambda)^2 \\ &= i((\mathbf{K}_0+\lambda)^2 - k^2)^{1/2}; \quad k^2 < (\mathbf{K}_0+\lambda)^2. \end{aligned} \tag{15}$$

Substituting (14) into (11) completes the stochastic representation of the wave field. By (A9) the coherent wave field becomes

$$\langle\psi(z,\mathbf{r},\omega)\rangle = e^{i\mathbf{K}_0\mathbf{r}}[-e^{-iK_z z} + (1+A_0)e^{iK_z z}]. \tag{16}$$

Clearly, $-(1+A_0)$ is the reflection coefficient for the coherent scattering in terms of which we will define the equivalent surface impedance Z_s below [see Eq. (22)].

B. An approximate solution

Once the A_n's are obtained, we find U by (14) and in turn $\psi(z,\mathbf{r},\omega)$ by (11), so that the problem is to determine a set of functions A_n's now.

In view of approximate analysis, we define the boundary-value error by the relation

$$e_b(\mathbf{r},\omega) = e^{-i\mathbf{K}_0\mathbf{r}}\left[\psi(0,\mathbf{r},\omega) + f(T^r\omega)\frac{\partial}{\partial z}\psi(0,\mathbf{r},\omega)\right]. \tag{17}$$

Taking in mind that $e_b(\mathbf{r},\omega)$ is a homogeneous random field, we look for that equations for A_n's which should yield $\langle|e_b|^2\rangle = 0$. Using (11), (14), (A8), and (A9), we find

$$\begin{aligned} \langle|e_b(\mathbf{r},\omega)|^2\rangle = &|A_0 + i\int_{R^2} k_z(\lambda)A_1(\lambda)F^*(\lambda)\,d\lambda|^2 \\ &+ \int_{R^2} |A_1(\lambda) + iK_z[2+A_0]F(\lambda) \\ &+ 2i\int_R k_z(\lambda+\lambda_1)A_2(\lambda,\lambda_1)F^*(\lambda_1)\,d\lambda_1|^2\,d\lambda \\ &+ 2!\int_{R^2\times R^2} |A_2(\lambda_1,\lambda_2) + (i/2)k_z(\lambda_1)A_1(\lambda_1)F(\lambda_2) \\ &+ (i/2)k_z(\lambda_2)A_1(\lambda_2)F(\lambda_1) \\ &+ 3i\int_{R^2} k_z(\lambda_1+\lambda_2+\lambda_3)A_3(\lambda_1,\lambda_2,\lambda_3) \\ &\times F^*(\lambda_3)\,d\lambda_3|^2\,d\lambda_1\,d\lambda_2 + \cdots. \end{aligned} \tag{18}$$

Here, each term in the right-hand side is non-negative. Thus putting $\langle|e_b|^2\rangle = 0$ yields the hierarchy of equations for A_n's:

$$A_0 + i\int_{R_2} k_z(\lambda)A_1(\lambda)F^*(\lambda)\,d\lambda = 0, \tag{19}$$

$$A_1(\lambda) + ik_z[2+A_0]F(\lambda) + 2i\int_{R^2} k_z(\lambda+\lambda_1)A_2(\lambda,\lambda_1)\,d\lambda_1 = 0, \tag{20}$$

$$\begin{aligned} A_2(\lambda_1,\lambda_2) &+ (i/2)k_z(\lambda_1)A_1(\lambda_1)F(\lambda_2) + (i/2)k_z(\lambda_2)A_1(\lambda_2)F(\lambda_1) \\ &+ 3i\int k_z(\lambda_1+\lambda_2+\lambda_3)A_3(\lambda_1,\lambda_2,\lambda_3)F^*(\lambda_3)\,d\lambda_3 = 0 \end{aligned}$$

etc.

Since $F(\lambda)$ is a function of the order of σ^1 by (3), these equations mean that A_0 and $A_n(n\geqslant 1)$ are at most of the order of σ^2 and σ^n, respectively. Therefore we can solve these equations by neglecting higher order functions for sufficiently small σ^2. Let us obtain the first order solution involving only A_0 and A_1. If we put $A_n = 0$ for $n\geqslant 2$, Eq. (20) gives the solution A_1

$$A_1(\lambda) = -iK_z[2+A_0]F(\lambda) \tag{21}$$

Inserting this into (19) yields

$$A_0 = -2Z_s/[1+Z_s] \tag{22}$$

where Z_s denotes the equivalent surface impedance given by

$$Z_s = K_z\int_{R^2} k_z(\lambda)|F(\lambda)|^2\,d\lambda. \tag{23}$$

Here, the solution (22) for the coherent wave (16) is similar to the results by other methods.[2,4]

$\mathrm{Re}(Z_s)$ represents energy dissipation of the coherent scattering due to the incoherent scattering, and $\mathrm{Im}(Z_s)$ suggest the existence of the surface wave. Since the real and imaginary parts of Z_s are non-negative and proportional to $K_z = k\cos\theta_0$, θ_0 being the angle of incidence (see Fig. 1), $\mathrm{Re}(A_0)$ and $\mathrm{Im}(A_0)$ are negative and tend to zero as $\theta_0\to\pi/2$. Figure 2 shows A_0 as a function of θ_0, where we have assumed the Gaussian roughness spectrum

$$|F(\lambda)|^2 = (\sigma^2\kappa^2/\pi)e^{-\kappa^2\lambda^2} \tag{24}$$

with κ the correlation distance of the random surface. Clearly, a negative $\mathrm{Re}(A_0)$ reduces the amplitude of the coherent

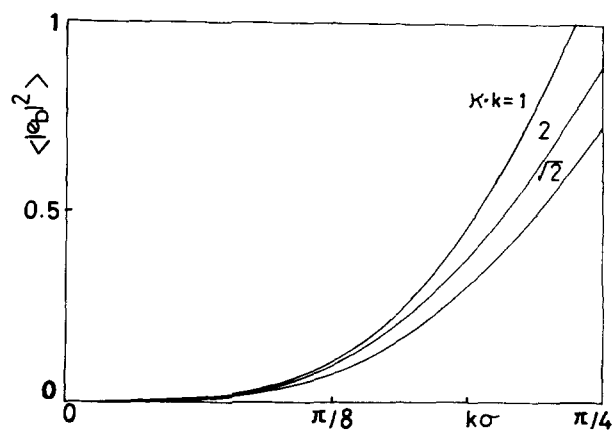


FIG. 3. Mean square value of the boundary-value error. Horizontal: the roughness parameter $k\sigma$.

scattering in (16) and also gives a feedback effect on A_1 in (21). This is due to multiple scattering.

In order to check the accuracy of the first order solution, let us calculate the mean square value of the boundary-value error; we will discuss the accuracy by another method in Sec. 4.A. Inserting (21) and (22) with $A_n = 0 (n \geqslant 2)$ into (18), we obtain

$$\langle |e_b|^2 \rangle = K_z^2 |2 + A_0|^2 \left(\sigma^2 \int_{R^2} |k_z(\lambda) F(\lambda)|^2 \, d\lambda \right) + \left| \int_{R^2} k_z(\lambda) |F(\lambda)|^2 \, d\lambda \right|^2, \tag{25}$$

which depends on K_z, σ^2, and the roughness spectrum. Figure 3 plots $\langle |e_b|^2 \rangle$ versus $k\sigma$ for normal incidence with $\theta_0 = 0$. The error is relatively small for $k\sigma < \pi/8$ but rapidly increases for $k\sigma > \pi/8$ as $k\sigma$ increasing. Thus the first order solution gives a good approximation for a slightly rough case. For $k\sigma > \pi/8$, higher order functions, such as A_2 and A_3, will yield a resonable solution for the effective boundary condition (6).

4. PROPERTIES OF THE RANDOM SURFACE SCATTERING

A. The optical theorem

From (4) we easily find the identity

$$\operatorname{div}[\operatorname{Im}(\psi^* \operatorname{grad} \psi / k)] = 0 \tag{26}$$

Integrating this over the columnar volume as shown in Fig. 4, then applying the divergence theorem, we obtain

$$\frac{1}{k} \int_{\pi l^2} \operatorname{Im}\left(\psi^* \frac{\partial}{\partial z} \psi|_{z=z_0}\right) d\mathbf{r} + \frac{1}{k} \int_{S_c} \operatorname{Im}\left(\psi^* \frac{\partial}{\partial \mathbf{n}} \psi\right) dS_c = 0.$$

Here S_c shows the side of the column, $\mathbf{n}$ an outward normal to S_c, and z_0 and l are the height and the radius of the column, respectively. The first integral and the second one are proportional to l^2 and l^1, respectively. Therefore, we obtain

$$\lim_{l \to \infty} \frac{1}{\pi l^2 k} \int_{\pi l^2} \operatorname{Im}\left(\psi^* \frac{\partial}{\partial z} \psi|_{z=z_0}\right) d\mathbf{r} = 0 \tag{27}$$

This means a vanishing power flux in the z direction per unit surface area, because the incident power is completely reflected by the surface. By (11) the integrand is a homogeneous random field and furthermore it is ergodic because $T^{\mathbf{r}}$ is a metric transitive transformation.[9,11] Hence, the space average can be replaced by the probabilistic average by virtue of the ergodic theorem. Consequently, we obtain for almost all ω

$$\frac{1}{k} \langle \operatorname{Im}\left(\psi^* \frac{\partial}{\partial z} \psi|_{z=z_0}\right) \rangle = 0, \tag{28}$$

which is the conservation law of power flux, generally applicable to the random surface scattering.

Substituting (11) and (14) into (28), we obtain the optical theorem in terms of A_n's

$$\frac{K_z}{k} = \frac{K_z}{k} |1 + A_0|^2 + \frac{1}{k} \sum_{n=1}^{\infty} n! \int_{(\mathbf{K}_0 + \lambda_1 + \lambda_2 + \cdots + \lambda_n)^2 < k^2} \cdots \int k_z(\lambda_1 + \lambda_2 + \cdots + \lambda_n) |A_n(\lambda_1, \lambda_2, \cdots, \lambda_n)|^2 \, d\lambda_1 \, d\lambda_2 \cdots d\lambda_n. \tag{29}$$

The left number is the incident power falling on a unit area, whereas the first term in the right-hand side is the coherently reflected power, and the series expansion is the incoherently scattered power from a unit area. This optical theorem gives another measure for estimating the accuracy of an approximate solution, because it is an exact relation. Note that the optical theorem explicitly relates to propagating waves, where surface waves affect it implicitly. On the other hand, the boundary-value error in Sec. 3.B directly relates to both of the waves.

Figure 5 shows the optical theorem for the first order solution. When $k\sigma$ increases, the coherent scattering (curve b) decreases and the incoherent scattering (curve c) increases; total power scattered (line a) remains constant equal to the incident power, because the first order solution satisfies the optical theorem.

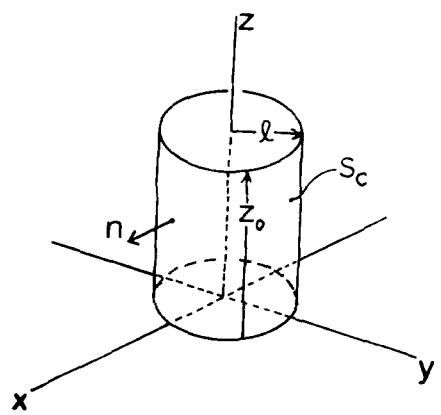


FIG. 4. Columnar volume for applying the divergence theorem. S_c the side of the column, $\mathbf{n}$ an outward normal to S_c, z_0 and l the height and the radius of the column, respectively.

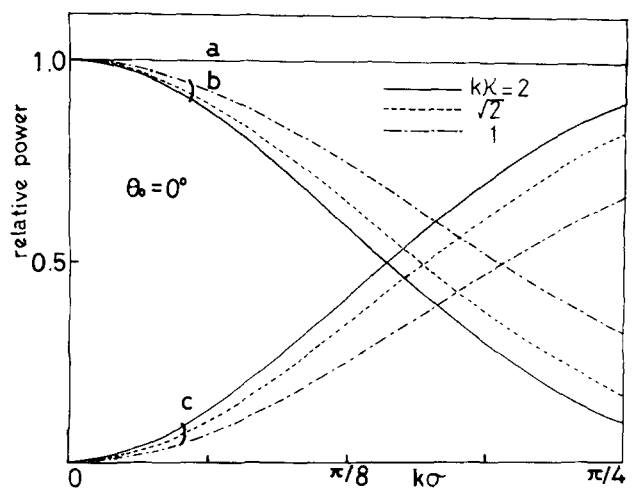


FIG. 5. Relative power flux vs $k\sigma$. a: incident flux and total scattered flux b + c; b: coherent scattering; c: incoherent scattering.

B. Angle distribution of the incoherent scattering

The incident plane wave is scattered into various directions. By $S(\Omega|\Omega_0)$ we denote the angle distribution of the incoherent scattering, that is, the average power scattered incoherently from unit surface area into unit solid angle of direction $\Omega = (\theta,\phi)$ when the angle of incidence is $\Omega_0 = (\theta_0,\phi_0)$ (see Fig. 1.) Since (29) describes power flow in the z direction per unit area, the series expansion in (29) equals the integral of $S(\Omega|\Omega_0)$ times $\cos\theta$ over 2π steradians

$$\int_{2\pi} S(\Omega|\Omega_0)\cos\theta d\,\Omega = \frac{1}{k}\sum_{n=1}^{\infty} n! \int\int_{(\mathbf{K}_0+\lambda_1+\cdots+\lambda_n)^2<k^2} \cdots \int k_z(\lambda_1+\lambda_2+\cdots+\lambda_n)|A_n(\lambda_1,\lambda_2,\cdots,\lambda_n)|^2\, d\,\lambda_1 d\,\lambda_2 \cdots d\,\lambda_n .$$

Putting $\mathbf{K} = \mathbf{K}_0 + \lambda_1 + \lambda_2 + \cdots + \lambda_n = (k\sin\theta\cos\phi, k\sin\theta\sin\phi)$, $\mathbf{K}_0 = (k\sin\theta_0\cos\phi_0, k\sin\theta_0\sin\phi_0)$ and hence $k_z(\lambda_1+\lambda_2+\cdots+\lambda_n) = k\cos\theta$ by (15). (The azimuth angle of incidence is measured from the $-x$ axis, hence Fig. 1 shows $\phi_0 = 0$.) We obtain

$$S(\Omega|\Omega_0) = k^2\cos\theta\{|A_1(\mathbf{K}-\mathbf{K}_0)|^2 + 2!\int_{R^2}|A_2(\mathbf{K}-\mathbf{K}_0-\lambda_2,\lambda_2)|^2\, d\,\lambda_2 + \cdots\}, \tag{30}$$

where $\mathbf{K}-\mathbf{K}_0$ is a Bragg vector of the scattering. Inserting (21) into this yields

$$S(\Omega|\Omega_0) \simeq k^4\cos\theta\cos^2\theta_0|2+A_0|^2|F(\Lambda)|^2;$$
$$\Lambda = |\mathbf{K}-\mathbf{K}_0| = k\,[\sin^2\theta+\sin^2\theta_0-2\sin\theta\sin\theta_0\cos(\phi-\phi_0)]^{1/2}, \tag{31}$$

which depends on the difference $(\phi-\phi_0)$ as expected from the isotropy of the random surface. Figure 6 shows the angle distribution of the incoherent scattering for the Gaussian roughness spectrum (24).

C. Power flow of the surface wave

There is some power flow of the surface wave which is written by the integrals in (14) over regions of λ's such that

$$(\mathbf{K}_0+\lambda_1+\cdots+\lambda_n)^2 > k^2.$$

In terms of the first order solution we write the surface wave $\Psi_s(z,\mathbf{r},\omega)$ as

$$\Psi_s(z,\mathbf{r},\omega) \simeq e^{i\mathbf{K}_0\mathbf{r}}\int_{(\mathbf{K}_0+\lambda)^2>k^2} A_1(\lambda)e^{i\lambda\mathbf{r}}e^{-[(\mathbf{K}_0+\lambda)^2-k^2]^{1/2}z}\,dB(\lambda). \tag{32}$$

In view of the isotropy of the random suface there is no averaged power in the direction perpendicular to $\mathbf{K}_0$, and the average power flux P_{sf} in the $\mathbf{K}_0$ direction becomes

$$P_{sf}(z) = (1/k\,|\mathbf{K}_0|)\langle \mathrm{Im}[\Psi_s^*(z,\mathbf{r},\omega)\mathbf{K}_0\mathrm{grad}\Psi_s(z,\mathbf{r},\omega)]\rangle \simeq \int_{(\mathbf{K}_0+\lambda)^2>k^2} \frac{(\mathbf{K}_0+\lambda)\mathbf{K}_0}{k\,|\mathbf{K}_0|}|A_1(\lambda)|^2 e^{-2[(\mathbf{K}_0+\lambda)^2-k^2]^{1/2}z}\, d\,\lambda. \tag{33}$$

Since this depends on z, it is convenient to define the power flow of the surface wave P_s by integrating $P_{sf}(z)$ over a plane which has unit width and an infinite height as shown in Fig. 7. Then we obtain

$$P_s \simeq \int_{(\mathbf{K}_0+\lambda)^2>k^2} \frac{(\mathbf{K}_0+\lambda)\mathbf{K}_0}{2k\,|\mathbf{K}_0|}\frac{|A_1(\lambda)|^2}{[(\mathbf{K}_0+\lambda)^2-k^2]^{1/2}}\, d\,\lambda. \tag{34}$$

Figure 8 shows P_s versus the angle of incidence. P_s vanishes at normal incidence because of symmetry and tends to zero as $\theta_0 \to \pi/2$ because the incident power falling on unit area vanishes.

D. Correlation function of the incoherently scattered field

By R_{is} we denote the correlation function of the incoherently scattered field in a plane of z = const. Using the first order solution, we obtain

$$R_{is}(\mathbf{a}) \simeq e^{i\mathbf{K}_0\mathbf{a}}\int_{R^2}|A_1(\lambda)|^2 e^{i\lambda\mathbf{a}}e^{-2\mathrm{Im}[k_z(\lambda)]z}\, d\,\lambda, \tag{35}$$

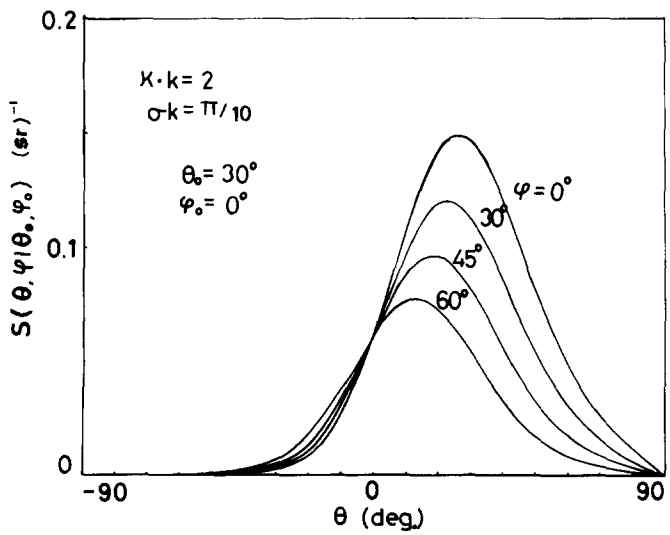


FIG. 6. Angle distribution of the incoherent scattering. (θ,ϕ): a scattering angle.

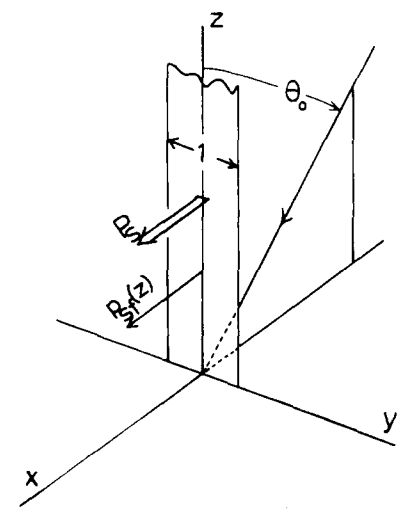


FIG. 7. Definition of power flow of the surface wave.

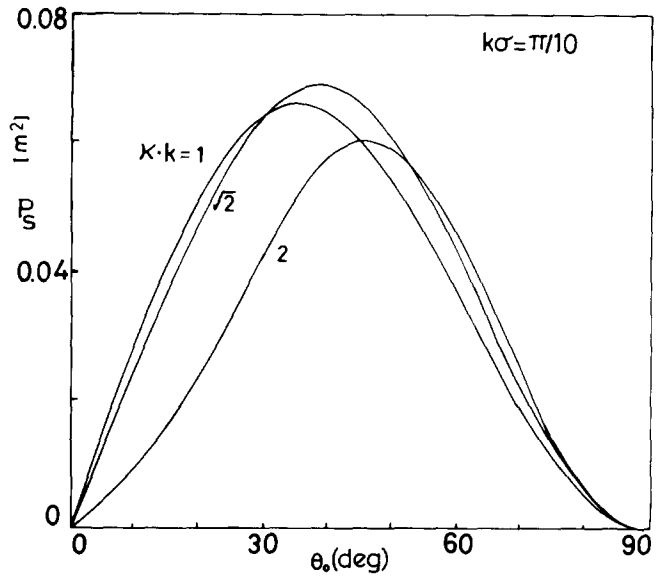


FIG. 8. Averaged power flow of the surface wave vs θ_0. Vertical: relative value normalized by the incident flux.

which for a large value of z becomes

$$R_{is}(\mathbf{a}) = e^{i\mathbf{K}_0\mathbf{a}} K_z^2 |2 + A_0|^2 \int_{(\mathbf{K}_0+\lambda)^2 < k^2} |F(\lambda)|^2 e^{i\lambda\alpha}\, d\lambda. \tag{36}$$

This is a band limited function. Thus the high frequency part of $|F(\lambda)|^2$ cannot be obtained from R_{is}.

5. CONCLUSIONS

We have proposed a new formulation of the random surface scattering, where we have made use of several concepts closely related with the stochastic homogeneity, such as the measure-preserving transformation in the sample space, the translation operator D, the nonlinear theory of the Brownian motion process, and the ergodic theorem. We find that the stochastic wave solution can be written in terms of a z dependent homogeneous random field and hence the problem is reduced to finding such a homogeneous random field as a functional of the random surface. This fact is to be understood as the starting point of the random surface scattering.

Assuming the effective boundary condition for a slightly rough case, we concretely obtained such a homogeneous random field and then calculated several statistical properties of the scattering.

Our formulation is essentially based on the homogeneity of an infinite random surface and hence it is applicable only to the scattering from such a homogeneous random surface. Provided the surface is described by a homogeneous random field, however, it can be extended to a more general boundary condition for electromagenetic waves without any difficulty.

APPENDIX

This Appendix summarizes formulas concerning Wiener's nonlinear theory of the Brownian-motion process. Concerning notations and definitions we follow the Appendix in Ref. 9. For detailed mathematical description, see Ogura,[9] Itô,[13,14] and Wiener.[11,12]

A. Complex Gaussian random measure

Let us denote by $d\lambda$ a rectangle at $\lambda = (\lambda_x, \lambda_y)$ having an infinitesimal area $d\lambda_x\, d\lambda_y$ in the two-dimensional plane $R^2 = (-\infty < \lambda_x, \lambda_y < \infty)$ and by ω a probability parameter describing a sample point in the sample space Ω_p. We introduce the complex Gaussian random measure $B(d\lambda, \omega) = dB(\lambda, \omega)$ on R^2 which satisfies the conditions:

(1) the real and imaginary parts of $dB(\lambda, \omega)$ have an identical independent Gaussian distributions with

$$\langle dB(\lambda, \omega)\rangle = 0, \quad \langle dB(\lambda_1, \omega) dB(\lambda_2, \omega)\rangle = \delta(\lambda_1 + \lambda_2)\, d\lambda_1\, d\lambda_2, \tag{A1}$$

where $\langle\ \rangle$ denotes the averaging over the sample space and the symbol $\delta(\lambda_1 + \lambda_2)\, d\lambda_1\, d\lambda_2$ stands for the area $d\lambda_1$ if $\lambda_1 = -\lambda_2$ or zero if $\lambda_1 \neq -\lambda_2$;

(2) for almost all ω, $dB(\lambda, \omega)$ satisfies

$$dB^*(\lambda, \omega) = dB(-\lambda, \omega), \tag{A2}$$

where the asterisk denotes the complex conjugate.

We define a shift of a sample field $dB(\lambda, \omega)$ by the relation

$$dB(\lambda, \omega) \to \exp(i\lambda\mathbf{a})\, dB(\lambda, \omega), \tag{A3}$$

where $\mathbf{a}$ is a vector in R^2. As easily verified, the right-hand side is a complex Gaussian random measure satisfying the above conditions, so that the shift (A3) generates a measure-preserving transformation $T^{\mathbf{a}}$ in Ω_p such that

$$\exp(i\lambda\mathbf{a})\, dB(\lambda, \omega) = dB(\lambda, T^{\mathbf{a}}\omega). \tag{A4}$$

Here $T^{\mathbf{a}}$ takes ω into $T^{\mathbf{a}}\omega$ with one-parameter group property: $T^0 = 1$(identity); $T^{\mathbf{a}}T^{\mathbf{b}} = T^{\mathbf{a}+\mathbf{b}}$.

If $F(\lambda)$ is a square-integrable function on R^2, then we can define a homogeneous Gaussian random field f by the Wiener integral

$$f(T^{\mathbf{r}}\omega) = \int_{R^2} e^{i\lambda\mathbf{r}} F(\lambda)\, dB(\lambda, \omega). \tag{A5}$$

By (A1) this has zero average and the correlation function

$$R_f(\mathbf{a}) = \langle f(T^{\mathbf{r}+\mathbf{a}}\omega) f^*(T^{\mathbf{r}}\omega)\rangle = \int_{R^2} e^{i\lambda\mathbf{a}} |F(\lambda)|^2 d\lambda. \tag{A6}$$

For simplification we often drop the parameter ω in the following.

B. Wiener–Hermite differentials

We define Wiener–Hermite differentials $h^{(n)}[dB(\lambda_1), dB(\lambda_2), \cdots, dB(\lambda_n)]$, $n = 0, 1, 2, \ldots$, associated with $dB(\lambda)$ by the relations

$$h^{(0)} = 1, \quad h^{(1)}[dB(\lambda)] = dB(\lambda), \quad h^{(2)}[dB(\lambda_1), \quad dB(\lambda_2)] = dB(\lambda_1)\,dB(\lambda_2) - \delta(\lambda_1 + \lambda_2)\,d\lambda_1\,d\lambda_2, \tag{A7}$$

etc. The nth degree Wiener–Hermite differential can be obtained by the recurrence formula

$$dB(\lambda_1)h^{(n-1)}[dB(\lambda_2), dB(\lambda_3), \cdots, dB(\lambda_n)]$$

$$= h^{(n)}[dB(\lambda_1), dB(\lambda_2), \cdots, dB(\lambda_n)] + \sum_{j=2}^{n} h^{(n-2)}[dB(\lambda_2), \cdots, dB(\lambda_{j-1}), db(\lambda_{j+1}), \cdots, dB(\lambda_n)]\delta(\lambda_1 + \lambda_j)\,d\lambda_1\,d\lambda_j. \tag{A8}$$

By these definitions the Wiener–Hermite differential satisfies the orthogonality relation

$$\langle h^{(n)}[dB(\lambda_{i_1}), dB(\lambda_{i_2}), \cdots, dB(\lambda_{i_n})]h^{(m)*}[dB(\lambda_{j_1}), dB(\lambda_{j_2}), \cdots, dB(\lambda_{j_m})]\rangle$$

$$= \delta_{nm}\delta^{n}_{ij}\,d\lambda_{i_1}d\lambda_{i_2}\cdots d\lambda_{i_n}\,d\lambda_{j_1}\,d\lambda_{j_2}\cdots d\lambda_{j_m}, \tag{A9}$$

where δ^{n}_{ij} equals the sum of all distinct products of n delta functions of the form $\delta(\lambda_{i_\nu} - \lambda_{j_\mu})$, $i = (i_1, i_2, \cdots, i_n), j = (j_1, j_2, \cdots, j_m)$, all i_ν and j_μ appearing just once in each product, for example

$$\delta^{2}_{ij} = \delta(\lambda_{i_1} - \lambda_{j_1})\delta(\lambda_{i_2} - \lambda_{j_2}) + \delta(\lambda_{i_2} - \lambda_{j_1})\delta(\lambda_{i_1} - \lambda_{j_2}).$$

By (A4) the Wiener–Hermite differential is translated as

$$h^{(n)}[dB(\lambda, T^{a}\omega), dB(\lambda, T^{a}\omega), \cdots, dB(\lambda, T^{a}\omega)]$$

$$= \exp[i(\lambda_1 + \lambda_2 + \cdots + \lambda_n)a]h^{(n)}[dB(\lambda, \omega), dB(\lambda, \omega), \cdots, dB(\lambda, \omega)]. \tag{A10}$$

C. Orthogonal development of a functional

If a functional $\Phi(\omega)$ of $dB(\lambda)$ has a finite variance, then it has the orthogonal development in terms of the multiple Wiener integrals (sometimes called the Wiener–Hermite expansion)

$$\Phi(\omega) = A_0 + \sum_{n=1}^{\infty} \int_{R^2\times R^2\times\cdots\times R^2} \int \cdots \int A_n(\lambda_1, \lambda_2, \ldots, \lambda_n)h^{(n)}[dB(\lambda_1), dB(\lambda_2), \cdots, dB(\lambda_n)], \tag{A11}$$

where A_0 and A_n are a constant and a deterministic function symmetric with respect to its arguments, respectively. By (A 9) the symmetric function A_n is uniquely given as

$$\langle \Phi(\omega)h^{(n)*}[dB(\lambda_1), dB(\lambda_2), \cdots, dB(\lambda_n)]\rangle = n!\,A_n(\lambda_1, \lambda_2, \cdots, \lambda_n)\,d\lambda_1\,d\lambda_2 \ldots d\lambda_n. \tag{A12}$$

By (A10) a random field $\Phi(T^{r}\omega)$ derived by the measure-preserving transformation is represented as

$$\Phi(T^{r}\omega) = A_0 + \sum_{n=1}^{\infty} \int_{R^2\times\cdots\times R^2} \int \cdots \int A_n(\lambda_1, \lambda_2, \cdots, \lambda_n)\exp[i(\lambda_1 + \lambda_2 + \cdots\lambda_n)r]h^{(n)}[dB(\lambda_1), dB(\lambda_2), \ldots, dB(\lambda_n)]. \tag{A13}$$

By (A9) we easily obtain the average and the correlation function of $\Phi(T^{r}\omega)$

$$\langle \Phi(T^{r}\omega)\rangle = A_0, \tag{A14}$$

$$\langle \Phi(T^{r+a}\omega)\Phi^{*}(T^{r}\omega)\rangle = |A_0|^2 + \sum_{n=1}^{\infty} n!\int_{R^2\times\cdots\times R^2} \int \cdots \int |A_n(\lambda_1, \lambda_2, \ldots S\lambda_n)|^2 e^{i(\lambda_1 + \lambda_2 + \cdots + \lambda_n)a}\,d\lambda_1\,d\lambda_2\cdots d\lambda_n. \tag{A15}$$

[1] S. O. Rice, Commun. Pure Appl. Math. **4** 351 (1951).
[2] F. G. Bass and I. M. Fuks, *Wave Sattering from Statistically Rough Surfaces* (Pergamon, New York, 1979).
[3] P. Backmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Pergamon, New York, 1963).
[4] V. D. Freilikher and I. M. Fuks, Izv. Vepsh. Uchebn. Zaved. **13** 98 (1970) [Radiophys. Electron. **13**, 73 (1970)].
[5] G. G. Zipfel and J. A. DeSanto, J. Math. Phys. **13**, 1903 (1972).
[6] Lord Rayleigh (J. W. Strutt), Proc R. Soc. Lond. Ser. A **79** 399, (1907).
[7] C. Eckart, Phys. Rev. **44** 12 (1933).
[8] J. L. Uretsky, Ann. Phys. **33**, 400 (1965).
[9] H. Ogura, Phys. Rev. A **11**, 942 (1975).
[10] H. Ogura and J. Nakayama, Phys. Rev. A **11**, 957 (1975).
[11] N. Wiener, Am. J. Math. **60**, 897 (1938).
[12] N. Wiener, *Nonliner Problems in Random Theory* (M.I.T Press, Cambridge, Mass. and Wiley, New York, 1958).
[13] K. Itô, J. Math. Soc. Jpn. **3**, 158 (1951).
[14] K. Itô, Jpn. J. Math. **22**, 63 (1952).

Covariant electrodynamics of a dyon in a medium

J. Cohn and N. Hong[a)]
Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73069

(Received 18 January 1980; accepted for publication 18 April 1980)

A covariant formulation of electrodynamics of a dyon in a uniform, isotropic, and transparent medium is discussed. Using generalized Maxwell's equations admitting magnetic charge, the 4-potentials of a dyon of electric charge e and magnetic charge m and the corresponding field tensor are calculated. This field tensor is then used to calculate the various stress tensors of Minkowski, Abraham, and Marx. And it is found that the dyon behaves like an electrically charged particle of "effective" charge e^*, where $e^* = e[1 + (\epsilon/\mu)(m/e)^2]^{1/2}$, where ϵ and μ are the electric and magnetic permeabilities of the medium.

PACS numbers: 03.50.De

1. INTRODUCTION

The electrodynamics of a particle with both electric and magnetic charge (dyons) has been discussed on various occassions. Beginning with Dirac's proposal of the possible existence of a magnetic monopole, however, most considerations have involved quantum mechanical treatments.[1]

On the classical side, Rund recently developed a systematic construction of generalized 4-potentials, to describe the electrodynamics of dyons in a vacuum.[2]

The purpose of the present work is to develop a covariant formulation of the electrodynamics of a dyon in a medium. Towards this purpose, we will first express the field tensor in terms of the relevant generalized 4-potentials. Then the field tensor is used to calculate the electromagnetic stress tensor. Here, we consider three types of macroscopic electromagnetic stress tensors, that of Minkowski, Abraham, and Marx.[3] It is shown that these electromagnetic stress tensors for the dyon case take the same form as those of the electron case except that charge e is replaced by "effective" charge e^*, where $e^* = e[1 + (\epsilon/\mu)(m/e)^2]^{1/2}$, where m is the magnetic charge and ϵ and μ are the electric and magnetic permeabilities of the medium.

We comment that the motivation for this work was the hope of finding some characteristic difference between the radiation fields (or stress tensors) of magnetically and electrically charged particles, which might then facilitate the detection of the former. However, in the vacuum case no such difference seems to exist, and so we turned our attention to the medium case where we do, as mentioned above, find such a difference. This will be elaborated upon in the last section of the paper.

2. FIELD OF A DYON

In this section we intend, first, to express the field tensors covariantly in terms of generalized 4-potentials. And then we will calculate the field generated by a dyon moving arbitrarily (with speed less than that of light in the medium) in a uniform transparent medium.

A. Maxwell's equations and 4-potentials

We begin by assuming the validity of the following generalized Maxwell's equation in the medium rest frame:

$$\nabla\times\mathbf{B} = \frac{4\pi\mu}{c}\mathbf{J} + \frac{\epsilon\mu}{c}\dot{\mathbf{E}}, \tag{2.1}$$

$$\nabla\times\mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}} - \frac{4\pi}{c}\mathbf{S}, \tag{2.2}$$

$$\nabla\cdot\mathbf{B} = 4\pi\sigma, \tag{2.3}$$

$$\nabla\cdot\mathbf{E} = (4\pi/\epsilon)\rho, \tag{2.4}$$

where $\mathbf{B}$ and $\mathbf{E}$ are the magnetic field and electric field, respectively; ϵ and μ are constant permeabilities of a transparent, uniform, and boundaryless medium; σ and ρ are the magnetic charge density and electric charge density, respectively; $\mathbf{S}$ and $\mathbf{J}$ are the magnetic current density and the electric current density, respectively; and c is the speed of light in vacuum.

One of the authors recently showed that in the electron case, i.e., when $\sigma = \mathbf{S} = 0$, the ordinary Maxwell's equations can be expressed as[4]

$$\tilde{F}^{\mu\nu}{}_{,\nu} = \frac{4\pi\mu}{c}J^{\mu}, \tag{2.5}$$

where

$$\tilde{F}^{\mu\nu} \equiv F^{\mu\nu} - (1/\alpha u)^2 V_{\alpha}(V^{\nu}F^{\mu\alpha} - V^{\mu}F^{\nu\alpha}), \tag{2.6}$$

where u is the speed of light in the medium; $\alpha = [1 - (u/c)^2]^{-1/2}$; V^{μ} is the 4-velocity of the medium; $F^{\mu\nu}$ is the antisymmetric field tensor which is in the medium rest frame (signified by the naught subscript),

$$F^{\mu\nu}_{(0)} = \begin{pmatrix} 0, & B_3, & -B_2, & -E_1 \\ -B_3, & 0, & B_1, & -E_2 \\ B_2, & -B_1, & 0, & -E_3 \\ E_1, & E_2, & E_3, & 0 \end{pmatrix}_{(0)}, \tag{2.7}$$

and

$$F^{*\mu\nu}{}_{,\nu} = 0, \tag{2.8}$$

where $F^{*\mu\nu}$ is defined as $F^{*\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$; $F^{*}_{\mu\nu} \equiv g_{\mu\alpha}\, g_{\nu\beta}F^{*\alpha\beta}$; and $g^{\mu\nu} = (1,1,1,-1)$.

Then $\tilde{F}^{\mu\nu}$ and $F^{*\mu\nu}$ are, in the medium rest frame, given by

a)Submitted in partial fulfillment of the PhD. degree at the University of Oklahoma, Norman, Oklahoma.

 022-2488/81/030478-06$1.00

$$\tilde{F}^{\mu\nu}_{(0)} = \begin{pmatrix} 0, & B_3, & -B_2, & -\epsilon\mu E_1 \\ -B_3, & 0, & B_1, & -\epsilon\mu E_2 \\ B_2, & -B_1, & 0, & -\epsilon\mu E_3 \\ \epsilon\mu E_1, & \epsilon\mu E_2, & \epsilon\mu E_3, & 0 \end{pmatrix}_{(0)} \tag{2.9}$$

and

$$F^{*\mu\nu}_{(0)} = \begin{pmatrix} 0, & E_3, & -E_2, & B_1 \\ -E_3, & 0, & E_1, & B_2 \\ E_2, & -E_1, & 0, & B_3 \\ -B_1, & -B_2, & -B_3, & 0 \end{pmatrix}_{(0)} . \tag{2.10}$$

Using the same definitions and notations, Eqs. (2.1)–(2.4) become

$$\tilde{F}^{\mu\nu}{}_{,\nu} = \frac{4\pi\mu}{c} J^\mu \tag{2.11}$$

and

$$F^{*\mu\nu}{}_{,\nu} = -\frac{4\pi}{c} S^\mu, \tag{2.12}$$

where $J^\mu_{(0)} = (\mathbf{J}, c\rho)_{(0)}$ and $S^\mu_{(0)} = (\mathbf{S}, c\sigma)_{(0)}$.

We now wish to express $F^{\mu\nu}$ in terms of generalized 4-potentials $\phi^\mu = (\boldsymbol{\phi}, U)$ and $\psi^\mu = (\boldsymbol{\psi}, V)$, which are supposed to satisfy the following equations in the medium rest frame:

$$\Box'\boldsymbol{\phi} = -\frac{4\pi\mu}{c}\mathbf{J}, \tag{2.13}$$

$$\Box' U = -\frac{4\pi}{\epsilon}\rho, \tag{2.14}$$

$$\Box'\boldsymbol{\psi} = -\frac{4\pi}{c} S, \tag{2.15}$$

$$\Box' V = -\frac{4\pi}{\epsilon\mu}\sigma, \tag{2.16}$$

where $\Box' \equiv \nabla^2 - (\epsilon\mu/\sigma^2)(\partial^2/\partial t^2)$.

Towards this end we will, firstly, express **B** and **E** in terms of the potentials.

By taking the curl of Eq. (2.1), we obtain

$$\nabla\times\nabla\times\mathbf{B} = \nabla(\nabla\cdot\mathbf{B}) - \nabla^2\mathbf{B} = \frac{4\pi\mu}{c}\nabla\times\mathbf{J} + \frac{\epsilon\mu}{c}\nabla\times\dot{\mathbf{E}}. \tag{2.17}$$

Substitution of Eq. (2.2) for $\nabla\times\mathbf{E}$ and Eq. (2.3) for $\nabla\cdot\mathbf{B}$ yields

$$\Box'\mathbf{B} = 4\pi\nabla\sigma - \frac{4\pi\mu}{c}\nabla\times\mathbf{J} + \frac{4\pi\epsilon\mu}{c^2}\dot{\mathbf{S}}. \tag{2.18}$$

This becomes, after using Eqs. (2.13)–(2.16)

$$\Box'\left[\mathbf{B} - \left\{-\epsilon\mu\nabla V + \nabla\times\boldsymbol{\phi} - \frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}}\right\}\right] = 0. \tag{2.19}$$

The argument of $\Box'$ is a certain vector, say **A**, which satisfies

$$\Box'\mathbf{A} = 0. \tag{2.20}$$

Then

$$\mathbf{B} = -\epsilon\mu\nabla V + \nabla\times\boldsymbol{\phi} - \frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}} + \mathbf{A}. \tag{2.21}$$

Similarly, by taking the curl of Eq. (2.2) and using the above method, we obtain the expression for **E** in terms of the potentials:

$$\mathbf{E} = -\nabla U - \nabla\times\boldsymbol{\psi} - \frac{1}{c}\dot{\boldsymbol{\phi}} + \mathbf{C}, \tag{2.22}$$

where $\Box'\mathbf{C} = 0$.

Since, from Eqs. (2.21) and (2.22), **B** and **E** are left unchanged by the transformations

$$\begin{aligned} &\boldsymbol{\phi}\to\boldsymbol{\phi} + \nabla\Lambda, \\ &\boldsymbol{\psi}\to\boldsymbol{\psi} + \nabla\Gamma, \\ &U\to U - \frac{1}{c}\dot{\Lambda}, \\ &V\to V - \frac{1}{c}\dot{\Gamma}, \end{aligned} \tag{2.23}$$

where Λ and Γ are arbitrary, we have the freedom of choosing the potentials such that[5]

$$\nabla\cdot\boldsymbol{\phi} + \frac{\epsilon\mu}{c}\dot{U} = 0 \tag{2.24}$$

and

$$\nabla\cdot\boldsymbol{\psi} + \frac{\epsilon\mu}{c}\dot{V} = 0. \tag{2.25}$$

The gauge Eqs. (2.24) and (2.25) can be written in covariant form as

$$\partial_\alpha\tilde{\phi}^\alpha_{(0)} = 0 \quad \text{and} \quad \partial_\alpha\tilde{\psi}^\alpha_{(0)} = 0, \tag{2.26}$$

where

$$\tilde{\phi}^\mu_{(0)} = (\boldsymbol{\phi}, \epsilon\mu U)_{(0)} = (\phi^\mu - (1/\alpha u)^2 V_\alpha \phi^\alpha V^\mu)_{(0)}, \tag{2.27}$$

$$\tilde{\psi}^\mu_{(0)} = (\boldsymbol{\psi}, \epsilon\mu V)_{(0)} = (\psi^\mu - (1/\alpha u)^2 V_\alpha \psi^\alpha V^\mu)_{(0)}, \tag{2.28}$$

Thus, in any inertial frame, the gauge for the potentials are such that

$$\partial_\alpha\tilde{\phi}^\alpha = 0, \quad \text{and} \quad \partial_\alpha\tilde{\psi}^\alpha = 0. \tag{2.29}$$

We now proceed to show that **A** and **C** can be absorbed into the gauge transformations, Eq. (2.23).

From Eq. (2.21), we have

$$\nabla\times\mathbf{A} = \nabla\times\mathbf{B} - \nabla(\nabla\cdot\boldsymbol{\phi}) + \nabla^2\boldsymbol{\phi} + \frac{\epsilon\mu}{c}\nabla\times\dot{\boldsymbol{\psi}}. \tag{2.30}$$

Using Eq. (2.1), and substituting $\Box' + (\epsilon\mu/c^2)(\partial^2/\partial t^2)$ for ∇^2, we obtain

$$\nabla\times\mathbf{A} = \frac{\epsilon\mu}{c^2}\frac{\partial^2\boldsymbol{\phi}}{\partial t^2} - \nabla(\nabla\cdot\boldsymbol{\phi}) + \frac{\epsilon\mu}{c}\frac{\partial}{\partial t}(\nabla\times\boldsymbol{\psi} + \mathbf{E}). \tag{2.31}$$

This equation finally becomes, utilizing Eq. (2.22),

$$\begin{aligned} \nabla\times\mathbf{A} &= -\nabla\left(\nabla\cdot\boldsymbol{\phi} + \frac{\epsilon\mu}{c}\dot{U}\right) + \frac{\epsilon\mu}{c}\frac{\partial\mathbf{C}}{\partial t} \\ &= \frac{\epsilon\mu}{c}\frac{\partial\mathbf{C}}{\partial t}. \end{aligned} \tag{2.32}$$

And similarly, from Eqs. (2.21), (2.3), and (2.16), we obtain the relation

$$\begin{aligned} \nabla\cdot\mathbf{A} &= \nabla\cdot\mathbf{B} + \epsilon\mu\nabla^2 V + \frac{\epsilon\mu}{c}\nabla\cdot\dot{\boldsymbol{\psi}} \\ &= \frac{\epsilon\mu}{c}\frac{\partial}{\partial t}\left(\nabla\cdot\boldsymbol{\psi} + \frac{\epsilon\mu}{c}\dot{V}\right) = 0. \end{aligned} \tag{2.33}$$

Thus, we can write

$$\mathbf{A} = \nabla \times \boldsymbol{\phi}' \tag{2.34}$$

for some vector function $\boldsymbol{\phi}'$.

Similarly, the curl and div of $\mathbf{C}$ are found to be

$$\nabla \times \mathbf{C} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \tag{2.35}$$

$$\nabla \cdot \mathbf{C} = 0. \tag{2.36}$$

Thus, $\mathbf{C}$ can be written as

$$\mathbf{C} = -\nabla \psi' \tag{2.37}$$

for some vector function ψ'.

Substituting Eq. (2.34) into Eq. (2.35), and Eq. (2.37) into Eq. (2.32), we find that

$$\nabla \times \mathbf{A} = -\frac{\epsilon\mu}{c}\nabla \times \dot{\boldsymbol{\psi}}' \tag{2.38}$$

and

$$\nabla \times \mathbf{C} = -\frac{1}{c}\nabla \times \dot{\boldsymbol{\phi}}'. \tag{2.39}$$

Therefore, $\mathbf{A}$ and $\mathbf{C}$ are, in general, of the form

$$\mathbf{A} = -\frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}}' - \epsilon\mu \nabla V' \tag{2.40}$$

and

$$\mathbf{C} = -\frac{1}{c}\boldsymbol{\phi}' - \nabla U \tag{2.41}$$

for some functions V' and U'.

Now we observe that $\mathbf{A}$ and $\mathbf{C}$ are left unchanged by the transformations:

$$\begin{aligned} &\boldsymbol{\phi}' \to \boldsymbol{\phi}' + \nabla \Lambda', \\ &\boldsymbol{\psi}' \to \boldsymbol{\psi}' + \nabla \Gamma', \\ &U' \to U' - \frac{1}{c}\dot{\Lambda}', \\ &V' \to V' - \frac{1}{c}\dot{\Gamma}'. \end{aligned} \tag{2.42}$$

Again we have the freedom of choosing these "potentials" to satisfy

$$\nabla \times \boldsymbol{\phi}' + \frac{\epsilon\mu}{c}\dot{U}' = 0 \tag{2.43}$$

and

$$\nabla \times \boldsymbol{\psi}' + \frac{\epsilon\mu}{c}\dot{V}' = 0. \tag{2.44}$$

By taking the curl of Eqs. (2.34) and (2.40), and using Eqs. (2.37) and (2.41), we find that

$$\Box' \boldsymbol{\phi}' = \nabla\left(\nabla \cdot \boldsymbol{\phi}' + \frac{\epsilon\mu}{c}\dot{U}'\right) = 0. \tag{2.45}$$

Similarly, by taking the curl of Eqs. (2.37) and (2.41), and using Eqs. (2.34) and (2.40), we obtain

$$\Box' \boldsymbol{\psi}' = \nabla\left(\nabla \cdot \boldsymbol{\psi}' + \frac{\epsilon\mu}{c}\dot{V}'\right) = 0. \tag{2.46}$$

And by taking the div instead of the curl, we obtain

$$\Box' V' = 0, \tag{2.47}$$

$$\Box' U' = 0. \tag{2.48}$$

Now we substitute $\mathbf{A}$ and $\mathbf{C}$ into Eqs. (2.21) and (2.22), respectively, to obtain

$$\mathbf{B} = -\epsilon\mu\nabla(V + \tfrac{1}{2}V') + \nabla \times (\boldsymbol{\phi} + \tfrac{1}{2}\boldsymbol{\phi}') - \frac{\epsilon\mu}{c}(\dot{\boldsymbol{\psi}} + \tfrac{1}{2}\dot{\boldsymbol{\psi}}'), \tag{2.49}$$

$$\mathbf{E} = -\nabla(U + \tfrac{1}{2}U') - \nabla \times (\boldsymbol{\psi} + \tfrac{1}{2}\boldsymbol{\psi}') - \frac{1}{c}(\dot{\boldsymbol{\phi}} + \tfrac{1}{2}\dot{\boldsymbol{\phi}}'), \tag{2.50}$$

since, from Eqs. (2.34), (2.40), (2.37) and (2.41),

$$\begin{aligned} \mathbf{A} = \nabla \times \boldsymbol{\phi}' &= -\frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}}' - \epsilon\mu\nabla V' \\ &= \tfrac{1}{2}\left\{\nabla \times \boldsymbol{\phi}' - \frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}}' - \epsilon\mu\nabla V'\right\} \end{aligned}$$

and

$$\begin{aligned} \mathbf{C} = -\nabla\psi' &= -\frac{1}{c}\dot{\boldsymbol{\phi}}' - \nabla U' \\ &= -\tfrac{1}{2}\left(\nabla \times \boldsymbol{\psi}' + \frac{1}{c}\dot{\boldsymbol{\phi}}' + \nabla U'\right). \end{aligned}$$

Furthermore, we observe that the "new" potentials (= "old" potentials $+\frac{1}{2}$ of "primed" potential) obey all the equations governing the "old" potentials, that is, Eqs. (2.13)–(2.16) and(2.24)–(2.29). Thus, by choosing the "old" potentials to denote the "new" potentials, we restore these equations and have[6]

$$\mathbf{B} = -\epsilon\mu\nabla V + \nabla \times \boldsymbol{\phi} - \frac{\epsilon\mu}{c}\dot{\boldsymbol{\psi}}, \tag{2.51}$$

$$\mathbf{E} = -\nabla U - \nabla \times \boldsymbol{\psi} - \frac{1}{c}\dot{\boldsymbol{\phi}}. \tag{2.52}$$

We now wish to express the field tensor in covariant form in terms of the above generalized potentials. From Eqs. (2.51) and (2.52), we obtain

$$\begin{aligned} F^{14} = -E_1 &= \frac{\partial U}{\partial x} + \left(\frac{\partial \psi_3}{\partial y} - \frac{\partial \psi_2}{\partial z}\right) + \frac{1}{c}\frac{\partial \phi_1}{\partial t}, \\ &= (\phi^{4,1} - \phi^{1,4}) + \epsilon^{14\alpha\beta}\psi_{\beta,\alpha} \end{aligned} \tag{2.53}$$

$$\begin{aligned} F^{12} = B_3 &= -\epsilon\mu\frac{\partial V}{\partial z} + \left(\frac{\partial \phi_2}{\partial x} - \frac{\partial \phi_1}{\partial y}\right) - \frac{\epsilon\mu}{c}\frac{\partial \psi_3}{\partial t} \\ &= (\phi^{2,1} - \phi^{1,2}) + \epsilon\mu\,\epsilon^{12\alpha\beta}\psi_{\beta,\alpha}, \end{aligned} \tag{2.54}$$

which may be generalized to give

$$F^{ij} = \phi^{j,i} - \phi^{i,j} + \epsilon\mu\,\epsilon^{ij\alpha\beta}\psi_{\beta,\alpha}, \qquad (i,j = 1,2,3) \tag{2.55}$$

$$F^{i4} = \phi^{4,i} - \phi^{i,4} + \epsilon^{i4\alpha\beta}\psi_{\beta,\alpha}. \tag{2.56}$$

Combining these two equations together, we finally have

$$\begin{aligned} F^{\mu\nu} = (\phi^{\nu,\mu} - \phi^{\mu,\nu}) &+ \{\epsilon\mu\,\epsilon^{\mu\nu\alpha\beta} \\ &+ (1/\alpha u)^2 V_\gamma(V^\nu\epsilon^{\mu\gamma\alpha\beta} - V^\mu\epsilon^{\nu\gamma\alpha\beta})\}\,\psi_{\beta,\alpha}. \end{aligned} \tag{2.57}$$

Inversely, it can be easily shown that the expression for $F^{\mu\nu}$ in Eq. (2.57) with the gauges given by Eq. (2.29) satisfies Maxwell's equations, Eqs. (2.11) and (2.12).

For brevity of notation let us introduce new quantities:

$$f^{\mu\nu} \equiv \phi^{\nu,\mu} - \phi^{\mu,\nu} \tag{2.58}$$

and

$$b^{*\mu\nu} \equiv \psi^{\nu,\mu} - \psi^{\mu,\nu}. \tag{2.59}$$

Then $F^{\mu\nu}$ can, in general, be written as

$$F^{\mu\nu} = f^{\mu\nu} - \epsilon\mu b^{\mu\nu} - (1/\alpha u)^2 V_\gamma (V^\nu b^{\mu\gamma} - V^\mu b^{\nu\gamma}, \tag{2.60}$$

where

$$b^{*\mu\nu} = \tfrac{1}{2}\epsilon^{\mu\nu\alpha\beta} b_{\alpha\beta}, \quad \text{and} \quad b^*_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} b^{*\alpha\beta}.$$

B. Field of a Dyon

The potentials ϕ^μ and ψ^μ for a dyon of magnetic charge m and electric charge e are the solutions to Eqs. (2.13)–(2.16). The solutions to these equations for the electron case ($\sigma = \mathbf{S} = 0$) are found in Cohn's paper.[7] Certain portions of that paper which are related to our problem shall be reviewed, and we shall use the same notation as in that work.

As in the vacuum case, we define $R^\mu_{(0)} = (\mathbf{x} - \mathbf{z}, c\Delta t)_{(0)}$, where $\mathbf{x}$ and $\mathbf{z}$ denote the field and particle locations, and Δt is the time required for the wave in the medium to go from $\mathbf{z}$ to $\mathbf{x}$. Then we may write $R^\mu_{(0)} = (U^\mu + (1/u)\, V^\mu)_{(0)}$ where $U^\mu_{(0)} = (\hat{r}, 0)_{(0)}$ is a unit position vector in the medium rest frame, and ρ is just $|\mathbf{x} - \mathbf{z}|_{(0)}$.

From the 4-velocity, v^μ, of the particle, new quantities are introduced:

$$\tilde{v}^\mu \equiv v^\mu + \lambda\gamma c^2 V^\mu, \quad \tilde{\tilde{v}}^\mu \equiv v^\mu - \frac{\gamma}{\alpha^2} V^\mu, \tag{2.61}$$

where

$$\gamma = (1 - (v/c)^2)^{-1/2} \text{ and } \lambda = (u - c)/c^3,$$

and

$$\tilde{a}^\mu \equiv (d\tilde{v}^\mu/d\tau), \quad \tilde{\tilde{a}}^\mu \equiv (d\tilde{\tilde{v}}^\mu/d\tau).$$

And various quantities are defined for simplicitly of formalism, as

$$\tilde{R}^\mu = \rho\left(U^\mu + \frac{1}{c} V^\mu\right); \quad \tilde{R}^\alpha \tilde{v}_\alpha = -\tilde{\rho} u;$$

$$\Lambda^\mu = \frac{\rho}{\mu\tilde{\rho}}\left(U^\mu + \frac{u}{c^2} V^\mu\right).$$

Using these quantities, the solutions for the electron case (which corresponds to just ϕ^μ) ($\sigma = \mathbf{S} = 0$) are found to be

$$\phi^\mu = \frac{e\mu}{c} \frac{\tilde{\tilde{v}}^\mu}{\rho}\Bigg]_{\text{ret}}, \tag{2.62}$$

and thus

$$f^{\mu\nu} = \frac{2e\mu}{cu}\left\{\left(\frac{\gamma c}{\tilde{\rho}}\right)^2 \left(\frac{1}{\alpha^2} - \frac{1}{\gamma^2}\right) \tilde{\tilde{v}}^{[\nu} \Lambda^{\mu]} + \frac{\mu}{\tilde{\rho}} (\tilde{\tilde{a}}^{[\mu} \Lambda^{\nu]} - \Lambda \cdot a \tilde{\tilde{v}}^{[\nu} \Lambda^{\mu]})\right\}, \tag{2.63}$$

where e is the electric charge, $A \cdot B \equiv A^\alpha B_\alpha$, $A^{[\mu} B^{\nu]} \equiv \frac{1}{2}(A^\mu B^\nu - A^\nu B^\mu)$, and all quantities on the right-hand side are retarded.

Similarly, the solutions to Eqs. (2.13)–(2.16) can be found as

$$\psi^\mu = \frac{m}{c} \frac{\tilde{\tilde{v}}^\mu}{\tilde{\rho}}\Bigg]_{\text{ret.}}, \tag{2.64}$$

where m is the magnetic charge of the dyon.

We also find, from Eqs. (2.58), (2.59), (2.62), and (2.64), that

$$b^{\mu\nu} = -(m/\mu e) f^{*\mu\nu}. \tag{2.65}$$

Then, the field tensor for a dyon becomes

$$F^{\mu\nu} = f^{\mu\nu} + (m\epsilon/e) f^{*\mu\nu} - (1/\alpha c)^2 (m\epsilon/e)\, y^{\mu\nu}, \tag{2.66}$$

where

$$y^{\mu\nu} \equiv V_\gamma (-f^{*\mu\gamma} V^\nu + f^{*\nu\gamma} V^\mu) \tag{2.67}$$

and where $f^{\mu\nu}$ is determined by the dyon's retarded kinematics according to Eq. (2.63).

3. ELECTROMAGNETIC STRESS TENSOR

In this section we will calculate the various electromagnetic stress tensors—Minkowski, Abraham, and Marx—for the field produced by a dyon moving arbitrarily in the medium. We take the tensors as[8]

$$T_{\text{M}}{}^{\mu\nu} = \frac{1}{4\pi\mu}(F^{\mu\alpha}\widetilde{F}_\alpha{}^\nu + \tfrac{1}{4} g^{\mu\nu} F^{\alpha\beta}\widetilde{F}_{\alpha\beta}), \tag{3.1}$$

$$T_{\text{A}}{}^{\mu\nu} = T_{\text{M}}{}^{\mu\nu} + \frac{1}{4\pi\mu}\left(\frac{1}{\alpha u}\right)^2 \times \left(F^{\mu\alpha} F_\alpha{}^\beta V_\beta V^\nu - \frac{1}{c^2} F^{\alpha\beta} F_\alpha{}^\gamma V_\beta V_\gamma V^\mu V^\nu\right), \tag{3.2}$$

$$T_{\text{S}}{}^{\mu\nu} = \left(\frac{u}{c}\right)^2 \left[T_{\text{M}}{}^{\mu\nu} - \frac{1}{4\pi\mu}\left(\frac{1}{\alpha u}\right)^2 \left\{V^\mu F^{\nu\alpha} F_\alpha{}^\beta V_\beta + \tfrac{1}{2} V^\mu \times V^\nu \left(\frac{1}{\alpha^2 u^2} F^{\alpha\beta} F_\alpha{}^\gamma V_\beta V_\gamma + \tfrac{1}{2} F^{\alpha\beta} F_{\alpha\beta}\right)\right\}\right], \tag{3.3}$$

where subscripts M, A, and S designate Minkowski, Abraham, and Marx, respectively. Using the definition, Eq. (2.6), these can be written as

$$T_{\text{M}}{}^{\mu\nu} = \frac{1}{4\pi\mu}[F^{\mu\alpha} F_\alpha{}^\nu + (1/\alpha u)^2 (F^{\mu\alpha} F^{\nu\beta} V_\alpha V_\beta - F^{\mu\beta} F_\beta{}^\alpha V_\alpha V^\nu) + \tfrac{1}{4} g^{\mu\nu}(F^{\alpha\beta} F_{\alpha\beta} - (2/\alpha^2 u^2) F^{\alpha\beta} F_\alpha{}^\gamma V_\beta V_\gamma)], \tag{3.4}$$

$$T_{\text{A}}{}^{\mu\nu} = T_{\text{A}}{}^{\nu\mu} = T_{\text{M}}{}^{\mu\nu} + (1/4\pi\mu)(1/\alpha u)^2 (F^{\mu\alpha} F_\alpha{}^\beta V_\beta V^\nu - (1/c^2) F^{\alpha\beta} F_\alpha{}^\gamma V_\beta V_\gamma V^\mu V^\nu), \tag{3.5}$$

$$T_{\text{S}}{}^{\mu\nu} = T_{\text{S}}{}^{\nu\mu} = \frac{u^2}{c^2}\left[T_{\text{M}}{}^{\mu\nu} - \frac{1}{4\pi\mu}(1/\alpha u)^2 \{V^\mu F^{\nu\alpha} F_\alpha{}^\beta V_\beta + \tfrac{1}{2} V^\mu V^\nu (F^{\alpha\beta} F_\alpha{}^\gamma V_\beta V_\gamma (1/\alpha^2 u^2) + \tfrac{1}{2} F^{\alpha\beta} F_{\alpha\beta})\}\right]. \tag{3.6}$$

Now these tensors will be evaluated by substituting Eq. (2.66) into Eqs. (3.4)–(3.6).

A. Minkowski

Substituting Eq. (2.66) into Eq. (3.4), we obtain

$$T_{\text{M}}{}^{\mu\nu} = \frac{1}{4\pi\mu}\left[(ff)^{\mu\nu} + \left(\frac{m\epsilon}{e}\right)^2 \{(f^* f^*)^{\mu\nu}\right.$$

$$-\left(\frac{1}{\alpha c}\right)^2(f^*y+yf^*)^{\mu\nu}+\left(\frac{1}{\alpha c}\right)^4(yy)^{\mu\nu}\Big\}$$
$$+\left(\frac{m\epsilon}{e}\right)\left\{(ff^*+f^*f)^{\mu\nu}-\left(\frac{1}{\alpha c}\right)^2(fy+yf)^{\mu\nu}\right\}\Big], \tag{3.7}$$

where

$$(ab)^{\mu\nu}\equiv a^{\mu\alpha}b_\alpha{}^\nu+\left(\frac{1}{\alpha u}\right)^2(a^{\mu\alpha}b^{\nu\beta}V_\alpha V_\beta-a^{\mu\beta}b_\beta{}^\alpha V_\alpha V^\nu)$$
$$+\tfrac{1}{4}g^{\mu\nu}\left(a^{\alpha\beta}b_{\alpha\beta}+\frac{2}{\alpha^2u^2}\,a^{\alpha\beta}b_\beta{}^\gamma V_\alpha V_\gamma\right), \tag{3.8}$$

$$(a+b)^{\mu\nu}\equiv a^{\mu\nu}+b^{\mu\nu}, \tag{3.9}$$

$$\{(a+b)(c+d)\}^{\mu\nu}\equiv(ac+ad+bc+bd)^{\mu\nu}. \tag{3.10}$$

The first term in Eq. (3.7) is the contribution of electric charge to the tensor, the second term of magnetic charge, and the third term of the interference between electric and magnetic charges. We now proceed to express the second term in terms of $f^{\mu\nu}$.

From the definition, Eq. (3.8), $(f^*f^*)^{\mu\nu}$ is

$$(f^*f^*)^{\mu\nu}$$
$$=f^{*\mu\alpha}f^{*\nu}_\alpha+\left(\frac{1}{\alpha u}\right)^2(f^{*\mu\alpha}f^{*\nu\beta}V_\alpha V_\beta-f^{*\mu\beta}f^{*\alpha}_\beta V_\alpha V^\nu)$$
$$+\tfrac{1}{4}g^{\mu\nu}\left(f^{*\alpha\beta}f^*_{\alpha\beta}+\frac{2}{\alpha^2u^2}f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma\right). \tag{3.11}$$

Using Eqs. (2.67) and (3.8), we find

$$(yy)^{\mu\nu}=\frac{c^4}{u^2}f^{*\mu\alpha}f^{*\nu\beta}V_\alpha V_\beta+\frac{c^2}{u^2}V^\mu V^\nu f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma$$
$$+\tfrac{1}{2}g^{\mu\nu}\left(\frac{c^4}{u^2}\right)f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma. \tag{3.12}$$

We also obtain, from Eqs. (2.67) and (3.8)–(3.10),

$$(f^*y+yf^*)^{\mu\nu}=\frac{2c^2}{u^2}f^{*\mu\alpha}f^{*\nu\beta}V_\alpha V_\beta-V^\mu f^{*\nu\alpha}f^{*\beta}_\alpha V_\beta$$
$$-\frac{c^2}{u^2}V^\nu f^{*\mu\alpha}f^{*\beta}_\alpha V_\beta$$
$$+\left(\frac{1}{\alpha c}\right)^2V^\mu V^\nu f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma$$
$$+\frac{c^2}{u^2}g^{\mu\nu}f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma. \tag{3.13}$$

By combining Eqs. (3.11)–(3.13), we obtain

$$(f^*f^*)^{\mu\nu}-\left(\frac{1}{\alpha c}\right)^2(f^*y+yf^*)^{\mu\nu}+\left(\frac{1}{\alpha c}\right)^4(yy)^{\mu\nu}$$
$$=f^{*\mu\alpha}f^{*\nu}_\alpha-\left(\frac{1}{\alpha c}\right)^2(f^{*\mu\alpha}f^{*\nu\beta}V_\alpha V_\beta-V^\mu f^{*\nu\alpha}f^{*\gamma}_\alpha V_\gamma)$$
$$+\tfrac{1}{4}g^{\mu\nu}\left(f^{*\alpha\beta}f^*_{\alpha\beta}-\frac{2}{\alpha^2c^2}f^{*\alpha\beta}f^{*\gamma}_\beta V_\alpha V_\gamma\right). \tag{3.14}$$

To express Eq. (3.14) in terms of $f^{\mu\nu}$, we will use the identities:

$$f^{*\mu\alpha}f^{*\nu}_\alpha=f^{\mu\alpha}f^\nu_\alpha+\tfrac{1}{2}g^{\mu\nu}f^{\alpha\beta}f_{\alpha\beta};\quad f^{*\alpha\beta}f^*_{\alpha\beta}=-f^{\alpha\beta}f_{\alpha\beta};$$
$$f^{*\mu\alpha}f^{*\nu\beta}V_\alpha V_\beta$$
$$=c^2f^{\mu\alpha}f_\alpha{}^\nu+V^\mu f^{\nu\alpha}f_\alpha{}^\beta V_\beta+V^\nu f^{\mu\alpha}f_\alpha{}^\beta V_\beta$$
$$+\tfrac{1}{2}V^\mu V^\nu f^{\alpha\beta}f_{\alpha\beta}-f^{\mu\alpha}f^{\nu\beta}V_\alpha V_\beta+\tfrac{1}{2}g^{\mu\nu}(c^2f^{\alpha\beta}f_{\alpha\beta}$$
$$-2f^{\alpha\beta}f_\beta{}^\gamma V_\alpha V_\gamma). \tag{3.15}$$

Then, finally, Eq. (3.14) becomes

$$(f^*f^*)^{\mu\nu}-(1/\alpha c)^2(f^*y+yf^*)^{\mu\nu}+(1/\alpha c)^4(yy)^{\mu\nu}$$
$$=(u^2/c^2)(ff)^{\mu\nu}. \tag{3.16}$$

Next, we will consider the last term in Eq. (3.7) which represents the interference between electric and magnetic charges. Straightforward calculation with the help of the identity

$$f^{\mu\alpha}f^{*\nu}_\alpha=-\tfrac{1}{4}g^{\mu\nu}f^{\alpha\beta}f^*_{\alpha\beta} \tag{3.17}$$

shows that

$$(ff^*+f^*f)^{\mu\nu}-(1/ac)^2(fy+yf)^{\mu\nu}=0. \tag{3.18}$$

Therefore, by substituting Eqs. (3.16) and (3.18) into Eq. (3.7), we obtain the Minkowski tensor for a dyon:

$$T_{\rm M}{}^{\mu\nu}=\frac{1}{4\pi\mu}\left\{1+(u/c)^2\left(\frac{m\epsilon}{e}\right)^2\right\}(ff)^{\mu\nu},$$
$$=\left\{1+(\epsilon/\mu)\left(\frac{m}{e}\right)^2\right\}T^{\mu\nu}_{{\rm M}(e)}, \tag{3.19}$$

where we substituted $(u/c)^2=1/\epsilon\mu$, and the subscript (e) designates the tensor for electric charge without any magnetic charge present.

B. Abraham and Marx

After a somewhat lengthy but straightforward calculation of the same fashion as for the Minkowski case, we arrive at the same result as Eq. (3.19) for both the Abraham and Marx tensors.

Thus, all three tensors for a dyon differ from those of the electrically charged particle without magnetic charge by a factor of $1+(\epsilon/\mu)(m/e)^2$. For all three tensors, explicitly, we have

$$T^{\mu\nu}_{\rm dyon}=\left\{1+\left(\frac{\epsilon}{\mu}\right)\left(\frac{m}{e}\right)^2\right\}T^{\mu\nu}_{(e)}. \tag{3.20}$$

In the next subsection, we will briefly consider the 4-divergences of the three tensors.

C. Calculation of 4-divergence

We first note that from the definition of $\tilde F^{\mu\nu}$, Eq. (2.6),

$$F_{\alpha\beta}\tilde F^{\alpha\beta}{}_{,\nu}=\tilde F^{\alpha\beta}F_{\alpha\beta,\nu} \tag{3.21}$$

and

$$F^{\mu\alpha}{}_{,\nu}\tilde F_\alpha{}^\nu=-\tfrac{1}{4}\epsilon^{\mu\alpha\beta\gamma}F^*_{\beta\gamma}{}^{,\nu}\epsilon_{\alpha\nu\xi\eta}\tilde F^{*\xi\eta}$$
$$=\tfrac{1}{4}\delta^{\mu\beta\gamma}_{\nu\xi\eta}F^*_{\beta\gamma}{}^{,\nu}\tilde F^{*\xi\eta}$$
$$=-\tfrac{1}{2}g^{\mu\nu}F^{\alpha\beta}\tilde F_{\alpha\beta,\nu}-(4\pi/c)\tilde F^{*\mu\alpha}S_\alpha, \tag{3.22}$$

$$F^{\mu\alpha}{}_{,\nu}F_\alpha{}^\nu=-\tfrac{1}{2}g^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta,\nu}-(4\pi/c)F^{*\alpha\beta}S_\alpha$$
$$=\tfrac{1}{2}g^{\mu\nu}F^{*\alpha\beta}F^*_{\alpha\beta,\nu}-(4\pi/c)F^{*\alpha\beta}S_\alpha. \tag{3.23}$$

Using these equations, we find that

$$T_{\rm M}{}^{\mu\nu}{}_{,\nu}=(1/c)(F^{\mu\alpha}J_\alpha-(1/\mu)\tilde F^{*\mu\alpha}S_\alpha) \tag{3.24}$$

and

$$T_{\rm S}{}^{\mu\nu}{}_{,\nu}=(u^2/c^3)\{(F^{\mu\alpha}J_\alpha-(1/\mu)\tilde F^{*\mu\alpha}S_\alpha)$$
$$-(1/\alpha u)^2(F^{\alpha\beta}J_\beta-(1/\mu)F^{*\alpha\beta}S_\beta)V_\alpha V^\mu\}. \tag{3.25}$$

In the sourceless case, therefore, both become identically zero, i.e.,

$$T_{\mathrm{M}}{}^{\mu\nu}{}_{,\nu} = T^{\mu\nu}_{\mathrm{S},\nu} \equiv 0, \quad \text{when } J^{\mu} = S^{\mu} = 0. \tag{3.26}$$

But $T_{\mathrm{A}}{}^{\mu\nu}{}_{,\nu}$ is not identically zero, when there is no charge present; this can be shown from the following equation

$$T_{\mathrm{A}}{}^{\mu\nu}{}_{,\nu} = T_{\mathrm{M}}{}^{\mu\nu}{}_{,\nu} + (1/4\pi\mu)(1/\alpha u)^2(F^{\mu\alpha}{}_{,\nu}F_{\alpha}{}^{\beta}V_{\beta}V^{\nu} - (2/c^2)F^{\alpha\beta}{}_{,\nu}F_{\alpha}{}^{\gamma}V_{\beta}V_{\gamma}V^{\mu}V^{\nu}). \tag{3.27}$$

4. DISCUSSION

In a previous paper, the authors have discussed the energy–momentum and angular momentum emission rates for an electrically charged particle moving arbitrarily in a uniform transparent medium with speed less than that of light in the medium.[9] Comparing that with this paper, we find very similar features. The electromagnetic stress tensors under consideration have the same form for both the electric and magnetic cases, except that e in the electric charge case is replaced by an "effective" charge e^* in the dyon case, where

$$e^* = e[1 + (\epsilon/\mu)(m/e)^2]^{1/2}. \tag{4.1}$$

Furthermore, the 4-divergences of the three tensors, in the charge free case, behave similarly, i.e., the Minkowski and Marx tensors are divergenceless in both cases, and the 4-divergence of the Abraham tensor is not identically zero in both cases.

Therefore, the energy-momentum and angular momentum emission rates for a dyon moving arbitrarily (whose speed is less than that of light in the medium) in a uniform transparent medium are the same as those for a pure electric charge, except that e is replaced by e^*. All the arguments discussed in the previous paper concerning energy–momentum and angular momentum emitted by a chaged particle are applicable to the dyon.[9] We will not repeat them here.

Despite their similarities, the difference between the electric charge and dyon is conspicuous. Since the difference arises due to the existence of magnetic charge, let us consider the pure magnetic monopole case. By setting $e = 0$, we find that the "effective" charge of a monopole, m^*, is

$$m^* = (\epsilon/\mu)^{1/2}m. \tag{4.2}$$

In a vacuum we cannot distinguish the magnetic monopole from the electron qualitatively, in the sense that the electromagnetic tensors for them are indistinguishable, except for their charges. But in a medium the magnetic monopole behaves like a particle of medium-dependent charge. Thus, the phenomena related to it and its motion would be significantly different from those of the electron. If we could arrange layers of media and observe certain phenomena such as radiation or the motion of a particle moving through the layers, it would then be possible to tell the difference between the electron and the magnetic monopole.

The theory developed in this paper does not provide us with the tools to analyze the radiation of a charged particle moving through an inhomogeneous medium, but we still expect that the effective charge of the magnetic monopole is medium dependent.[10]

It should be remarked here that, even though the theory presented in this paper reveals certain properties of magnetic charge, we cannot claim that the tensors considered represent the *real* situation correctly. Nevertheless, it is suggestive that all three tensors, Minkowski, Abraham, and Marx, agree on at least two points: A dyon behaves like a charged particle of effective charge e^* and there is no interference effect due to the simultaneous existence of the two types of charge.

It should be remarked that, if the existence of the dyon is assumed, it is possible that ordinary matter may contain a certain amount of magnetic charge, and Maxwell's equations, Eqs. (2.1)–(2.4), may have to be changed accordingly.[11]

[1] See, for example, P. A. M. Dirac, Proc. R. Soc. London, Ser. A **133**, 60–72 (1931); J. Schwinger, Phys. Rev. D **12**, 3105–11 (1975).
[2] Hanno Rund, J. Math. Phys. **18**, 84–95 (1977).
[3] For a fuller discussion of the various stress tensors and their merits see I. Brevik, Dan. Viden. Selsk. **27**, No. 11 (1970); **37**, No. 13 (1970), and also his preprint *Experiments in Phenomenological Electrodynamics and the Electromagnetic Energy–Momentum Tensor* (University of Trondheim, 1978).
[4] Jack Cohn, Ann. Phys. (N.Y.) **114**, 467–78 (1978).
[5] These gauges are a generalized form of the gauge introduced in, for example, J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1967), pp. 180–181.
[6] This result is obtained differently by Rund, see Ref. 2.
[7] J. Cohn, see Ref. 4.
[8] I. Brevik, see Ref. 3.
[9] J. Cohn and N. Hong, to be published in Ann. Phys. (N.Y.).
[10] P. B. Price *et al.*, Phys. Rev. Lett. **35** (1975), in analyzing the image on a Cherenkov detector, used the formula, Intensity $\propto m^2[n^2 - (u/v)^2]$.
[11] L. L. Vant-Hull, Phys. Rev. **173**, 1412–3 (1968), measured the magnetic charge on the neutron and the difference of the magnetic charge of the proton and the electron, and found them to be less than 2×10^{-41} Wb. He also found that the magnetic charge of the electron is less than 8×10^{-39} Wb.

A multipole formalism for material media

Charles P. Frahm
Illinois State University, Normal, Illinois 61761
Naval Coastal Systems Center, Panama City, Florida 32407

(Received 15 July 1980; accepted for publication 19 September 1980)

It is shown that it is possible to make a meaningful multipole expansion for the electromagnetic fields produced by localized charge and current distributions embedded in material media—provided the media possess certain reasonable properties. The result is similar to the well-known multipole expansions for localized sources in vacuum but differs from it primarily because of induced charge and current density contributions to the various multipole coefficients.

PACS numbers: 03.50.De

I. INTRODUCTION

The usefulness of multipole expansions for electromagnetic fields is well known. They are especially helpful in the treatment of radiation from localized charge and current distributions. In addition they are convenient for characterizing known fields as well as fields obtained from numerical calculations and from empirical results.

The classic textbook presentation of multipole fields appears in Appendix B of *Theoretical Nuclear Physics* by Blatt and Weisskopf.[1] These authors give numerous references to the original literature. A more recent and thorough treatment is given in *Classical Electrodynamics* by Jackson.[2] Both these texts discuss multipole expansions for charge and current distributions in vacuum.

Intuitively, it would seem that multipole expansions also ought to exist for source distributions in material media—especially if the medium becomes linear and of uniform character at sufficiently large distances from the source distributions. For in that case it would appear that the medium at large distances would behave like a modified vacuum, differing from the true vacuum only in the numerical values assigned to the constitutive parameters. There would, of course, be complications due to induced charges and currents which would have to be properly taken into account.

It will be shown in this paper that such a multipole expansion does in fact exist for material media. Unfortunately, it has the unpleasant property of being an expansion of an integral equation so that the terms in the expansion involve the fields themselves. The utility of the expansion is thus somewhat restricted. However, it will be shown that the expansion terms depend on the fields only in localized regions so that its usefulness may be greater than first impressions might indicate. In particular the expansion may at least be helpful in characterizing the fields produced by sources embedded within material media.

To establish notation and to prepare the way for the derivation of the multipole expansion for material media, Sec. II reviews various aspects of the field equations. Section III introduces the concept of an asymptotic background or "modified vacuum" and then gives the form of the field equations as well as the wave equations appropriate for a system with an asymptotic background. The derivation itself is outlined in Sec. IV. It parallels closely the derivation for sources in vacuum as presented in Jackson. Consequently only substantive differences are indicated. In making the derivation it will be convenient to consider the medium to have continuous properties. An extension to discontinuous media is then made in Sec. V in conjunction with an interpretation of the effective source densities which contribute to the various multipole terms. SI units will be used throughout.

II. FIELD EQUATIONS

It will be sufficient to consider harmonic fields having a time-dependence of the form

$$\mathbf{E}\sim e^{i\omega t}, \tag{1}$$

where ω is the angular frequency. Maxwell's equations in vacuum then become

$$\nabla\cdot\mathbf{B}=0, \tag{2}$$

$$\nabla\times\mathbf{E}+i\omega\mathbf{B}=0, \tag{3}$$

$$\nabla\times\mathbf{B}-i\omega\mathbf{E}/c^2=\mu_0\mathbf{J}_t, \tag{4}$$

$$\nabla\cdot\mathbf{E}=\rho_t/\epsilon_0, \tag{5}$$

where c is the speed of light in vacuum, and μ_0 and ϵ_0 are the permeability and permittivity of free space, respectively. The subscript t on the charge density ρ_t and current density $\mathbf{J}_t$ is there to emphasize that these are the total source densities.

Equations (2)–(5) are of general validity and can be interpreted either as microscopic expressions or appropriately averaged macroscopic relations. The averaging process has been discussed at length by numerous authors.[2–5] It will suffice here to simply point out that the continuity and differentiability properties of the macroscopic fields are determined by the weight function used to perform the averaging. It will be assumed during most of this paper that the weight function is sufficiently smooth to assure the existence everywhere of all necessary derivatives. The usual procedure of allowing media discontinuities can be obtained as a limiting case in the end.

In material media it is convenient to isolate from the total source densities those parts which depend upon the total polarization $\mathbf{P}$ and total magnetization $\mathbf{M}$ of the medium. These parts are given by

$$\rho_p=-\nabla\cdot\mathbf{P}, \tag{6}$$

$$\mathbf{J}_p=i\omega\mathbf{P}, \tag{7}$$

$$\mathbf{J}_m=\nabla\times\mathbf{M}, \tag{8}$$

 022-2488/81/030484-05$1.00

and are called the polarization charge density, the polarization current density, and the magentization current density, respectively. The total source densities then exhibit the decompositions

$$\rho_t = \rho + \rho_p, \tag{9}$$

$$\mathbf{J}_t = \mathbf{J} + \mathbf{J}_p + \mathbf{J}_m, \tag{10}$$

where ρ and $\mathbf{J}$ are referred to as the macroscopic charge and current densities, respectively.

It will be assumed that the medium is linear and isotropic, but not necessarily homogeneous. Thus the polarization, magnetization, and current density can be expressed in terms of the electric and magnetic fields by

$$\mathbf{P} = (\epsilon - \epsilon_0)\mathbf{E}, \tag{11}$$

$$\mathbf{M} = (1/\mu_0 - 1/\mu)\mathbf{B}, \tag{12}$$

$$\mathbf{J} = \mathbf{J}_s + \sigma\mathbf{E}, \tag{13}$$

where $\mathbf{J}_s$ is the "specified," or field independent, part of the current density. The properties of the medium are now completely contained in the three constitutive parameters ϵ, μ, and σ. These, of course, are the permittivity, permeability, and conductivity, respectively, of the medium. In the case of inhomogeneous media they are functions of position.

Maxwell's two equations containing the sources then become

$$\nabla\cdot(\epsilon\mathbf{E}) = \rho, \tag{14}$$

$$\nabla\times(\mathbf{B}/\mu) - \eta\mathbf{E} = \mathbf{J}_s \tag{15}$$

where the parameter η has been introduced for convenience and is defined by

$$\eta = \sigma + i\omega\epsilon. \tag{16}$$

III. ASYMPTOTIC BACKGROUND

In order to establish a multipole expansion for material media it is necessary that the specified source density $\mathbf{J}_s$ be localized and that the medium become homogeneous at large distances. Thus it will be assumed that there exists a spherical surface S_0 (radius R_0) such that outside of S_0 the medium is homogeneous and $\mathbf{J}_s$ is zero. The system will then be said to possess an asymptotic background or modified vacuum having constitutive parameters equal to those of the medium outside S_0.

Taking the necessary derivatives in Eqs. (14) and (15) and using a subscript 2 to denote parameters associated with the asymptotic background, Maxwell's two source dependent equations can now be written in the form

$$\nabla\cdot\mathbf{E} = \rho_e/\epsilon_2, \tag{17}$$

$$\nabla\times\mathbf{B} - i(k_2^2/\omega)\mathbf{E} = \mu_2\mathbf{J}_e. \tag{18}$$

The quantities ρ_e and $\mathbf{J}_e$ are the effective charge and current densities defined by

$$\rho_e = -(\epsilon_2/\eta)(\nabla\cdot\mathbf{J}_s + \nabla\eta\cdot\mathbf{E}), \tag{19}$$

$$\mathbf{J}_e = (\mu/\mu_2)[\mathbf{J}_s - \nabla(1/\mu)\times\mathbf{B} + i(\delta k^2/\omega\mu)\mathbf{E}]. \tag{20}$$

Furthermore,

$$k^2 = -i\omega\mu(\sigma + i\omega\epsilon) \tag{21}$$

and

$$\delta k^2 = k^2 - k_2^2. \tag{22}$$

Clearly the effective sources contain contributions not only from the specified current density but from medium inhomogeneities. It will be seen in a later section that the latter contributions are just the familiar induced charge and current contributions.

It is now straightforward to derive the wave equations in the usual way, thereby obtaining

$$(\nabla^2 + k_2^2)\mathbf{E} = \mathbf{F}, \tag{23}$$

$$(\nabla^2 + k_2^2)\mathbf{B} = \mathbf{G}, \tag{24}$$

where

$$\mathbf{F} = \nabla\rho_e/\epsilon_2 + i\omega\mu_2\mathbf{J}_e \tag{25}$$

and

$$\mathbf{G} = -\mu_2\nabla\times\mathbf{J}_e. \tag{26}$$

IV. MULTIPOLE EXPANSION

It is evident from Eqs. (25) and (26) along with (19) and (20) that outside the surface S_0 both $\mathbf{F}$ and $\mathbf{G}$ are zero so that $\mathbf{E}$ and $\mathbf{B}$ satisfy the Helmholtz equation for the asymptotic background. Furthermore $\mathbf{E}$ and $\mathbf{B}$ are divergenceless outside S_0. It thus follows that outside S_0 the fields can be represented in the form of a multipole expansion

$$\mathbf{E} = \sum_{l,m}\{(-i\omega/k_2^2)a_E(l,m)\nabla\times[f_l(k_2r)\mathbf{X}_{l,m}] + a_M(l,m)g_l(k_2r)\mathbf{X}_{l,m}\}, \tag{27a}$$

$$\mathbf{B} = \sum_{l,m}\{a_E(l,m)f_l(k_2r)\mathbf{X}_{l,m} - (1/i\omega)a_M(l,m)\nabla\times[g_l(k_2r)\mathbf{X}_{lm}]\}. \tag{27b}$$

Except for the use of the asymptotic value k_2 in place of the vacuum value and some changes necessitated by the use of SI units instead of Gaussian units, these expressions and their derivations are identical to those given in Jackson (pp. 744–7). In particular f_l and g_l are solutions of Bessel's equation while $\mathbf{X}_{lm}$ is a vector spherical harmonic defined in terms of the spherical harmonics Y_{lm} by

$$\mathbf{X}_{lm} = -i\mathbf{r}\times\nabla Y_{lm}/(l(l+1))^{1/2}. \tag{28}$$

The quantities $a_E(l,m)$ and $a_M(l,m)$ are referred to as the electric and magnetic multipole coefficients, respectively. They can be expressed in terms of the fields by

$$a_E(l,m)f_l(k_2r) = \frac{k_2^2}{\omega(l(l+1))^{1/2}}\int Y^*_{lm}\,\mathbf{r}\cdot\mathbf{E}\,d\Omega, \tag{29a}$$

$$a_M(l,m)g_l(k_2r) = \frac{-\omega}{(l(l+1))^{1/2}}\int Y^*_{lm}\,\mathbf{r}\cdot\mathbf{B}\,d\Omega, \tag{29b}$$

where it is understood that r exceeds the radius of S_0 and the integral is over all solid angles.

Equation (29) can be used to express the multipole coefficients in terms of the effective sources. To accomplish this the vector identity

$$\mathbf{r}\cdot\nabla^2\mathbf{A} = \nabla^2(\mathbf{r}\cdot\mathbf{A}) - 2\nabla\cdot\mathbf{A} \tag{30}$$

may be used to cast the inhomogeneous wave equations (23) and (24) into the form

$$(\nabla^2 + k_2^2)(\mathbf{r}\cdot\mathbf{E}) = \mathbf{r}\cdot\mathbf{F} + 2\rho_e/\epsilon_2, \tag{31a}$$

$$(\nabla^2 + k_2^2)(\mathbf{r}\cdot\mathbf{B}) = \mathbf{r}\cdot\mathbf{G}. \tag{31b}$$

For (damped) outgoing waves at large distances these differential equations are equivalent to the following integral equations:

$$\mathbf{r}\cdot\mathbf{E}(\mathbf{r}) = -\frac{1}{4\pi}\int_{\text{a.s.}} \frac{e^{-ik_2|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,[\mathbf{r}\cdot\mathbf{F}(\mathbf{r}') + 2\rho_e(\mathbf{r}')/\epsilon_2]\,dV', \tag{32a}$$

$$\mathbf{r}\cdot\mathbf{B}(\mathbf{r}) = -\frac{1}{4\pi}\int_{\text{a.s.}} \frac{e^{-ik_2|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,\mathbf{r}'\cdot\mathbf{G}(\mathbf{r}')\,dV', \tag{32b}$$

where the integrals are over all space. The equivalence is easily verified using the well-known identity

$$(\nabla^2 + k_2^2)\left[\frac{e^{-ik_2|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\right] = -4\pi\delta(\mathbf{r}-\mathbf{r}'), \tag{33}$$

where δ is the Dirac delta-function. Substituting equations (32) into Eqs. (29) and using the identity[6]

$$\int Y_{lm}^* \frac{e^{-ik_2|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,d\Omega = -4\pi k_2 h_l^2(k_2 r) j_l(k_2 r') Y_{lm}^*(\theta'\phi') \tag{34}$$

yields after some simplification

$$a_E(l,m) = \frac{ik_2^3}{\omega(l(l+1))^{1/2}}\int_{\text{a.s.}} j_l(k_2 r) Y_{lm}^*(\mathbf{r}\cdot\mathbf{F} + 2\rho_e/\epsilon_2)\,dV, \tag{35a}$$

$$a_M(l,m) = \frac{-i\omega k_2}{(l(l+1))^{1/2}}\int_{\text{a.s.}} j_l(k_2 r) Y_{lm}^*\,\mathbf{r}\cdot\mathbf{G}\,dV. \tag{35b}$$

In these expressions the j_l and h_l^2 are the spherical Bessel and Hankel functions, respectively. The quantities f_l and g_l in Eqs. (29) have been identified with h_l^2 in obtaining Eqs. (35).

It is now straightforward to substitute for **F** and **G** from Eqs. (25) and (26), simplify, and obtain the desired expressions for the multipole coefficients in terms of the effective sources,

$$a_E(l,m) = \frac{-\mu_2 k_2}{(l(l+1))^{1/2}}\int_{\text{a.s.}} Y_{lm}^*\left\{\frac{\eta_2}{\epsilon_2}\rho_e\frac{\partial}{\partial r}[r j_l(k_2 r)] + k_2^2\,\mathbf{r}\cdot\mathbf{J}_e\, j_l(k_2 r)\right\}dV, \tag{36a}$$

$$a_M(l,m) = \frac{-i\omega\mu_2 k_2}{(l(l+1))^{1/2}}\int_{\text{a.s.}} Y_{lm}^*\,\nabla\cdot(\mathbf{r}\times\mathbf{J}_e)\, j_l(k_2 r)\,dV. \tag{36b}$$

Taking into account the changes required by the use of SI units these results are identical in form to those exhibited in Jackson except that no allowance has been made for intrinsic magnetization. The presence of the effective source densities, however, does produce a substantive difference from the traditional multipole expansion. It should be noted that although the integrals in Eqs. (36) are over all space, the vanishing of the source densities outside S_0 effectively limits the integrations to the region interior to S_0.

In the long wavelength limit

$$|k_2 R_0| \ll 1, \tag{37}$$

Eqs. (36) become

$$a_E(l,m) = \frac{-\eta_2\mu_2 k_2^{l+1}}{\epsilon_2(2l+1)!!}\left(\frac{l+1}{l}\right)^{1/2} Q_{lm}, \tag{38a}$$

$$a_M(l,m) = \frac{i\omega\mu_2 k_2^{l+1}}{(2l+1)!!}\left(\frac{l+1}{l}\right)^{1/2} M_{lm}, \tag{38b}$$

where Q_{lm} and M_{lm} are the electric and magnetic multipole moments, respectively. They are defined by

$$Q_{lm} = \int_{\text{a.s.}} r^l Y_{lm}^* \rho_e\,dV, \tag{39a}$$

$$M_{lm} = \frac{-1}{l+1}\int_{\text{a.s.}} r^l Y_{lm}^*\,\nabla\cdot(\mathbf{r}\times\mathbf{J}_e)\,dV. \tag{39b}$$

V. EFFECTIVE SOURCES

The discussion to this point has assumed that all quantities, including the constitutive parameters, are sufficiently smooth that all necessary derivatives exist. However, for the purpose of interpreting the significance of the effective sources it is convenient to consider homogeneous media separated by sharp boundaries. Such systems, of course, have discontinuous constitutive parameters. Nevertheless they can be treated as limiting cases of smooth systems by the introduction of generalized functions. This must be done with care, so that the resulting expressions are all well defined.

For purposes of illustration it will be assumed that the system consists of two homogeneous regions separated by a simply-connected closed surface S with the specified current density confined well within S. Quantities like $1/\eta$ appearing in the first term of the effective charge density, Eqs. (19), can then be written in the form

$$1/\eta = 1/\eta_1 + (1/\eta_2 - 1/\eta_1)\theta_s(\mathbf{r}), \tag{40}$$

where the subscript 1 refers to the region interior to S and $\theta_s(\mathbf{r})$ is the surface unit step function defined by

$$\theta_s(\mathbf{r}) = \begin{cases} 1 & \text{for } \mathbf{r} \text{ outside } S \\ 0 & \text{for } \mathbf{r} \text{ inside } S \end{cases}. \tag{41}$$

The second term in the effective charge density requires special attention. If the discontinuous limits of the various factors are taken separately the resulting product of generalized functions is meaningless.[7] On the other hand, for smooth systems the following two forms of the terms are obviously equivalent:

$$-(1/\eta)\nabla\eta\cdot\mathbf{E} = \nabla(1/\eta)\cdot(\eta\mathbf{E}). \tag{42}$$

The form on the right, however, is well defined in the discontinuous limit. To see this it is convenient to note from Eq. (40) that

$$\nabla(1/\eta) = \Delta(1/\eta)\delta_s(\mathbf{r})\hat{n}, \tag{43}$$

where $\hat{n}$ is the unit outward normal to the surface S, $\Delta(1/\eta)$ is the change in $1/\eta$ across the surface

$$\Delta(1/\eta) = 1/\eta_2 - 1/\eta_1, \tag{44}$$

and $\delta_s(\mathbf{r})$ is the surface delta function. The surface delta function has the property of converting a volume integral into a surface integral so that for an arbitrary function f

$$\int_{\text{a.s.}} f(\mathbf{r})\delta_s(\mathbf{r})\,dV = \int_S f(\mathbf{r})\,dS. \tag{45}$$

Using Eq. (43) the right-hand side of Eq. (42) becomes

$$\nabla(1/\eta)\cdot(\eta\mathbf{E}) = \Delta\,(1/\eta)\delta_s(\mathbf{r})\hat{n}\cdot(\eta\mathbf{E}). \tag{46}$$

On and near the boundary Eq. (15) implies that

$$\nabla\cdot(\eta\mathbf{E}) = 0, \tag{47}$$

which in turn implies that $\hat{n}\cdot(\eta\mathbf{E})$ is continuous across the boundary. The right-hand side of Eq. (46) is thus well defined.

By utilizing the continuity of $\hat{n}\cdot(\eta\mathbf{E})$ and the fact that only its value on the boundary is required, Eq. (46) can be written in a form which makes evident the significance of the second term in the effective charge density. Thus, using $\mathbf{E}_1$ and $\mathbf{E}_2$ to denote the fields just inside and outside the boundary, respectively,

$$\begin{aligned}\nabla(1/\eta)\cdot(\eta\mathbf{E}) &= \delta_s(\mathbf{r})\hat{n}\cdot[(1/\eta_2 - 1/\eta_1)(\eta_2\mathbf{E}_2 + \eta_1\mathbf{E}_1)/2] \\ &= \tfrac{1}{2}\delta_s(\mathbf{r})\hat{n}\cdot[\mathbf{E}_2 - \mathbf{E}_1 + (1/\eta_2 - 1/\eta_1)(\eta\mathbf{E})] \\ &= \tfrac{1}{2}\delta_s(\mathbf{r})\hat{n}\cdot\Delta\,\mathbf{E} + \tfrac{1}{2}\nabla(1/\eta)\cdot(\eta\mathbf{E}).\end{aligned}$$

Hence

$$\nabla(1/\eta)\cdot(\eta\mathbf{E}) = \delta_s(\mathbf{r})\hat{n}\cdot\Delta\,\mathbf{E}. \tag{48}$$

From Eq. (5) it is evident that

$$\hat{n}\cdot\Delta\,\mathbf{E} = \Sigma_t/\epsilon_0, \tag{49}$$

where Σ_t is the total surface charge density. Thus the effective charge density can be written in the form

$$\rho_e = [i\omega\epsilon_2/(\sigma_1 + i\omega\epsilon_1)]\rho_s + (\epsilon_2/\epsilon_0)\Sigma_t\delta_s(\mathbf{r}), \tag{50}$$

where ρ_s is the specified charge density defined by

$$\nabla\cdot\mathbf{J}_s = -i\omega\rho_s. \tag{51}$$

The effective current density can be treated in a similar fashion. In the first term of Eq. (20) the permeability can be written

$$\mu = \mu_1 + \Delta\mu\theta_s(\mathbf{r}). \tag{52}$$

Then the second term may be rewritten and treated as follows:

$$\begin{aligned}&-\mu\nabla(1/\mu)\times\mathbf{B} = \nabla\mu\times(\mathbf{B}/\mu), \\ &\nabla\mu\times(\mathbf{B}/\mu) = \Delta\mu\delta_s(\mathbf{r})\hat{n}\times(\mathbf{B}/\mu).\end{aligned} \tag{53}$$

Because $\mathbf{J}_s$ has been assumed nonzero only inside S, Eq. (15) implies that $\hat{n}\times(\mathbf{B}/\mu)$ is continuous across the boundary. The right-hand side of Eq. (53) is thus well defined.

By following a procedure similar to that used to obtain Eq. (48), Eq. (53) can be shown to yield

$$\nabla\mu\times(\mathbf{B}/\mu) = \delta_s(\mathbf{r})\hat{n}\times\Delta\,\mathbf{B}. \tag{54}$$

Equation (14) now implies that

$$\hat{n}\times\Delta\,\mathbf{B} = \mu_0\mathbf{K}_t, \tag{55}$$

where $\mathbf{K}_t$ is the total surface current density. According to Eqs. (10) and (13), $\mathbf{K}_t$ can be thought of as composed of four separate terms: a specified term $\mathbf{K}_s$, an ohmic term $\mathbf{K}_\Omega$, a polarization term $\mathbf{K}_p$, and a magnetization term $\mathbf{K}_m$. However, the specified current density has been assumed zero on and near S so that $\mathbf{K}_s$ vanishes. For finite conductivities and permittivities the ohmic and polarization terms, being proportional to the finite electric field, also vanish. Hence the only remaining term is the magnetization term. This can be verified directly by using Eqs. (8) and (12) to show that

$$\mathbf{K}_m = \hat{n}\times\Delta\,\mathbf{M} = \hat{n}\times\Delta\,\mathbf{B}/\mu_0 \tag{56}$$

and then comparing with Eq. (55). Thus

$$\hat{n}\times\Delta\,\mathbf{B} = \mu_0\mathbf{K}_m \tag{57}$$

and

$$\nabla\mu\times(\mathbf{B}/\mu) = \mu_0\mathbf{K}_m\delta_s(\mathbf{r}). \tag{58}$$

The third term in Eq. (20) for the effective current density presents no special problem. It is convenient, however, to write δk^2 in the form

$$\delta k^2 = (k_1^2 - k_2^2)[1 - \theta_s(\mathbf{r})]. \tag{59}$$

Combining all the various results and simplifying gives the following expression for the effective current density:

$$\begin{aligned}\mathbf{J}_e = {} & (\mu_1/\mu_2)\mathbf{J}_s + (\mu_0/\mu_2)\,\mathbf{K}_m\delta_s(\mathbf{r}) \\ & + [(\mu_1/\mu_2)(\sigma_1 + i\omega\epsilon_1) - (\sigma_2 + i\omega\epsilon_2)]\,[1 - \theta_s(\mathbf{r})]\mathbf{E}.\end{aligned} \tag{60}$$

VI. SUMMARY

It is evident form Eqs. (27) and (36) that it is possible to make a meaningful multipole expansion for charge and current distributions embedded in a material medium provided that the medium can be thought of as being homogeneous at large distances and that the "specified" sources are well localized; i.e. that the system can be thought of as having an asymptotic background. The form of the expansion and the expressions for the coefficients are the same as those for the usual multipole expansion in vacuum except for the use of the constitutive parameters of the asymptotic background (instead of those of the vacuum) and the use of effective source densities in place of the usual macroscopic source densities.

For homogeneous media separated by sharp boundaries the effective charge density has two contributions—one from the specified charge density and one from the total surface charge density—each appropriately modified by the constitutive parameters. The effective current density, on the other hand, contains three contributions—one from the specified current density, another the surface current density arising from magnetization, and one that results from the difference between the actual ohmic and displacement currents and those that would be present if the asymptotic background pervaded all space. It should be noted that all contributions to the effective source densities vanish outside S_0.

Clearly the multipole expansion presented in this paper reduces to the conventional expansion when the exterior region is the vacuum. However, when the exterior region differs from the vacuum, e.g., sources in an ocean environment, the two formalisms are distinct. In fact, in the latter case the conventional formalism fails altogether whereas the present formalism is still applicable.

[1] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952).

[2]J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975).
[3]S. R. deGroot and L. G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).
[4]G. Rusakoff, Am. J. Phys. **38**, 1188 (1970).
[5]F. N. H. Robinson, *Macroscopic Electromagnetism* (Pergamon, New York, 1973).
[6]G. Goertzel and N. Tralli, *Some Mathematical Methods of Physics* (McGraw-Hill, New York, 1960), p. 179. Also see Ref. 2, p. 756.
[7]M. J. Lighthill, *Fourier Analysis and Generalized Functions*, Student Edition (Cambridge U. P., New York, 1964), p. 19.

Energy-particle-number inequality in nonlinear complex-scalar field theory [a)]

Gerald Rosen
Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(Received 3 October 1980; accepted 7 November 1980)

For a self-interaction energy density $U(|\psi|^2)$ that is positive-definite, monotone-increasing with increasing $|\psi|^2$, and concave-saturating, it is shown that the total field energy E and particle number N satisfy the general inequality $E \geqslant$ (function of ξ) $|N|^{(3-\xi)/2(2-\xi)}$ in which the positive parameter $\xi \equiv \min [d \ln U(\rho)/d \ln \rho]$ is less than unity.

PACS numbers: 03.50.Kk

I. INTRODUCTION

In recent years there has been renewed interest in singularity-free finite-energy solutions to nonlinear complex-scalar field theories with Lagrangian densities of the generic form

$$\mathscr{L} = -\partial^\mu \psi^* \partial_\mu \psi - U(|\psi|^2). \tag{1}$$

Particular attention has been given to self-interaction energy densities $U(|\psi|^2)$ that are positive-definite, monotone-increasing with increasing $|\psi|^2$, and concave-saturating

$$U(0) = 0, \quad U(\rho) > \rho U'(\rho) > 0 \quad \text{for all } \rho > 0, \tag{2}$$

and $\rho U^{-1} U'$ nondecreasing with increasing ρ. As originally intimated by the present author[1] and demonstrated by Roffman[2] and Morris,[3] the concavity conditions in (2) are sufficient to guarantee existence of singularity-free localized particlelike solutions to the field equation

$$\ddot{\psi} - \nabla^2 \psi + U'(|\psi|^2)\psi = 0, \tag{3}$$

which follows from (1). Both the (constant) energy

$$E = \int [|\dot{\psi}|^2 + |\nabla \psi|^2 + U(|\psi|^2)]\, d^3x \tag{4}$$

and (constant) particle number

$$N = i \int (\psi^* \dot{\psi} - \dot{\psi}^* \psi)\, d^3x \tag{5}$$

are finite quantities for these particlelike (soliton) solutions.

It is noteworthy that the conditions in (2) imply that

$$1 > \xi \equiv \min [d \ln U(\rho)/d \ln \rho] > 0, \tag{6}$$

where $(1 - \xi)$ is a measure of the concavity manifest in U. The differential inequality $d \ln U(\rho)/d \ln \rho \geqslant \xi$ can be integrated to yield a lower bound on the self-interaction energy density, namely

$$U(\rho) \geqslant m^{4-2\xi} \rho^\xi \quad \text{for all } \rho \geqslant 0, \tag{7}$$

where m (units cm^{-1} with $\hbar = c = 1$) is an additional positive constant parameter associated with U. Assuming that the self-interaction energy density obeys the conditions in (2) and hence admits a lower bound of the form (7), then the constants of the motion (4) and (5) are related by the energy-particle-number inequality derived in the following and exhibited in (15).

II. DERIVATION OF THE ENERGY-PARTICLE-NUMBER INEQUALITY

Consider smooth and localized ψ such that the integrals

$$I_k \equiv \int |\psi|^k\, d^3x \tag{8}$$

exist for $2\xi \leqslant k \leqslant 6$. Then the Schwarz inequality applied to (5) yields

$$|N| \leqslant 2\left(I_2 \int |\dot{\psi}|^2\, d^3x\right)^{1/2}. \tag{9}$$

Hence, by evoking the Sobolev inequality[4-6]

$$\int |\nabla \psi|^2\, d^3x \geqslant 3\left(\frac{\pi}{2}\right)^{4/3} (I_6)^{1/3}, \tag{10}$$

and making use of (7), it follows that the energy (4) has the lower bound

$$E \geqslant \tfrac{1}{4} N^2 (I_2)^{-1} + 3(\pi/2)^{4/3} (I_6)^{1/3} + m^{4-2\xi} I_{2\xi}. \tag{11}$$

In combination with (11), the Hölder inequality for $\xi < 1$

$$(I_2)^{3-\xi} \leqslant (I_{2\xi})^2 (I_6)^{1-\xi}, \tag{12}$$

implies that

$$\begin{aligned} E &\geqslant \tfrac{1}{4} N^2 (I_2)^{-1} + 3(\pi/2)^{4/3} (I_6)^{1/3} \\ &\quad + m^{4-2\xi} (I_2)^{(3-\xi)/2} (I_6)^{-(1-\xi)/2} \\ &\equiv F(I_2, I_6) \geqslant \min F(I_2, I_6). \end{aligned} \tag{13}$$

Thus, by solving the minimization equations

$$\frac{\partial F}{\partial I_2} = 0, \quad \frac{\partial F}{\partial I_6} = 0 \tag{14}$$

algebraically for I_2, I_6 and substituting the latter values into $F(I_2, I_6)$, the final member in (13) yields the general inequality

$$\begin{aligned} E \geqslant (2)^{\frac{1}{2}} &[(1-\xi)^{-3(1-\xi)/4(2-\xi)}] \\ &\times (2-\xi)[(3-\xi)^{-(3-\xi)/4(2-\xi)}] \\ &\times \pi^{(1-\xi)/(2-\xi)} [|N|^{(3-\xi)/2(2-\xi)}] m. \end{aligned} \tag{15}$$

III. DISCUSSION

In view of the sharpness[4,5] of the Sobolev inequality (10), the right side of (15) can be expected to be close to the minimum admissible energy for any prescribed value of $|N|$. The exponent $(3-\xi)/2(2-\xi)$ in (15) gives a power-law dependence on $|N|$ somewhat weaker than linear; this is the hallmark of field-theoretic "binding" in a minimum-energy solution to (3) which involves two or more solitons, presum-

[a)] Work supported by NASA grant NSG 7491.

 022-2488/81/030489-02$1.00

ably in a bound-state configuration. More pronounced binding is generally allowable between a soliton ($N>0$) and an antisoliton ($N<0$), with N closer to zero in the composite solution.

To illuminate the numerical content of (15), consider the special cases of possible physical interest:

$$\begin{aligned}
&\xi\to 0: E\geqslant 2^{3/2}\times 3^{-3/8}\times\pi^{1/2}\,|N|^{3/4}m \\
&\qquad = 3.320\,47\,|N|^{3/4}m \\
&\xi=\tfrac{1}{2}: E\geqslant 2^{1/6}\times 3\times 5^{-5/12}\times\pi^{1/3}\,|N|^{5/6}m \\
&\qquad = 2.522\,16\,|N|^{5/6}m \qquad (16)\\
&\xi=\tfrac{3}{4}: E\geqslant 2^{-3/10}\times 3^{-9/10}\times 5\times\pi^{1/5}\,|N|^{9/10}m \\
&\qquad = 1.899\,69\,|N|^{9/10}m \\
&\xi\to 1: E\geqslant |N|m.
\end{aligned}$$

As in (15), the numerical coefficients that appear in (16) decrease with increasing values of ξ; the $\xi\to 1$ limiting case [fourth entry show in (16)] is patently consistent with linear theory i.e., $U(\rho)=m^2\rho$, box normalization for the plane-wave solutions, and strict superposition.

Although the inequality (15) is derived here in the context of classical field theory, this result also applies to certain nonlinear complex-scalar quantum field theories. For applicability of (15), the Hamiltonian operator H must be normal-ordered to give $\langle H\rangle_{\text{vac}}=0$ and contain suitable counter terms which serve to cancel self-interaction fluctuations and thus engender a lower bound on the (vacuum-based) energy expectation values of the form obtained by combining (4) and (9),

$$E\equiv\langle H\rangle\geqslant\frac{1}{4}N^2\left(\int|\psi|^2\,d^3x\right)^{-1} + \int[|\nabla\psi|^2+U(|\psi|^2)]\,d^3x. \qquad (17)$$

In (17), ψ denotes a disposable parameter- or variational-function in the representative state functional,[7] and the kinetic energy of the field is bounded from below as a consequence of the uncertainty principle, as in particle mechanics.[5,8] The values of N in (17) are restricted to the eigenvalue spectrum $N=0,\ \pm 1,\ \pm 2,\cdots$ of the self-adjoint operator corresponding to (5). For stationary states that feature N quanta in a single dynamical mode of the field, (17) is likely to hold with the equality sign. Thus, for example, the precise energy for linear-theoretic monochromatic "beam" states (with all quanta having the same particle-character and the same wave vector $\mathbf{k}$) is derivable by minimizing the effective energy with box normalization

$$E_{\text{eff}}\equiv\frac{1}{4}N^2\left(\int_V|\psi|^2\,d^3x\right)^{-1}+\int_V(|\nabla\psi|^2+m^2|\psi|^2)\,d^3x, \qquad (18)$$

$$\delta E_{\text{eff}}=0\ \Rightarrow\ E_{\text{eff}}=(|\mathbf{k}|^2+m^2)^{1/2}|N|.$$

The minimization procedure employed in (13) to get (15) is clearly analogous to the minimization of the effective energy in (18).

[1] G. Rosen, J. Math. Phys. **9**, 996 (1968).
[2] E. H. Roffman, Bull. Amer. Math. Soc. **76**, 70 (1970).
[3] T. F. Morris, Phys. Lett. B **76**, 337 (1978).
[4] G. Rosen, SIAM J. Appl. Math. **21**, 30 (1971).
[5] E. H. Lieb, Rev. Mod. Phys. **48**, 553 (1976).
[6] G. Talenti, Ann. Mat. Pura Appl. **110**, 353 (1976).
[7] See, for example, G. Rosen, *Path Integrals and their Applications in Quantum, Statistical, and Solid State Physics*, edited by G. J. Papadopoulos and J. T. Devreese (Plenum, New York, 1978), pp. 201–235.
[8] G. Rosen, Phys. Rev. A **20**, 1287 (1979).

Measurement in stochastic mechanics [a)]

David S. Shucker
Department of Mathematics, University of California Irvine, Irvine, California 92717

(Received 5 December 1979; accepted for publication 8 February 1980)

Stochastic mechanics is an explanation of nonrelativistic quantum phenomena in terms of stochastic differential equations. In this note a simple example of a measurement is constructed and the behavior of the sample paths of the corresponding stochastic differential equation is examined. The sample demonstrates that stochastic mechanics provides a natural explanation of the "reduction of the wave packet."

PACS numbers: 03.65.Bz

I. BASIC THEORY

Let us begin by considering the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = \tfrac{1}{2}\sum_{k=1}^{n}\frac{\partial^2\psi}{\partial x_k^2} + V(t)\psi = -\tfrac{1}{2}\nabla^2\psi + V(t)\psi. \tag{1}$$

Here $\psi = \psi(x_1,x_2,...,x_n;t) = \psi(x;t)$ is the wavefunction; $V(t)$ is the operation of multiplication by the real valued function $V(x_1,x_2,..,x_n;t)$; and ∇ is the vector operator

$$\left(\frac{\partial}{\partial x_1},\frac{\partial}{\partial x_2},...,\frac{\partial}{\partial x_n}\right).$$

We have set $\hbar$ and all masses equal to 1 for simplicity.

Let us write $\psi = \exp(R + iS)$ and define the vector valued function

$$b^* = \nabla R + \nabla S. \tag{2}$$

We assume ψ is sufficiently smooth for this to be done. Note that b^* is the sum of the real part and the imaginary part (without the $\sqrt{-1}$) of the logarithmic gradient $\nabla\psi/\psi$.

Then it is easy to verify that $\rho(x;t) = |\psi(x;t)|^2$ which is interpreted as a probability density in quantum mechanics, satisfies

$$\frac{\partial\rho}{\partial t} = -\nabla\cdot(b^*\rho) + \tfrac{1}{2}\nabla^2\rho, \tag{3}$$

and is thus the probability density of a solution to the stochastic differential equation

$$dx(t) = b^*(x;t)\,dt + dw(t). \tag{4}$$

Here $w(t)$ is the Wiener process, normalized so that the variance of $w(s_2) - w(s_1)$ is $n|s_2 - s_1|$, i.e., the probability density of $w(s_2) - w(s_1)$ is

$$(2\pi|s_2 - s_1|)^{-n/2}\exp\left(-\frac{x^2}{|s_2 - s_1|}\right).$$

We refer to b^* as the mean forward velocity of this process.

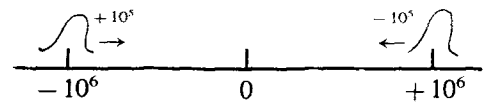
In stochastic mechanics we consider a quantum mechanical system to have a definite, though unknown, position at each time. The time evolution of this position is governed by the probabilistic law (4).

II. AN EXAMPLE OF A MEASUREMENT

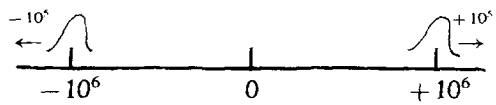
Consider a free particle of mass 1 with the following (un-normalized) wavefunction at time 0:

$$\psi(x;0) = \exp[10^5 ix - (x + 10^6)^2] + \exp[-10^5 ix - (x - 10^6)^2]. \tag{5}$$

The wavefunction is a superposition of two Gaussian wave packets centered at ∓ 10 and moving with velocities ± 10 (the widths of the wave packets are not drawn to scale):



At $t = 20$, the two wave packets have widened somewhat and exchanged positions:



It is interesting that while it appears that the two wave packets passed through each other, in the corresponding stochastic process the particle reverses direction with probability greater than $\frac{1}{2}$.

This is seen by an easy symmetry argument. Note that $\psi(x;0) = \psi(-x;0)$ and hence $\psi(x;t) = \psi(-x;t)$ for all t and all x. It follows that $b^*(x;t) = -b^*(-x;t)$ where b^* is the mean forward velocity of the corresponding stochastic process. Therefore, if the particle is at the origin at time t', it is equally likely to be at x or $-x$ at time $t > t'$.

Almost all sample paths of the process are continuous. After throwing out the discontinuous paths, we divide the paths into two sets:

(1) Those paths that pass through the origin for some t, $0 \leqslant t \leqslant 20$. Let μ denote the measure of this set, $0 < \mu < 1$.

(2) Those paths that don't pass through the origin for $0 \leqslant t \leqslant 20$. The measure of this set is $1 - \mu$.

The probability that the particle reverses direction (i.e., that the positions of the particle at times 0 and 20 have the same sign) is $\frac{1}{2}\mu + 1 - \mu > \frac{1}{2}$ because half the paths in the first set and all the paths in the second set have this property.

This result seems to contradict classical physics and our experience, for a free particle has never been observed to exhibit such strange behavior. But in stochastic mechanics,

[a)]This work was partially supported by an NSF graduate fellowship and by NSF grant MCS-7906633.

 022-2488/81/030491-04$1.00

if the particle is observed (i.e., its position measured) then this strange behavior disappears. Specifically, we consider a measurement by interaction with a second particle, also of mass 1, with the (un-normalized) wavefunction at time 0:

$$\psi(y;0) = \exp(-y^2). \tag{6}$$

The two-particle system has wavefunction:

$$\psi(x,y;0) = \exp(-y^2)\{\exp[10^5 ix - (x+10^6)^2] + \exp[-10^5 ix - (x-10^6)^2]\}. \tag{7}$$

If there were no interaction between the two particles (i.e., zero potential), then no measurement would be made and each particle would move as if the other didn't exist. In fact, the wavefunction would factor at all times t as

$$\psi(x,y;t) = \psi_1(x;t)\psi_2(y;t), \tag{8}$$

where ψ_1 and ψ_2 are solutions of the free one-dimensional Schrödinger equation, and the mean forward velocity, now a 2-vector, would be

$$\begin{aligned} b^+(x,y;t) &= \left([\mathrm{Re}+\mathrm{Im}]\frac{\partial}{\partial x}\log\psi(x,y;t),\right. \\ &\qquad \left.[\mathrm{Re}+\mathrm{Im}]\frac{\partial}{\partial y}\log\psi(x,y;t)\right) \\ &= \left([\mathrm{Re}+\mathrm{Im}]\frac{\partial}{\partial x}\log\psi_1(x;t),\right. \\ &\qquad \left.[\mathrm{Re}+\mathrm{Im}]\frac{\partial}{\partial y}\log\psi_2(y;t)\right). \end{aligned} \tag{9}$$

Here [Re + Im] is the sum of the real and imaginary parts: $[\mathrm{Re}+\mathrm{Im}](a+ib) = a+b$. Thus the mean forward velocity in the x direction is independent of the y coordinate and is the same as the mean forward velocity of the 1-particle system given by (5), and the mean forward velocity in the y direction is similarly the same as in the 1-particle system given by (6).

Now let the particles interact with the potential

$$V(x,y;t) = \frac{1}{20\alpha}|x-y|^2\chi_{(0,\alpha)}(t) \tag{10}$$

in the limit $\alpha \to 0_+$. The function $\chi_{(0,\alpha)}$ is the characteristic function of the interval $(0,\alpha)$. The effect of this potential is approximately to give the particles a velocity increment of 10^3 towards each other, since the particles are separated by a distance of approximately 10^6. This is only a one percent change in the momentum of the first particle.

To solve explicitly, we solve separately for

$$\exp(-y^2)\exp[10^5 ix - (x+10^6)^2], \tag{11}$$

$$\exp(-y^2)\exp[-10^5 ix - (x-10^6)^2], \tag{12}$$

and then add the two solutions. The solutions are effected by the change of variables $R = (x+y)/2$, $r = x-y$. The wavefunctions are then Gaussian wave functions in quadratic central potentials.

The wavefunction (11) is, omitting a constant factor,

$$\exp[-2R^2 + (10^5 i - 2\times 10^6)R] \times \exp\left(\frac{-r^2}{2} + \frac{10^5 i - 2\times 10^6}{2}r\right). \tag{13}$$

At time $t = 0_+$ (after the impulse) the wavefunction is

$$\exp[-2R^2 + (10^5 i - 2\times 10^6)R] \times \exp\left(\frac{-1+10^{-3}i}{2}r^2 + \frac{10^5 i - 2\times 10^6}{2}r\right). \tag{14}$$

At $t > 0$, the wavefunction is

$$\begin{aligned} &\exp\left\{(-\tfrac{1}{2}-it)^{-1}R^2 + (10^5 i - 2\times 10^6)(1+2it)^{-1}R \right. \\ &\quad + \left[-\left(\frac{1+10^{-3}i}{2}\right)^{-1} - 4it\right]^{-1} r^2 \\ &\quad \left. + \left(\frac{10^5 i - 2\times 10^6}{2}\right)[1+2it(1+10^{-3}i)]^{-1} r\right\} \\ &= \exp\left(\tfrac{1}{4}(-\tfrac{1}{2}-it)^{-1} + \left[-\left(\frac{1+10^{-3}i}{2} - 4it\right)^{-1}\right]\right. \\ &\quad \times (x^2+y^2) + \left\{\tfrac{1}{2}(-\tfrac{1}{2}-it)^{-1} - 2\right. \\ &\quad \left.\left.\times\left[-\left(\frac{1+10^{-3}}{2}\right)^{-1} - 4it\right]^{-1}\right\} xy \right. \\ &\quad + \frac{10^5 i - 2\times 10^6}{2}\{(1+2it)^{-1} \\ &\quad + [1+2it(1+10^{-3}i)]^{-1}\} x \\ &\quad + \frac{10^5 i - 2\times 10^6}{2}\{(1+2it)^{-1} \\ &\quad \left. - [1+2it(1+10^{-3}i)]^{-1}\} y\right). \end{aligned} \tag{15}$$

To this must be added the solution for the other half of the wavefunction (12). Omitting the same constant factor, this is

$$\begin{aligned} &\exp\left(\left\{\tfrac{1}{4}(-\tfrac{1}{2}-it)^{-1} + \left[-\left(\frac{1+10^{-3}i}{2}\right)^{-1} - 4it\right]^{-1}\right\}(x^2+\right. \\ &+ \left\{\tfrac{1}{2}(-\tfrac{1}{2}-it)^{-1} - 2\left[-\left(\frac{1+10^{-3}i}{2}\right)^{-1} - 4it\right]^{-1}\right\} xy \\ &- \frac{10^5 i - 2\times 10^6}{2}\{(1+2it)^{-1} + [1+2it(1+10^{-3}i)]^{-1}\} x \\ &\left. - \frac{10^5 i - 2\times 10^6}{2}\{(1+2it)^{-1} - [1+2it(1+10^{-3}i)]^{-1}\} y\right). \end{aligned} \tag{16}$$

The wavefunction is the sum of (15), which is a wave packet centered at $(-10^6 + 10^5 t + 10^3 t, -10^3 t)$ and moving with velocity $(10^5 + 10^3, -10^3)$, and (16), which is a wave packet centered at $(10^6 - 10^5 t - 10^3 t, 10^3 t)$ and moving with velocity $(-10^5 - 10^3, +10^3)$.

Let us analyze the first wave packet with the change of coordinates

$$\hat{R} = \tfrac{1}{2}(x+y+10^6+10^5 t), \tag{17a}$$

$$\hat{r} = x - y + 10^6 - 10^5 t - 2\times 10^3 t. \tag{17b}$$

In this coordinate system the first wave packet is centered at (0,0). The mean forward velocity in this coordinate system is

$$b^+(\hat{R},\hat{r};t) = \left\{\frac{4t-2}{1+4t^2}\hat{R}, \frac{-2(1+10^{-3}) + 4(1+10^{-6}t)}{1-4\times 10^{-3}t + 4(1+10^{-6})t^2}\hat{r}\right\}, \tag{18}$$

if we ignore interference from the second wave packet.

From the original wavefunction (7) and the change of

variables (17), we see easily with the use of a table of the error function that, if the x coordinate of the system is negative at time 0, then the $\hat{R}$ coordinate at time 0 has absolute value less than 1.5 with probability $\geqslant 0.9999$. Also, using a table of the error function and the following inequality for the Wiener process w:

$$\text{Prob.}\{w:\exists t,s \text{ such that } 0\leqslant t,s\leqslant\delta \text{ and } |w(t)-w(s)|>2\epsilon\}$$
$$\leqslant 2\,\text{Prob.}\left\{|w(\delta)-w(0)|>\frac{\epsilon}{2}\right\}, \tag{19}$$

we find that the Wiener process contribution to $\hat{R}$ is, with probability at least $(0.9998)^3$, less than:

3 for $0\leqslant t\leqslant\frac{1}{2}$,

9 for $\frac{1}{2}\leqslant t<5$,

16 for $5\leqslant t\leqslant 20$.

[The inequality (19) is esssentially Lemma 3 in Appendix A of [5]].

Let us now assume that the $\hat{R}$ coordinate is $\leqslant 1.5$ in absolute value at time 0 and that the Wiener process contribution to $\hat{R}$ is less than 3 for $0\leqslant t\leqslant\frac{1}{2}$, less than 9 for $\frac{1}{2}\leqslant t\leqslant 5$, and less than 16 for $5\leqslant t\leqslant 20$. Let us further assume that the total contribution to $\hat{R}$ from the perturbation of mean forward velocity caused by interference with the second wave packet is less than $\frac{1}{2}$ for $0\leqslant t\leqslant 20$. This assumption will be checked later.

Now for $0\leqslant t\leqslant\frac{1}{2}$, the mean forward velocity $[(4t-2)/(1+4t^2)]\hat{R}$ tends to decrease $|\hat{R}|$, while for $t>\frac{1}{2}$ it tends to increase $|\hat{R}|$ at a rate proportional to $|\hat{R}|$. The worst possible (i.e., largest) values for $|\hat{R}|$ under the above assumptions are

$$1.5+0.5+3=5 \qquad \text{for } t\leqslant\tfrac{1}{2},$$
$$(5+9)\exp\left(\int_{1/2}^{5}\frac{4t-2}{1+4t^2}\,dt\right)<82 \qquad \text{for } t\leqslant 5,$$
$$(82+16)\exp\left(\int_{5}^{20}\frac{4t-2}{1+4t^2}\,dt\right)<400 \qquad \text{for } t\leqslant 20.$$

Thus, assuming that the x coordinate of position is negative at time 0, and assuming that the total contribution to $\hat{R}$ from the perturbing effect of the second wave packet is less than $\frac{1}{2}$, the probability is greater than $(0.9999)(0.9998)^3$ that $|\hat{R}|<400$ for $t\leqslant 20$. Similarly, if we assume that the perturbing effect of the second wave packet contributes less than 1 to $\hat{r}$, then the probability is greater than $(0.9999)(0.9998)^3$ that $|\hat{r}|<800$ for $t\leqslant 20$. Thus the probability is greater than $(0.9999)^2(0.9998)^6>0.996$ that $|\hat{R}|<400$ and $|\hat{r}|<800$ for $0\leqslant t\leqslant 20$. To confirm this result, we need only check the assumption that the perturbation in mean forward velocity caused by the second wave packet contributes less than $\frac{1}{2}$ in the $\hat{R}$ coordinate and less than 1 in the $\hat{r}$ coordinate.

Both wave packets are defined as exponentials, so we must find the perturbation

$$[\mathrm{Re}+\mathrm{Im}]\{\nabla\log[\exp(a_1)+\exp(a_2)]-\nabla\log[\exp(a_1)]\}$$
$$=[\mathrm{Re}+\mathrm{Im}]\{\nabla\log[1+\exp(a_2-a_1)]\}$$
$$=[\mathrm{Re}+\mathrm{Im}]\{\nabla[\exp(a_2-a_1)-\tfrac{1}{2}\exp(2a_2-2a_1)$$
$$+\tfrac{1}{3}\exp(3a_2-3a_1)\cdots]\}$$
when $\mathrm{Re}(a_1)>\mathrm{Re}(a_2)$
$$=[\mathrm{Re}+\mathrm{Im}][(\exp(a_2-a_1)-\exp(2a_2-2a_1)$$
$$+\exp(3a_2-3a_1)\cdots)(\nabla(a_2-a_1))]$$
$$[\mathrm{Re}+\mathrm{Im}]\left[\frac{\exp(a_2-a_1)}{1+\exp(a_2-a_1)}\nabla(a_2-a_1)\right]. \tag{20}$$

In the case at hand, a_1 is the exponent in (15) and a_2 is the exponent in (16). We are working in the region $|\hat{R}|<400$ and $|\hat{r}|<800$, which means that in the x–y plane, the distance from the center $(-10^6+10^5t+10^3t, 10^3t)$ of the wave packet (15) is at most $800\sqrt{2}<1200$. Now the closest approach of the centers of the wave packets (15) and (16) is approximately a distance of 19,800. Trivial estimates show that $\mathrm{Re}(a_2-a_1)\ll -100$, and also that $|\nabla(a_2-a_1)|<10^7$. Thus the perturbation (20) is indeed negligible.

What we have shown is, that if the x coordinate of the system is negative at time 0, then the probability is greater than 0.996 that the position of the system lies within a distance of 1200 in the x–y plane from $(-10^6t+10^5t+10^3t, -10^3t)$ for $0\leqslant t\leqslant 20$. In particular, the probability is greater than 0.996 that the first particle does not reverse direction. The measure on paths looks very much like that attributable to the "reduced" wave packet (15) because the perturbing effect of (16) is, with probability greater than 0.996, less than $\exp(-100)$.

III. ANALYSIS OF THE EXAMPLE

In stochastic mechanics, the dynamics of a system is given by a stochastic differential equation and is thus a Markov process. The system has no "memory" of the past. It is this property which is responsible for the strange behavior of the sample paths of the stochastic differential equation corresponding to the single particle system considered in the previous section. The particle, upon reaching the origin, has no memory of which direction it came from, and is just as likely to reverse direction as to continue in the same direction.

In a two particle system, it is the position of the two particles taken together that constitute a Markov process. It does not follow, and indeed is generally not true, that the position of one particle constitutes a Markov process. (A sufficient condition for a position of each particle to constitute separately a Markov process is that the corresponding wave function factor as in Eq. (8) at all times.) The position of one particle is not a Markov process because the "memory" of its past is contained in the present position of the second particle.

In the two particle system presented earlier, for example, the first particle is, with high probability, near either -10^6 or $+10^6$ at time 0. At time $10^6/(10^5+10^3)$, the particle is with high probability near the origin. However, if this particle was near -10^6 at time 0, then the second particle (the "measuring device") is near -10^4 [or more precisely, $-10^3\times 10^6/(10^5+10^3)$] at time 10, and if the first particle was near $+10^6$ at time 0, then our "measuring device" is near $+10^4$ at time 10. The mean forward velocity of the particle depends not only on its own present position, but also on the present position of the "measuring device,"

which in turn depends on the *past* position of the measured particle.

Furthermore, in the stochastic process corresponding to the wavefunction given by (15) and (16), we saw that, if the measured particle were to the left of the origin at time zero, then, to a very good approximation, its future would evolve in accordance with the stochastic process corresponding to the wavefunction (15). Thus, to compute the probability density in the future, conditioned upon the knowledge that the particle was to the left of the origin at time zero (which may be ascertained by our measurement), we need only to consider the "reduced" wave packet (15) instead of the full wavefunction (15) plus (16). The perturbing influence of the wave packet (16) still exists; equation (20) and the discussion following show however that this influence is negligible. It is negligible because we have made a "macroscopic" measurement. At time $t = 0$, the measuring device (the second particle) is located near $+ 10^4$ or $- 10^4$; it is precisely this macroscopic difference between $+ 10^4$ and $- 10^4$ that accounts for the negligible perturbing influence of the wave packets upon each other.

The reduction of the wave packet when a macroscopic measurement is made has been the subject of much discussion and controversy among physicists. Schrödinger himself was unhappy with the situation, and criticized the theory he helped found.[6] It is intriguing that something so difficult to explain in quantum mechanics admits so straightforward and natural an explanation in stochastic mechanics.

ACKNOWLEDGMENT

The author wishes to thank Edward Nelson for his advice and encouragement.

[1]T. G. Dankel, Jr., Arch. Ration. Mech. Anal. **37**, 192 (1970).
[2]I. Fényes, Z. Phys. **132**, 81 (1952).
[3]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[4]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, New Jersey, 1967).
[5]E. Nelson, J. Math. Phys. **5**, 332 (1964).
[6]E. Shcrödinger, "Die Gegenwärtige Situation in der Quanten-mechanic," Naturwiss. **28**, 807, 823, 844 (1935).
[7]D. Shucker, "Stochastic Mechanics of Systems with Zero Potential," to appear, J. of Funct. Anal.

Quantum mechanical Hamiltonian models of discrete processes

Paul Benioff
Institute for the History and Philosophy of Science, Tel Aviv University, Tel Aviv, Israel
Division of Environmental Impact Studies, Argonne National Laboratory, Argonne, Illinois 60439[a]

(Received 29 April 1980; accepted for publication 19 September 1980)

Here the results of other work on quantum mechanical Hamiltonian models of Turing machines are extended to include any discrete process T on a countably infinite set A. The models are constructed here by use of scattering phase shifts from successive scatterers to turn on successive step interactions. Also a locality requirement is imposed. The construction is done by first associating with each process T a model quantum system M with associated Hilbert space $\mathscr{H}_M$ and step operator U_T. Since U_T is not unitary in general, M, $\mathscr{H}_M$, and U_T are extended into a (continuous time) Hamiltonian model on a larger space which satisfies the locality requirement. The construction is compared with the minimal unitary dilation of U_T. It is seen that the model constructed here is larger than the minimal one. However, the minimal one does not satisfy the locality requirement.

PACS numbers: 03.65.Bz, 03.65.Nk

I. INTRODUCTION

In recent work,[1] hereafter referred to as I, quantum mechanical Hamiltonian models of Turing machines were constructed. The constructions were such that the model systems depended both on which machine was being considered and on the number of computation steps for which the model was valid. Also, the Coleman approximation[2] which makes the kinetic energy linear in the momentum was used. The construction also depended on the fact that each Turing machine corresponds to a step transfer function on a countably infinite set A of instantaneous descriptions or overall machine states.[3] Although T could be many-one and into, the fact that it was restricted to (standard) Turing machines allowed for some simplifications in the model construction.

Now most maps T from A to A do not correspond to Turing machines, thus it is natural to ask if the work of I can be generalized both by removal of the Coleman approximation and to apply to any abstract discrete process with step transfer function $T{:}A\rightarrow A$. From now on A denotes an arbitrary countably infinite set of process-system states or descriptions.

The purpose of this paper is to show that such a generalization is possible. In particular, it will be seen that for any abstract discrete process T on A where T can be many-one and into, one can construct a quantum-mechanical Hamiltonian model of the process. The construction is carried out in much the same spirit as is that applied to various models of the measurement process.[2,4,5]

In essence, what is done here is to first associate to each process T, a model process described by a discrete semigroup of operators on a Hilbert space $\mathscr{H}_M$. The model is then extended and expanded into a larger system whose continuous time evolution is described by a Hamiltonian. In the extended model, the discrete stepwise nature of the process modeled is reflected in the fact that the part of the model state which corresponds to the process system is essentially stationary for a finite time interval when it describes a completed model-process step. The importance of this stationarity in models of the measurement process has been emphasized by Emch.[4]

Section II discusses two general features of the model construction, the use of scattering phase shifts to turn on and off successive step interactions in the model, and the requirement of locality. In essence, this requirement says that if the model system is in a state corresponding to a process state Φ_j at stage j, then as the system evolves towards a state Φ_{j+1} corresponding to the process state at stage $j+1$, the system state must have nonzero components only in the subspace spanned by Φ_j and Φ_{j+1}. In particular, it must not include states which the model process will not reach until several steps in the future or which the model process has already passed through several steps previously.

Section III discusses in detail the use of scatterings from fixed centers to turn on successive model-process interactions. The scattering phase shifts which turn on the successive model-process interactions are calculated in the eikonal approximation.[6] One-dimensional scattering is assumed. Inequalities which the various model parameters must satisfy are discussed.

Section IV describes a model extension which is sufficient for the inclusion of processes whose step functions T can be many-one and into, and which satisfies the locality requirement. Each step of the original process is expanded into three successive steps, record, T process, then shift in the expanded model. In essence, as the extended-expanded model evolves, it generates a history of its evolution. Explicit expressions in terms of simple operators are given for the interaction operators for each of the types of steps and are used to generate an overall model Hamiltonian H_{3n}^{T} for describing $3n$ model steps. The Schrödinger evolution of the model is discussed. It is seen that the model does in fact do what is claimed, namely, describe the expanded process and satisfy the locality requirement.

The dilation theorem of Sz. Nagy[7] as applied to the model processes is discussed in Sec. V. The minimal unitary

[a] Present address.

 022-2488/81/030495-13$1.00

dilation of a model process operator U_T is constructed. The resultant model describes successive steps of the model T process, however the locality requirement is not satisfied. Finally, Sec. VI discusses some further aspects of the constructions of this paper.

II. GENERAL CONSIDERATIONS

A. Phase shifts as interaction parameters

The models will be constructed here using the following idea. Consider a discrete reversible process on a countable set A which is represented by a bijection $T:A\to A$. That is, given some a in A as the overall state or description of the process at stage j, $T(a)$ gives the state at stage $j+1$ and $T^l(a)$, the state at stage $j+l$. Since T is a bijection l can take any value in Z, the set of integers. The extension to processes for which T is not a bijection will be done later in Sec. IV.

Let $\mathscr{H}_M$ be a separable Hilbert space of a system M and let $a\mapsto\psi_a^M$ be a one-one correspondence of A onto an orthonormal basis set $\{\psi_a^M|a\in A\}$ which spans $\mathscr{H}_M$. Then one can define an operator U_T in $B(\mathscr{H}_M)$, the set of all bounded linear operators over $\mathscr{H}_M$ by

$$U_T\psi_a^M=\psi_{T(a)}^M \tag{1}$$

for all a in A. Since T is a bijection, U_T is unitary. Let H' be some self-adjoint operator which satisfies

$$U_T=e^{iKH'}, \tag{2}$$

where K is a real number.

It is desired now to construct a Hamiltonian system which starts with M in state ψ_a^M and ends up and remains with M in state $\psi_{T(a)}^M$. This is done here by adding another system w and a fixed center c and using the scattering of w off the fixed scattering center c to turn on the interaction H'. The incident momentum of w and the w-c interaction potential strength and range are adjusted so that the total scattering phase shift given by the eikonal approximation[6] equals K.

Here and in what follows the term "M system state" refers to those parts of the overall $M+w$ system state which lie in $\mathscr{H}_M$. It will always be clear to which component in any expression the discussion is referring. Of course, the real M system state can only be obtained as a trace of the $M+w$ system state over the system w Hilbert space.

Successive steps are included by replacing c by a sequence C of widely spaced fixed scattering centers. The spacing between the centers and the range of the interaction between w and each system c of C are such that the scattering from the jth system c_j of C is essentially complete before w moves into the range of the $j+1$st scatterer in C. Since w is described by a moving wave packet, one has the following picture: if the M system is started in state ψ_a^M, it will evolve to $\psi_{T(a)}^M$ while the w wave packet is within the interaction range of c_1. The M system state will remain (essentially) stationary in state $\psi_{T(a)}^M$ during those times for which the w wave packet is (essentially) outside the interaction range of c_1 and c_2. The M system state evolves away from $\psi_{T(a)}^M$ and towards $\psi_{T(a)}^{M_2}$ as the w wave packet begins interacting with c_2, and so on. Fig. 1 shows M, C, and w at a stage when the w system wave

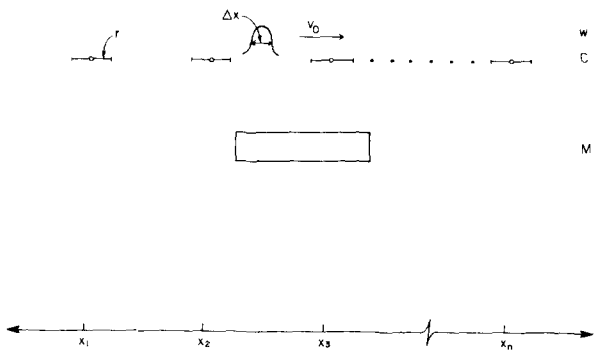


FIG. 1. A schematic representation of the model. The small circles represent the scattering centers of C at fixed positions $x_1,x_2\ldots x_n$. r denotes the w-c interaction range. The w wave packet of width Δx is shown between two centers of C and moving to the right with velocity v_0. The form of M is arbitrarily denoted by a rectangle.

packet is essentially between scatterers.

To keep the model as simple as possible without losing essential details, the w-C scattering will be described in one dimension only. Extension to three dimensions is straightforward.

B. Localization and extension

The construction just described can be represented as follows. Let $f(x,v_0,t)$ represent the w-wave packet amplitude with group velocity v_0 at position x at time t. Assume that $f(x,v_0,0)$ is peaked at x_0 which is distant from the centers of C, and that the spread Δx does not change during the time interval of interest. The centers in C are sufficiently widely spaced so that at any time w interacts with at most one center of C. For each position x let $\delta(x)$ denote the accumulated w-C scattering phase shift. (The dependence of $\delta(x)$ on the momentum of w is ignored in this discussion as it is small in the eikonal approximation.) Then if M is initially in state ψ_a^M at time 0, the overall system state at time t for w at position x is given approximately by

$$\psi_a^{M+w}(x,t)\approx f(x,v_0,t)\psi_a^M(x), \tag{3}$$

where $\psi_a^M(x)=\exp[i\delta(x)H']\psi_a^M$. If H' is such that Eq. (2) is satisfied, one has

$$\psi_a^M(x)\approx\sum_{a'\in A}d_{a'}^{a}(\delta(x))\psi_{a'}^M, \tag{4}$$

where the coefficients $d_{a'}^{a}(\delta(x))$ are in general different from 0 if $\delta(x)\neq jK$ for all integers j. If $\delta(x)=jK$ then $d_{a'}^{a}(jK)=1$ if $a'=T^j(a)$ and 0 otherwise.

Now the model parameters can be adjusted so that there exists a sequence of nonoverlapping time intervals of width Δ centered about the times $t_0,t_1,\ldots,t_n$ such that for all t in the interval $[t_j-\Delta/2,\ t_j+\Delta/2]$, $\delta(x)=jK$ for all x for which $f(x,v_0,t)$ is appreciable. Thus for these times, one has

$$\psi_a^{M+w}(x,t)\approx f(x,v_0,t)\psi_{T^j(a)}^M.$$

However, the description given is not "local" in the following sense: Let t lie in an interval $[t_j+\Delta/2,\ t_{j+1}-\Delta/2]$ for some $j<n$. Then $f(x,v_0,t)$ is appreciable for values of x at which w interacts with center c_{j+1} of C. For such x, $jK<\delta(x)<(j+1)K$ and $\psi_a^M(x)$ is sup-

posed to describe M in transition from $\psi^M_{T^j(a)}$ to $\psi^M_{T^{j+1}(a)}$. For any such x, e.g., $x = v_0 t + x_0$, Eq. (4) shows that $\psi^M_a(x)$ evolves from a state which is localized at $\psi^M_{T^j(a)}$ when $[t_j - \Delta/2 \leqslant t \leqslant t_j + \Delta/2]$ to an "unlocalized state" [that is $d^a_{a'}(\delta(x)) \neq 0$ for many a' for values of x for which $f(x,v_0,t)$ is appreciable]. As t increases and w completes interaction with c_{j+1}, $\delta(x) \simeq (j+1)K$ for all x for which $f(x,v_0,t)$ is appreciable, and the unlocalized state $\psi^M_a(x)$ relocalizes at $\psi^M_{T^{j+1}(a)}$ as all other components in the sum of Eq. (4) disappear.

This is unsatisfactory because it means that if one were to interfere with the model process at such times by measurement to determine what state the M system was in, one might find that it was in some state $\psi^M_{a'}$ which the undisturbed model process will not reach until many steps in the future, or if $j > 1$, one might find on measurement that M is in a state in which it has already been several steps earlier.

One would like the model to be such that interference by measurement does not destroy the model representation to this extent. It would be more satisfactory if measurements made during times in which w was interacting with c_{j+1} gave the result that M was either in state $\psi^M_{T^j(a)}$ or in state $\psi^M_{T^{j+1}(a)}$ and no other.

More generally, for all times t for which w is interacting with c_{j+1} with $0 \leqslant j < n$, one requires that for all x for which $f(x,v_0,t)$ is appreciable,[8]

$$\psi^M_a(x) = \gamma(\delta(x))\psi^M_{T^j(a)} + \lambda(\delta(x))\psi^M_{T^{j+1}(a)} \tag{5}$$

for each a in A for which $T^j(a) \neq T^{j+1}(a)$. Here $\gamma(\delta(x))$ and $\lambda(\delta(x))$ are a-independent coefficients such that $\gamma(jK) = \lambda((j+1)K) = 1$, $\gamma((j+1)K) = \lambda(jK) = 0$, and $\gamma(\delta(x)) \neq 0 \neq \lambda(\delta(x))$ if $jK < \delta(x) < (j+1)K$. Also $|\gamma(\delta(x))|^2 + |\lambda(\delta(x))|^2 = 1$ is required as a normalization condition.

The problem with a model which satisfies a locality condition such as that given by Eq. (5) is that any operator V_ϵ with $jK < \epsilon < (j+1)K$ for $0 \leqslant j < n-1$, which for each a in A satisfies

$$V_\epsilon \psi^M_a = \gamma(\epsilon)\psi^M_{T^j(a)} + \lambda(\epsilon)\psi^M_{T^{j+1}(a)}$$

is not in general unitary. [To see this note that $(\psi^M_{T(a)}, V^\dagger_\epsilon V_\epsilon \psi^M_a) \neq 0$ in general]. Thus it would appear that one cannot both keep the locality requirement and construct a Hamiltonian model of the $w + C + M$ interaction as outlined above.

It turns out, though, on closer inspection that if T is such that $T^2 = 1$ then V_ϵ is unitary provided that $\lambda^*(\epsilon)\gamma(\epsilon) + \lambda(\epsilon)\gamma^*(\epsilon) = 0$. This result can be obtained from Eq. (4) by noting that, in general, only those coefficients $d^a_{a'}(\delta(x))$ are different from 0 for which a' lies in the T-invariant irreducible subset of A which contains a. (This is discussed in more detail in Sec. V and the Appendix.) $T^2 = 1$ if and only if all such subsets contain at most two elements.

This result suggests that one might satisfy the locality requirement by adding auxiliary systems to M and thereby extending A to $A \times X$, where X is the set of states of the auxiliary systems. Then if necessary, one expands T by constructing one or more types of processes T_i on $A \times X$ such that for each i, $T_i^2 = 1$, T_i is a bijection, and the action of $T_m(T_{m-1}(\cdots(T_1(-)\cdots))$ on selected elements of $A \times X$ corresponds to the action of T on A. It will be seen in Sec. IV, that this is indeed possible for $m = 3$ and furthermore that in such an extension, T can be many-one and into.

III. USE OF SCATTERINGS TO TURN ON MODEL INTERACTIONS

A. One scattering

To investigate scattering from one fixed center C is replaced by one scattering center c fixed at position x_c. The $w + c + M$ system Hamiltonian is given by

$$H = H_0 + V(x - x_c)H', \tag{6}$$

where H_0 is the free Hamiltonian. $V(x - x_c)$ is the interaction potential between w and c with w at position x and is such that $\int_{-\infty}^{\infty} V(y)\,dy \simeq \int_{-r}^{r} V(y)\,dy$ is finite. r is the range of V.

A restriction on the model systems M considered in this paper is now imposed. This is that they be such that all interactions between the component particles of M are included in H'. Such models include noninteracting lattice systems of particles with or without spin. As a result, the free Hamiltonian H_0 includes a term for the kinetic energy of w only or

$$H_0(x) = -\hbar^2\partial^2/2m\partial x^2, \tag{7}$$

where m is the mass of w.

The initial state $\psi(0)$ of the $w + M$ system at time $t = 0$ is taken to be a wave packet for w which is far removed from c, and an initial state ψ^M_a for M. That is

$$\psi(x,0) = \psi^w(x,0)\psi^M_a, \tag{8}$$

where

$$\psi^w(x,0) = \int dk\, \phi(k - k_0)e^{ik(x - x_0)} \tag{9}$$

and $\phi(k - k_0)$ is a normalized momentum space wave packet centered about k_0. $\psi^w(x,0)$ is a coordinate space wave packet centered at $x = x_0$ at time 0.

Let $\Delta x(0)$ be the x space spread of $\psi^w(x,0)$ at time 0. Then one requires that at time 0 essentially no scattering has occurred or that the wave packet be essentially outside the range of interaction with c at x_c. This gives the requirement on x_0 that

$$x_0 \ll x_c - r - \Delta x(0)/2. \tag{10}$$

At any time t, the overall state $\psi(x,t)$ is given by[9]

$$\psi(x,t) = \int dk\, e^{-iE_k t/\hbar}\phi(k - k_0)\psi_{+k}(x)e^{-ikx_0}. \tag{11}$$

Here $E_k = \hbar^2 k^2/2m$ is the total energy of the system and $\psi_{+k}(x)$ is the scattering solution of the Schrödinger Hamiltonian of Eq. (6) with incoming plane waves $\exp(ikx)$ for w and M in state ψ^M_a. From now on we set $x_c = 0$.

$\psi_{+k}(x)$ is given by the Lippman–Schwinger equation as

$$\psi_{+k}(x) = e^{ikx}\psi^M_a + (2m/\hbar^2)\int_{-\infty}^{\infty} G_{+k}(x,x')V(x')H'\psi_{+k}(x')\,dx', \tag{12}$$

where the one-dimensional outgoing Green's function $G_{+k}(x,x')$ is given by

$$G_{+k}(x,x') = (i/2k)e^{ik|x-x'|}. \tag{13}$$

To obtain the eikonal approximation solution, one defines $\theta_{+k}(x)$ by

$$\theta_{+k}(x) = e^{-ikx}\psi_{+k}(x). \tag{14}$$

This gives, from Eqs. (12) and (13)

$$\theta_{+k}(x) = \psi_a^M + (mi/\hbar^2 k) \times \int_{-\infty}^{\infty} e^{ik|x-x'| - ik(x-x')} V(x')H'\theta_{+k}(x')\,dx', \tag{15}$$

which can also be written as

$$\theta_{+k}(x) = \psi_a^M + (mi/\hbar^2 k)\int_{-\infty}^{x} V(x')H'\theta_{+k}(x')\,dx' + (mi/\hbar^2 k)\int_{x}^{\infty} e^{-2ik(x-x')}V(x')H'\theta_{+k}(x')\,dx'. \tag{16}$$

The eikonal approximation is obtained by neglecting the far right-hand term of Eq. (16). This gives

$$\theta_{+k}(x) \simeq \psi_a^M + (mi/\hbar^2 k)\int_{-\infty}^{x} V(x')H'\theta_{+k}(x')\,dx'. \tag{17}$$

This approximation is valid if $V(x')H'\theta_{+k}(x')$ changes little over a distance $1/k$, which is the case if $|V(x'+1/k) - V(x')|E_k^{-1} \ll 1$ for all x'.[6] We thus require that k_0 be large and the wave packet function $\phi(k-k_0)$ be fairly narrow ($\Delta k/k_0 \ll 1$) so that the eikonal approximation, which is a high-energy approximation, hold for all values of k which have appreciable components in Eq. (11).

The requirement is now imposed that H' be bounded. Under this requirement, Eq. (17) is immediately solved to give

$$\theta_{+k}(x) = \exp\left[\frac{mi}{\hbar^2 k}\int_{-\infty}^{x} V(x')\,dx'H'\right]\psi_a^M, \tag{18}$$

which gives

$$\psi_{+k}(x) = \exp\left[ikx + (mi/\hbar^2 k)\int_{-\infty}^{x} V(x')\,dx'H'\right]\psi_a^M. \tag{19}$$

Since $V(x)$ is negligible for values of x outside the range interval $[-r,r]$, one can use the fact that $\int_{-\infty}^{x} V(y)\,dy \simeq \int_{-\infty}^{\infty} V(y)\,dy$ for all $x > r$ to obtain

$$\psi_{+k}(x) = e^{ikx}e^{iDH'/k}\psi_a^M, \tag{20}$$

where

$$D = (m/\hbar^2)\int_{-\infty}^{\infty} V(x')\,dx'. \tag{21}$$

From Eqs. (11) and (19), one sees that for a given time t, $\psi(x,t)$ is appreciable for those values of x in the interval of width $\Delta x(t)$ centered at $x_0 + v_0 t$. $\Delta x(t)$ is the coordinate space spread of the w system wave packet at the time t and $v_0 = \hbar k_0/m$ is the group velocity of the packet. The small effect of the scattering phase shift on the determination of the value of x at which the phase of $\psi(x,t)$ is stationary has been neglected.

Thus for all time t such that

$$x_0 + v_0 t - \Delta x(t)/2 \geqslant r, \tag{22}$$

$\psi(x,t)$ is appreciable for only those values of x for which Eq. (20) is valid. For such t, one can use Eq. (20) in Eq. (11) instead of Eq. (19). Furthermore, it is easy to arrange the model parameters so that $\Delta x(t) \simeq \Delta x(0)$ for all t during which w is interacting with c. Thus, one can replace $\Delta x(t)$ by $\Delta x(0)$ in Eq. (22).

One now recalls that the desired goal of the above model construction is that the completed w-c scattering should correspond to one step of the process. To this end, the range and strength of the potential V and value of k_0 must be such that

$$D/k_0 = K, \tag{23}$$

where K is the constant appearing in Eq. (2). Then using Eqs. (1), (2), and (20), one has

$$\psi_{+k_0}(x) = e^{ik_0 x}\psi_{T(a)}^M, \tag{24}$$

which is the desired result.

However, account must be taken of the fact that $\psi(x,t)$ is a wave packet with momentum dispersion Δk and there are appreciable contributions to $\psi(x,t)$ for values of k in $(k_0 - \Delta k, k_0 + \Delta k)$, where $\Delta k \geqslant 1/\Delta x(0)$. ($t$ is such that $\Delta x(t) \simeq \Delta x(0)$.) In order to minimize these components Δk must be quite small.

To get an estimate of the magnitude of these unwanted components one notes that for k close to k_0 with $\delta = k_0 - k$

$$e^{i(D/k)H'} = e^{iKH'}e^{i\delta KH'/k} \simeq e^{iKH'}(1 + i\delta KH'/k) \tag{25}$$

for sufficiently small δ, that is for δ such that $K\delta\|H'\|/k \ll 1$. One can thus get an upper estimate of the magnitude of these components by replacing $\psi_{+k}(x)$ in Eq. (11) by

$$\psi_{+k}(x) \simeq e^{ikx}e^{iKH'}(1 + i\Delta kKH'/k_0)\psi_a^M = e^{ikx}(1 + i\Delta kKH'/k_0)\psi_{T(a)}^M. \tag{26}$$

Such a replacement is valid if $\Delta kK\|H'\|/k_0 \ll 1$. This inequality is easily satisfied since the eikonal approximation requires that $\Delta k/k_0 \ll 1$ and H' can be chosen so that $K\|H'\|$ is of order unity. From Eq. (26) one sees that the upper estimate of the magnitude of the unwanted components satisfies $(K\Delta k/k_0)\|H'\psi_{T(a)}^M\| \ll \|\psi_{T(a)}^M\|$ which is the desired condition.

B. Successive scatterings

The results obtained so far describe a model for one step of a process. The model can be extended easily to correspond to n steps of a process. To this end, one first replaces the fixed scattering center c by a one-dimensional lattice C of n fixed scattering centers with the jth center fixed at position x_j. Each center can be regarded as a spinless system fixed at x_j. Initially, the system w is localized well to the left of the leftmost center at x_1. w moves along the C lattice interacting with the scattering centers until it passes out of range of the rightmost center at x_n. The spacing of the centers is arranged so that w interacts with at most one center at a time.

The overall Hamiltonian for the $w + C + M$ system is given by

$$H = H_0 + \sum_{j=1}^{n} V(x - x_j)H', \tag{27}$$

where $V(x - x_j)$ is the interaction potential between w and

the jth system of C. As before, the overall system states at times 0 and t are given by Eqs. (8) and (11).

$\psi_{+k}(x)$ is given by Eq. (12) with appropriate changes made in the interaction potential:

$$\psi_{+k}(x) = e^{ikx}\psi_a^M + (2m/\hbar^2)\int_{-\infty}^{\infty} G_{+k}(x,x') \times \sum_{j=1}^{n} V(x'-x_j)H'\psi_{+k}(x')\,dx', \tag{28}$$

with $G_{+k}(x,x')$ given by Eq. (13).

To obtain a solution of Eq. (28) as a sequence of eikonal scatterings from the successive centers of C, one must require that at no time is w interacting with more than one center. This is conveniently satisfied by letting the adjacent systems of C have a constant spacing $d = x_{j+1} - x_j$ and setting $x_1 = d$ and requiring that

$$d \gg \Delta x(0) + 2r \tag{29}$$

and that there be essentially no wave packet spreading during the time the w-C scatterings are occurring. This means that there be no spreading for all times $t \leqslant t_n$, where

$$t_n = (-x_0 + nd + \Delta x(0)/2 + r)/v_0\,. \tag{30}$$

Note that with $x_1 = d$, $x_j = jd$ for each $j \leqslant n$.

Under this requirement, one can solve Eq. (28) using the eikonal approximation to give

$$\psi_{+k}(x) \simeq \exp\left[ikx + (im/\hbar^2 k)\int_{-\infty}^{x} V(x'-n(x)d)\,dx'H'\right] \times \exp(iDH'/k)^{n(x)-1}\psi_a^M. \tag{31}$$

$n(x)$ is defined as follows: At position x, w is either interacting with some system of C or it is not. If it is, $n(x)$ = place label, l, of the system with which it is interacting. If w is not interacting, then $n(x)$ = place label, l, of the system with which w has just finished interacting. That is,

$$n(x) = \left.\begin{array}{l} l \text{ if } x \geqslant ld - r \text{ and } x < (l+1)d - r \text{ and } l < n \\ n \text{ if } x \geqslant nd - r \end{array}\right\}. \tag{32}$$

Equation (31) can be obtained as the iterative solution of Eq. (28) under the restriction of Eq. (29) and use of arguments similar to those used in Eqs. (16)–(19). [The iteration converges since H' is bounded.] On an intuitive basis, Eq. (31) can be understood as follows: for any position x of w, such that $d + r \leqslant x < 2d - r$, the scattering from the first center can be regarded as complete. For any such x, $\psi_{+k}(x)$ is given by Eq. (20). Now the scattering from the second center at $x_2 = 2d$ can be described by Eq. (12) with $V(x'-x_2)$ replacing $V(x')$ and the outgoing solution, Eq. (20) replacing $\exp(ikx)\psi_a^M$ as the incoming state in Eq. (12). Following the solution of Eq. (12), one gets for all $x > x_1 + r$, a solution similar to Eq. (19) or

$$\psi_{+k}(x) = e^{ikx}e^{(im/\hbar^2 k)}\int_{-\infty}^{x} V(x')\,dx'H'e^{iDH'/k}\psi_a^M. \tag{33}$$

This solution is good for all x for which $x_1 + r < x < x_3 - r$. In particular, for x such that $x_2 + r \leqslant x < x_3 - r$, $\psi_{+k}(x)$, given by Eq. (33) with the upper limit of the x integral replaced by r (or ∞), describes the completed scattering off the second system. By iteration of this description, one can arrive at the solution given by Eq. (31).

If x is such that $x - n(x)d > r$, then the upper limit x of the integral in Eq. (31) can be replaced by r (or ∞) to give

$$\psi_{+k}(x) \simeq e^{ikx}(e^{iDH'/k})^{n(x)}\psi_a^M. \tag{34}$$

This describes the situation when w at x has completed scattering from the first $n(x)$ C systems and has not yet interacted with the $n(x)$ + 1st C system. If x is such that $x - nd > r$, then $n(x)$ in Eq. (34) is replaced by n. For all such x, the scattering is finished. This happens for all times $t > t_n$, Eq. (30), for which the overall state evolves with no further interaction.

Finally, one can extend this quantum mechanical model to take into account the possibility that the interaction Hamiltonian H' for M is stage-dependent, i.e., $H' = H'_j$. Such a model corresponds to a discrete process on A for which the bijection $T_j: A \to A$, which gives the process transformation from stage j to stage $j+1$, is j-dependent. Such models were also used in I, in which quantum mechanical models of Turing machines were constructed.

In this case, one replaces $\Sigma_{j=1}^{n} V(x'-x_j)H'$ in Eqs. (27) by $\Sigma_{j=1}^{n} V(x'-x_j)H'_j$. The solution is carried out in the same manner as that which gives Eq. (31), with the exception that the ordering of the H'_j must be preserved since H'_j and H'_l with $l \neq j$ may not commute. Taking this into account, one obtains the solution as

$$\psi_{+k}(x) \simeq \exp\left[ikx + (im/\hbar^2 k)\times\int_{-\infty}^{x} V(x'-n(x)d)\,dx'H'_{n(x)}\right] \times \exp(iDH'_{n(x)-1}/k)\exp(iDH'_{n(x)-2}/k) \cdots \exp(iDH'_1/k)\psi_a^M, \tag{35}$$

which replaces Eq. (31) as the eikonal approximation solution. Note that even though the process operator T_j and the corresponding operator H'_j are stage-dependent, the Hamiltonian for the extended system $M + C + w$ obtained by replacing H' in Eq. (27) by H'_j, is stage- and time-independent.

C. Limitations on input parameters

It is worthwhile to take a closer look at the limitations on the input parameters which are necessary for Eqs. (31), (33), and (34) to be good approximations. One requirement was that the w-particle wave packet not spread appreciably during the scatterings from the n centers of C. If one assumes that w is free and is described by a minimum wave packet at time 0 (the former is satisfied in that the eikonal approximation requires that the potential be small in comparison to the initial energy and the latter means that the time derivative of $\langle x^2\rangle - \langle x\rangle^2 = 0$ at time 0), then the spread at time t is related to the spread at time 0 by[10]

$$\Delta x(t) = [(\Delta x(0))^2 + (\hbar\Delta k\,t/m)^2]^{1/2}. \tag{36}$$

The requirement $\Delta x(t) \simeq \Delta x(0)$ gives, on expanding the square root and rearranging, $t \ll (\sqrt{2})m\Delta x(0)/(\hbar\Delta k)$.

This must hold for all times during which scattering occurs up to t_n. Thus by Eq. (30), one has with $v_0 = \hbar k_0/m$ and Eq. (29), $-x_0 + (2n+1)(\Delta x(0)/2 + r) \ll (\sqrt{2})k_0\Delta x(0)/\Delta k$

or

$$n \ll [\sqrt{2}k_0\Delta x(0) + x_0\Delta k\ ]/(\Delta x(0) + 2r)\Delta k\ , \tag{37}$$

where x_0 is the position of the center of the w wave packet at time 0.

Thus, there is an upper limit on the number of model steps which can be carried out before the w packet spreads enough to give interference between the scattering from the different centers. This requirement on n can be relaxed by letting the spacing between the jth and $j+1$th centers of C increase as j increases. That is one that simply requires $x_{j+1} - x_j \gg \Delta x(t) + 2r$ instead of $x_{j+1} - x_j = d$ and relaxes the requirement that $\Delta x(t) \simeq \Delta x(0)$. This has the effect that the time it takes the jth model step to be completed increases as j increases.

Another limitation on n comes from the momentum space spread in the w wave packet and resulting unwanted components of the scattering solution $\psi_{+k}(x)$. To obtain this limit, one replaces δ/k in Eq. (25) by $\Delta k/k_0$ as in Eq. (26), and then uses the result in Eq. (34) to get for $n(x) = n$,

$$\psi_{+k}(x) \sim e^{ikx}[(1 + i\Delta kKH'/k_0)e^{iKH'}]^n\psi_a^M\ . \tag{38}$$

The leading unwanted term in the expansion of Eq. (38) is $i(n\Delta kKH'/k_0)(\exp[iKH'])^n\psi_a^M$. In order that this be small in comparison to $[\exp(iKH')]^n\psi_a^M$ it is sufficient to require that

$$n\Delta kK\,\|H'\|/k_0 \ll 1\ , \tag{39}$$

which gives another upper limit on n.

It is of interest to replace the model parameters with actual values which satisfy the various limitations. To this end, let w be an incident 0.25 Mev proton which gives $k_0 \simeq 10^{12}\ \text{cm}^{-1}$. Let $\Delta k = 10^5\ \text{cm}^{-1}$ which gives $\Delta x(0) \geqslant 10^{-5}$ cm. Let $K = 1$ and $r = 10^{-5}$ cm. Define the average strength V_0 of the interaction potential by $2rV_0 = \int_{-r}^{r} V(y)\,dy$. One then has from Eqs. (21) and (23), that $V_0/E_0 \simeq 10^{-7} \ll 1$. Thus, the conditions for the validity of the eikonal approximation can be satisfied. Set $d = 10^{-3}$ cm, which satisfies Eq. (29), and $n = 10^6$. Then Eq. (37) ($|x_0| < 1$) is satisfied and Eq. (39) is satisfied if one can arrange the model so $\|H'\| \sim 1$. The total length of the C lattice is $10^6 \times 10^{-3}$ cm $= 10$ m. A 0.25 MeV proton travels this distance in about 10^{-6} sec which means that each model step is carried out in about 10^{-12} sec. These calculations refer to the parameters of the $w + C$ scattering only and take essentially no account of the possible limitations of the H' interaction in the M system.

IV. MODEL EXTENSION

A. The extended system

As was noted before, in order to include models of processes for which the transfer function T is *not* one-one onto A and to satisfy the locality requirement, one must extend the system. What is done here is to add auxiliary systems and expand the interaction so that the added systems function essentially as a recording system to record the successive steps of the T process. In more detail, let R be a tape with an infinite number of cells labelled by integers in Z, and which is such that all cells are initially blank. Let h_R be a recording head which is initially at R lattice position 0 and let the process system M be in some state a in A. The original process step corresponds to three steps in the expanded system as follows: first, record into the R cell scanned by h_R, the process-system state; next read the R cell scanned, and carry out on the process state, the T change operation corresponding to what was read in the scanned R cell. Thus, if a is read in the R cell scanned by h_R carry out the transformation V_a on the process-system state. $V_a: A \to A$ is a map such that $V_a(a) = T(a)$, $V_a(T(a)) = a$, and $V_a(a') = a'$ otherwise. V_a exchanges the process-system states a and $T(a)$ and leaves others alone. If the R cell scanned by h_R is blank, then V_b is the identity map on A and nothing happens. Finally, shift h_R one cell to the right.

If the process system is initially in state a then repetition of the above steps in the order given generates the record a, $T(a)\cdots T^n(a)\cdots$ in successive cells of R as h_R moves down the R tape and as the process system simultaneously evolves stepwise through successive states $a \to T(a) \to \cdots \to T^n(a) \to \cdots$.

The added systems R and h_R can be modelled as follows: R is represented by a Z lattice of systems r_j for each j in Z. To each r_j is associated a Hilbert space $\mathscr{H}_{r_j}$ which is spanned by an orthonormal basis set $\{\psi_a^{r_j} | a \in A\} \cup \{\psi_b^{r_j}\}$. $\psi_a^{r_j}$ is the state of r_j corresponding to the symbol a being recorded in the jth cell of R and $\psi_b^{r_j}$ is the state corresponding to a blank in the jth cell of R.

The Hilbert space $\mathscr{H}_R$ of the lattice is the space spanned by all infinite product states ψ_γ^R of the form

$$\psi_\gamma^R = \bigotimes_{j \in Z} \psi_{\gamma(j)}^{r_j}\ , \tag{40}$$

where γ is any function from Z to $A \cup \{b\} \equiv A_b$ such that $\gamma(j) \neq b$ for at most a finite number of j values. Let $(A_b^Z)_b$ denote the set of all such functions. $\mathscr{H}_R$ is the incomplete tensor product subspace of the infinite tensor product space $\otimes_{j \in Z} \mathscr{H}_{r_j}$ and is separable.[11]

The recording head h_R is represented in the model by a spinless system whose states lie in the Hilbert space $\mathscr{H}_{h_R} = l^2(Z)$, the space of all square-summable functions on Z. The extended model system Hilbert space $\mathscr{H}'$ is separable, where

$$\mathscr{H}' = \mathscr{H}_M \otimes \mathscr{H}_R \otimes \mathscr{H}_{h_R}\ . \tag{41}$$

From now on M, R, and h_R will denote either the extended process systems or the corresponding systems in the quantum mechanical model. It will be clear from context which is being referred to. The overall $w + C + M + R + h_R$ system Hilbert space is given by

$$\mathscr{H} = \mathscr{H}_w \otimes \mathscr{H}'\ ,$$

where $\mathscr{H}_w = L^2(R, dx)$. C is represented as a lattice of fixed scattering centers, or systems whose state does not change, so no Hilbert space is needed for C.

Figure 2 represents the overall system. The w-wave packet and the positions of the C scattering centers are shown. The lattice position of M is arbitrary.

In order to use this model with the results of Sec. III, one requires that Eq. (7) represent the free Hamiltonian of the $w + C + M + R + h_R$ extended system. This means

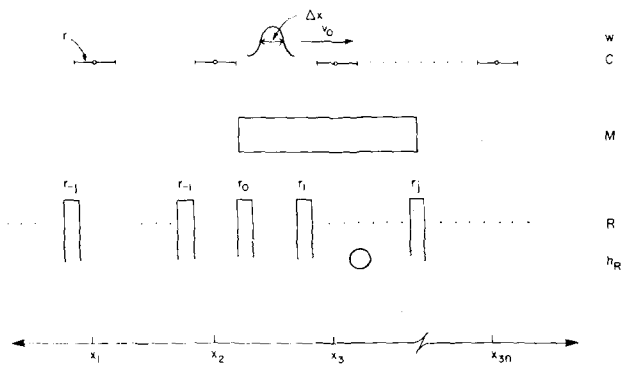


FIG. 2. A schematic representation of the extended model. w and the centers of C are the same as in Fig. 1. M is denoted by the elongated rectangle, the lattice systems r_j of R by smaller vertical rectangles and h_R by a larger circle. An explicit representation of M and each r_j can be given by letting M and r_j be N lattices of spin $\frac{1}{2}$ systems where M extends to the right and the r_j's extend down. The w wave packet lies in between and out of range of c_2 and c_3.

that all systems in R must be noninteracting and that all states ψ_a^M of M must have the same energy. Also, since the Hilbert spaces $\mathscr{H}_M$ and $\mathscr{H}_{r_j}$ for each j in Z are separable, one cannot represent M and each r_j by a spin system as this requires that $\mathscr{H}_M$ and $\mathscr{H}_{r_j}$ be finite-dimensional.

One way to satisfy these requirements explicitly is to let the model systems M and r_j be infinite N lattices of noninteracting spin 1/2 particles. Let the Hilbert spaces $\mathscr{H}_M$ and $\mathscr{H}_{r_j}$ be spanned by all infinite product states Φ_λ of the form

$$\Phi_\lambda = \otimes_{j\in N} \psi^{1/2}_{\lambda(j)j} ,$$

where $\psi^{1/2}_{+j}$ and $\psi^{1/2}_{-j}$ are respective spin up and spin down eigenstates for the $(j+1)$th lattice particle and λ is any $+,-$ sequence such that $\lambda(j) = +$ for at most a finite number of j values. Let the set of such sequences be denoted by $\{\{+,-\}^N\}_b$.

A bijection (or one-one coding) from A onto $(\{+,-\}^N)_b$ can be constructed as follows: Let $a \to l(a)$ be any bijection from A onto N. These exist since A is countably infinite. For each n in N let $\theta(n)$ be the inverted binary representation of n. That is, $0\to 0$, $1\to 1$, $2\to 01$, $3\to 11$, $4\to 001$, $5\to 101$, and so on. Let $+$ correspond to 1, and $-$ to 0; then each $\theta(n)$ becomes a finite $+,-$ sequence. Extend $\theta(n)$ to an infinite sequence $\mu(n)$ in $(\{+,-\}^N)_b$ by adding an infinite sequence of $-$'s only to the right-hand end of $\theta(n)$. Thus, for example, 4 becomes $--+---\cdots$. The desired bijection is then $a\to\mu(l(a))$.

From this one sets

$$\psi_a^M = \Phi_{\mu(l(a))}$$

for each a in A. Since $a\to\mu(l(a))$ is a bijection, the $\{\Phi_{\mu(l(a))} | a\in A\}$ span $\mathscr{H}_M$.

$\psi_y^{r_j}$ and $\mathscr{H}_{r_j}$ are constructed in a similar way where y denotes an element of A_b. The only difference is that one requires a bijection from A_b onto $(\{+,-\}^N)_b$. This is conveniently given by $a\to\mu(l(a)+1)$ and $b\to\mu(0)$. The other details are the same. Note that for this example R is a Z lattice of noninteracting systems each of which is an N lattice of spin $\frac{1}{2}$ systems. Thus $\mathscr{H}_R = (\otimes_{j\in Z}\mathscr{H}_{r_j})_b$ is an incomplete tensor product of the $\mathscr{H}_{r_j}$ and each $\mathscr{H}_{r_j} = (\otimes_{j\in N}\mathscr{H}_{1/2})_b$ is an incomplete tensor product of the two-dimensional space $\mathscr{H}_{1/2}$.

B. The elementary operators

One must now define the necessary elementary operators which are used to build the interactions. For each a in A, define the projection operator P_a on $\mathscr{H}_M$ by $P_a\psi_{a'}^M = \psi_a^M\delta_{a,a'}$. Also define the operator U_a on $\mathscr{H}_M$ by

$$U_a\psi_{a'}^M \begin{cases} = \psi_{T(a)}^M & \text{if } a' = a \\ = \psi_a^M & \text{if } a' = T(a) \\ = \psi_{a'}^M & \text{otherwise} \end{cases} . \tag{42}$$

Also, define $U_b = 1$, where b denotes the blank. U_a exchanges ψ_a^M and $\psi_{T(a)}^M$ and leaves the other basis functions alone. Note that U_a is well defined even if T is many-one and into. Also U_a is both unitary and self-adjoint.

For each real number δ define U_a^δ by

$$U_\alpha^\delta = e^{i\delta H}a , \tag{43}$$

where H_a is any self-adjoint operator which satisfies $\exp(iKH_a) = U_a$. One clearly has for each a for which $a\neq T(a)$

$$U_a^\delta\psi_a^M = \alpha(\delta)\psi_a^M + \beta(\delta)\psi_{T(a)}^M , \tag{44}$$

where $\alpha(\delta)$ and $\beta(\delta)$ are possibly a-dependent coefficients such that $\alpha(0) = \beta(K) = 1$, $\alpha(K) = \beta(0) = 0$, $|\alpha(\delta)|^2 + |\beta(\delta)|^2 = 1$ and $\alpha^*(\delta)\beta(\delta) + \alpha(\delta)\beta^*(\delta) = 0$. One also has $U_a^\delta\psi_{T(a)}^M = \alpha(\delta)\psi_{T(a)}^M + \beta(\delta)\psi_a^M$. For all a' such that $U_a\psi_{a'}^M = \psi_{a'}^M$ [including those a' for which $a' = a = T(a)$], one has $U_a^\delta\psi_{a'}^M = \gamma(\delta)\psi_{a'}^M$, where $|\gamma(\delta)| = 1$ and $\gamma(0) = \gamma(K) = 1$.

There are many possible choices for H_a which satisfy the above conditions. As regards the model aspects of concern here, it does not matter which choice is made other than the requirement that all the H_a have an a-independent bound. To make the presentation simpler, it will also be required that H_a is such that $\alpha(\delta)$ and $\beta(\delta)$ are independent of a and $\gamma(\delta) = 1$ for all δ. A choice which satisfies this requirement is

$$H_a = \sum_{l=0}^{1} (\pi l/K)P_a^l , \tag{45}$$

where $P_a^l = \psi_a^l)(\psi_a^l$ and $\psi_a^l = (\sqrt{2})^{-1}(\psi_a^M + (-1)^l\psi_{T(a)}^M)$ if $a\neq T(a)$ and $H_a = 0$, if $a = T(a)$. With this choice of H_a, one has $\alpha(\delta) = (1 + \exp(i\pi\delta/K))/2$ and $\beta(\delta) = (1 - \exp(i\pi\delta/K))/2$.

Another possible advantage of this choice of H_a is, in brief, as follows: If one considers the density operator state $\rho_a(\delta) = U_a^\delta\psi_a^M)(\psi_a^M U_a^{\delta\dagger}$, Eq. (45) gives the result that there are no values of δ such that $0<\delta<K$ for which $\rho_a(\delta) = \rho_{T(a)} \equiv \psi_{T(a)}^M)(\psi_{T(a)}^M$. Thus as the model system evolves under one w-c scattering in C, there is no cycling of $\rho_a(\delta)$ between ρ_a and $\rho_{T(a)}$ and back as δ increases from 0 to K. This lack of cycling is consistent with the locality requirement discussed in Sec. II because it means that $\rho_a(\delta)$ arrives at $\rho_{T(a)}$ for the first time only at the completion of one w-c scattering. Other choices of H_a, e.g., $H_a = (2n\pi/K)P_a^0$

$+ ((2m+1)\pi/K)P^{\frac{1}{a}}$ with $n \neq m$, give such cycling.

For each a in A and k in Z define the projection operator P_{ak} on $\mathscr{H}_R$ by $P_{ak}\psi^R_\gamma = \psi^R_\gamma \delta_{\gamma(k),a}$ where for each function γ in $(A^{Z}_{b})_b$, ψ^R_γ is given by Eq. (40). Also define the operator U_{ak} by

$$U_{ak}\psi^R_\gamma = \psi^R_{\gamma'}, \tag{46}$$

where γ' is such that $\gamma(j) = \gamma'(j)$ for all $j \neq k$ and $\gamma'(k) = a[b]$ if $\gamma(k) = b[a]$, and $\gamma = \gamma'$ if $\gamma(k) \neq a,b$. U_{ak} acting on ψ^R_γ changes at most the k th factor of ψ^R_γ and that only if $\gamma(k) = a$ or b. In this case, it interchanges $\psi^{r_k}_a$ and $\psi^{r_k}_b$. U_{ak} is both unitary and self-adjoint.

For each real number δ, define U^δ_{ak} by

$$U^\delta_{ak} = e^{i\delta H_{ak}}, \tag{47}$$

where H_{ak} is any self-adjoint operator which satisfies $\exp[iKH_{ak}] = U_{ak}$. Then one has for each γ for which $\gamma(k) = a$ or b,

$$U^\delta_{ak}\psi^M_\gamma = \alpha(\delta)\psi^M_\gamma + \beta(\delta)\psi^M_{\gamma'}, \tag{48}$$

where $\alpha(\delta)$ and $\beta(\delta)$ are possibly γ-dependent coefficients which satisfy the conditions given for Eq. (44). Also let $U^\delta_{ak}\psi^M_\gamma = \psi^M_\gamma$ if $\gamma(k) \neq a,b$.

As was the case for U_a, there are many possible choices for H_{ak}. However, one does require that $\|H_{ak}\| < m$ for some m for all a and k. A possible choise for H_{ak}, which satisfies the conditions already discussed for H_a and is nonzero only in the component space $\mathscr{H}_{r_k}$ of $\mathscr{H}_R$, is defined by

$$H_{ak} = \sum_{l=0}^{1} (\pi l/K) P^l_{ak}, \tag{49}$$

where $P^l_{ak} = \psi^l_{ak})(\psi^l_{ak}$ and $\psi^l_{ak} = (\sqrt{2})^{-1}(\psi^{r_k}_b + (-1)^l\psi^{r_k}_a)$.

One also needs the projection operator P^R_k defined on $\mathscr{H}_R$ for each k in Z by

$$P^R_k = \sum_{\gamma \in (A^{Z}_{b})_b}{}' P^R_\gamma, \tag{50}$$

where $P^R_\gamma = \psi^R_\gamma)(\psi^R_\gamma$, Eq. (40)), and the prime on the sum means that it is restricted to those γ in $(A^{Z}_{b})_b$ for which $\gamma(k) \neq b$ and $\gamma(j) = b$ for all $j > k$. P^R_k projects out all record expression states for which the last (in the direction of increasing Z) nonblank record cell is at position k.

In a similar fashion, one defines on $\mathscr{H}_{h_R}$, the projection operator $P_k = \psi^{h_R}_k)(\psi^{h_R}_k$, where $\psi^{h_R}_k$ is the state in $l^2(Z)$ which corresponds to h_R being localized at lattice position k. Also, let H_k and U^k_{+1} be such that

$$U^k_{+1}\psi^{h_R}_l = e^{iKH_k}\psi^{h_R}_l = \psi^{h_R}_{l'}, \tag{51}$$

where $l' = k+1[k]$, if $l = k[k+1]$ and $l = l'$ otherwise. U^k_{+1} exchanges $\psi^{h_R}_k$ and $\psi^{h_R}_{k+1}$ and leaves all other states in the basis alone. U^k_{+1} is both unitary and self-adjoint. Let H_k be given by $H_k = \Sigma^1_{l=0}(\pi l/K)P^l_k$, where $P^l_k = \psi^l_k)(\psi^l_k$ and $\psi^l_k = (\sqrt{2})^{-1}(\psi^{h_R}_k + (-1)^l\psi^{h_R}_{k+1})$.

As was done for U_a, define $U^{k\delta}_{+1}$ by

$$U^{k\delta}_{+1} = e^{i\delta H_k} \tag{52}$$

for each real number δ. This gives

$$U^{k\delta}_{+1}\psi^{h_R}_k = \alpha(\delta)\psi^{h_R}_k + \beta(\delta)\psi^{h_R}_{k+1}, \tag{53}$$

where $\alpha(\delta)$ and $\beta(\delta)$ satisfy the conditions given for Eq. (44). Clearly $U^{k\delta}_{+1}\psi^{h_R}_l = \psi^{h_R}_l$ if $l \neq k, k+1$. Also $\alpha(\delta)$ and $\beta(\delta)$ are independent of k.

C. The step interaction operators

The operators defined so far can be used to define overall record, process, and shift step operators as follows: Define the recording step operator U_1 on $\mathscr{H}'$ by

$$U_1 = \sum_{\substack{a\in A\\ k\in Z}} P_a \otimes U_{ak} \otimes P_k, \tag{54}$$

where U_{ak} is defined by Eq. (46). U_1 corresponds in the model to recording the state of the process into that cell of R which is scanned by h_R provided that the cell scanned is initially blank. U_1 acting on a state of the form $\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_k$, where $\gamma(k) = b$ converts it to $\psi^M_a \otimes \psi^R_{\gamma'} \otimes \psi^{h_R}_k$, where $\gamma'(k) = a$ and $\gamma(j) = \gamma'(j)$ for all $j \neq k$. U_1 is clearly unitary since U_{ak} is unitary and $\Sigma_a P_a = 1 = \Sigma_k P_k$. Also U_1 is self-adjoint since $U^2_{ak} = 1$.

Define the process-step operator U_2 by

$$U_2 = \sum_{\substack{y\in A_b\\ k\in Z}} U_y \otimes P_{yk} \otimes P_k, \tag{55}$$

where U_{ak} is defined by Eq. (46). U_1 corresponds in the model to recording the state of the process into that cell of R which is scanned by h_R provided that the cell scanned is initially blank. U_1 acting on a state of the form $\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_k$, where $\gamma(k) = b$ converts it to $\psi^M_a \otimes \psi^R_{\gamma'} \otimes \psi^{h_R}_k$, where $\gamma'(k) = a$ and $\gamma(j) = \gamma'(j)$ for all $j \neq k$. U_1 is clearly unitary since U_{ak} is unitary and $\Sigma_a P_a = 1 = \Sigma_k P_k$. Also U_1 is self-adjoint since $U^2_{ak} = 1$.

The shift operator U_3 is defined on $\mathscr{H}'$ from Eqs. (50) and (51) by

$$U_3 = \sum_{k\in Z} 1 \otimes P^R_k \otimes U^k_{+1}. \tag{56}$$

U_3 corresponds to shifting the head to a fresh record cell provided it is scanning the last filled record cell. U_3 acting on a state of the form $\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_k$, where $\gamma(k) \neq b$ and for all $j > k$, $\gamma(j) = b$ converts it into $\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_{k+1}$. Since U^k_{+1} is unitary and self-adjoint and $\Sigma_k P^R_k = 1$, U_3 is both unitary and self-adjoint.

Using the definitions of H_{ak}, H_a, and H_k already given, one can define three interaction operators H_1, H_2, and H_3 by

$$H_1 = \sum_{\substack{a\in A\\ k\in Z}} P_a \otimes H_{ak} \otimes P_k, \tag{57}$$

$$H_2 = \sum_{\substack{a\in A\\ k\in Z}} H_a \otimes P_{ak} \otimes P_k, \tag{58}$$

and

$$H_3 = \sum_{k\in Z} 1 \otimes P^R_k \otimes H_k \tag{59}$$

Using these operators, one can show that for $j = 1, 2, 3$,

$$U_j = e^{iKH_j} \tag{60}$$

holds with U_j given by Eqs. (54)–(56). This can be seen by substitution into the exponential, expanding in a power series (which is valid since there exist bounds on H_{ak}, H_a, and H_k which are independent of the indices) and noting that the terms of Eqs. (57)–(59) are pairwise orthogonal.

D. The final model

The results obtained here can be combined directly with those of Sec. III as follows: From the viewpoint of Sec. IIA and III, the process system must be regarded as $M + R + h_R$ with an overall state space $A \times (A^{\,z}_{\,b})_b \times Z$. One has three transition functions T_1, T_2, and T_3 for the record, process, and shift steps, which are defined so they satisfy

$$U_j(\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_k) = \psi^{MRh_R}_{T_j(a\gamma k)} \tag{61}$$

for each $j = 1, 2, 3$ and each $(a\gamma k)$ in $A \times (A^{\,z}_{\,b})_b \times Z$. In particular, $T_1(a\gamma k) = (a\gamma' k)$ and $\psi^{MRh_R}_{T_1(a\gamma k)} = \psi^M_a \otimes \psi^R_{\gamma'} \otimes \psi^{h_R}_k$, where for all $j \neq k, \gamma(j) = \gamma'(j)$ and $\gamma'(k) = a[b]$ if $\gamma(k) = b$ [a], and $\gamma'(k) = \gamma(k)$ otherwise. For T_2 if $\gamma(k) = b$, then $T_2(a\gamma k) = (a\gamma k)$ for all a. If $\gamma(k) = a$, then $T_2(a'\gamma k) = (a_1\gamma k)$, where $a_1 = T(a)[a]$ if $a' = a[T(a)]$, and $a' = a_1$ otherwise. $T_3(a\gamma k) = (a\gamma k')$ where $k' = k + 1[k-1]$ if $[k-1]$ = position of last nonblank symbol in γ and $k' = k$ otherwise. It is clear from these definitions and the fact that U_j is both unitary and self-adjoint that $T_j^2 = 1$ and T_j is a bijection on $A \times (A^{\,z}_{\,b})_b \times Z$ for $j = 1, 2, 3$.

The final model is supposed to evolve under a Hamiltonian in such a way that each step of the original process T on M becomes three successive steps, record, process, then shift, in the extended model. Each step in the extended model takes place by turning on the appropriate interaction operator by means of w interacting with a scattering center in the control lattice C.

To this end, one defines for each process T on M an extended model Hamiltonian by Eq. (27)

$$H^T_{3n} = H_0 + \sum_{j=1}^{3n} V(x - x_j) H^T(j), \tag{62}$$

where H_0 is given by Eq. (7) and $H^T(j)$ is defined from Eqs. (57)–(59) by

$$H^T(j) = \left.\begin{matrix} H_1 \\ H_2 \\ H_3 \end{matrix}\right\} \quad \text{if } j = \begin{matrix} 1 \bmod 3 \\ 2 \bmod 3 \\ 0 \bmod 3 \end{matrix}. \tag{63}$$

In these expressions, the dependence of the Hamiltonian on the process T is made explicit. Also for each j, $H^T(j)$ is bounded as is required for use in the results of Sec. III.

The final result of this work is obtained by combining the results of the previous sections. That is, to each discrete process T on M with a countable set A of states and for each n, one can associate an extended quantum mechanical model $w + C + M + R + h_R$ such that the Schrödinger evolution of the model describes the process in the following sense:

For any time T and w-position x, the overall model state is given by Eq. (11) as

$$\psi(x,t) = \int dk\, e^{-iE_k t/\hbar} \phi(k - k_0) \psi_{+k}(x) e^{-ikx_0} \tag{11}$$

and $\psi_{+k(x)}$ is given by Eq. (35) as $\psi_{+k}(x) = e^{ikx} \psi^{MRh_R}_k(x)$, where

$$\psi^{MRh_R}_k(x) \simeq \exp\left[(im/\hbar^2 k) \int_{-\infty}^{x} V(x' - n(x)d)\, dx' H^T(n(x))\right] \times \exp\left[\frac{iD}{k} H^T(n(x) - 1)\right] \cdots \exp[iDH^T(1)/k]\, \Phi^{MRh_R}, \tag{35}$$

where $n(x)$ [Eq. (32)] is the place label of the scattering center in C which is either interacting with w at x or, if none is interacting, has just finished interacting with w. D is given by Eq. (21) and $H^T(j)$ by Eq. (63). This result also requires that the separation d between successive scattering centers satisfies Eq. (29), or $d \gg \Delta x + 2r$, where Δx is the w-wave packet spread at time 0 and r is the range of the interaction potential.

Let $\Phi^{MRh_R} = \psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_l$, where $\gamma(j) = b$ for all $j \geq l$. Then if $\phi(k - k_0)$ is sufficiently narrow, $\psi^{MRh_R}_k(x)$ is essentially independent of k over the important values in Eq. (11). As a result, one has (Eq. 3)

$$\psi(x,t) \approx \psi^{MRh_R}_{k_0}(x) F_{k_0 x_0}(x,t), \tag{64}$$

where

$$\psi^{MRh_R}(x) = \exp\left[(im/\hbar^2 k_0) \int_{-\infty}^{x} V(x' - n(x)d)\, dx' H^T(n(x))\right] \times \exp[iKH^T(n(x) - 1)] \cdots \exp[iKH^T(1)](\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_l) \tag{65}$$

and

$$F_{k_0 x_0}(x,t) = \int dk\, e^{-iE_k t/\hbar} \phi(k - k_0) e^{ik(x - x_0)}. \tag{66}$$

For each time t, the only values of x that contribute appreciably to Eq. (64) are those in the interval of width Δx centered on $x_0 + v_0 t$. If t is such that the interval end points also satisfy $pd + r < x_0 + v_0 t - \Delta x/2$ and $x_0 + v_0 t + \Delta x/2 < (p+1)d - r$ for some $p < 3n$, then the w-wave packet has finished interacting with the pth center and is not yet interacting with the $p+1$th center, (Fig. 2). This shows that for all times t in the interval $\Delta = [d - (\Delta x + 2r)]/v_0$ centered on $[(p + \frac{1}{2})d - x_0]/v_0$, $\psi^{MRh_R}_{k_0}(x)$ is independent of x for the relevant values of x and can be replaced by

$$\psi^{MRh_R}_{k_0}(p) = e^{iKH^T(p)} e^{ikH^T(p-1)} \cdots e^{ikH^T(1)} (\psi^M_a \otimes \psi^R_\gamma \otimes \psi^{h_R}_l), \tag{67}$$

where $n(x) = p$.

This state describes the completion of p model steps and can also be written as

$$\psi^{MRh_R}(p) = \psi^M_{T^f(a)} \otimes \psi^R_{\gamma^f} \otimes \psi^{h_R}_{l+s}. \tag{68}$$

Here e, f, and s are the largest integers in $(p+1)/3$, $(p+2)/3$, and $(p/3)$ respectively and γ^f is related to γ by $\gamma^f(j) = \gamma(j)$ for $j < l$ and $j > l + f$ and $\gamma^f(j) = T^{j-l}(a)$ for $l \leq j \leq l + f$. This corresponds to the situation in which e steps of the process T have been carried out on M initially in state a, the history of f steps is recorded on R which was initially in

state γ, and h_R was shifted from l to $l+s$. The length of time Δ for which the system remains in this state can be set by adjusting the model parameters, in particular $\Delta = d - (\Delta x + 2r)$.

For all times $t > (3nd - x_0 + \frac{1}{2}\Delta x + r)/v_0$, $\psi_k^{MRh_R}(x)$ can be replaced in Eq. (11) by $\psi_k^{MRh_R}(3n)$. This corresponds to the completion of $3n$ model steps when w has moved out of range of all the $3n$ scattering centers in C and shows that no further changes occur in $M + R + h_R$.

For the times described above when the w wave packet is essentially entirely between c_p and c_{p+1} and not interacting with either, one has for the overall state

$$\psi(x,t) \simeq \psi^{MRh_R}(p) F_{k_0 x_0}(x,t). \tag{69}$$

For those times for which Eq. (69) is valid, the w system state is uncorrelated with the $M + R + h_R$ system state.

For those times t for which at least some of the w wave packet is within interaction range of some center c_p of C, the overall state is obtained from Eqs. (64)–(67) as

$$\psi(x,t) \begin{cases} \simeq \psi_{k_0}^{MRh_R}(p-1) F_{k_0 x_0}(x,t) & \text{if } x \leqslant pd - r \\ \simeq e^{i\delta(x) H^T(p)} \psi_{k_0}^{MRh_R}(p-1) F_{k_0 x_0}(x,t) & \\ \qquad \text{if } pd - r < x < pd + r & \\ \simeq \psi_{k_0}^{MRh_R}(p) F_{k_0 x_0}(x,t) & \text{if } x \geqslant pd + r \end{cases}, \tag{70}$$

where $\delta(x) = (M/\hbar^2 k_0) \int_{-\infty}^{x} V(x' - pd)\, dx'$. The first part refers to those components of the w wave packet which have completed interaction with c_{p-1} and have not yet begun interaction with c_p. The second and third parts describes those components which are interacting with and have completed interacting with c_p, respectively.

It is of interest to examine $\exp(i\delta(x) H^T(p)) \psi_{k_0}^{MRh_R}(p-1)$ further. From the definitions of H_1, H_2, and H_3, and Eqs. (57)–(59) one can easily see that

$$e^{i\delta(x) H^T(p)} \begin{cases} = \sum_{\substack{a \in A \\ k \in Z}} P_a \otimes U_{ak}^{\delta(x)} \otimes P_k & \text{if } p = 1 \bmod 3 \\ = \sum_{\substack{y \in A_b \\ k \in Z}} U_y^{\delta(x)} \otimes P_{yk} \otimes P_k & \text{if } p = 2 \bmod 3 \\ = \sum_{k \in Z} 1 \otimes P_k^R \otimes U_{+1}^{k\delta(x)} & \text{if } p = 0 \bmod 3 \end{cases}, \tag{71}$$

where $U_y^{\delta(x)}$, $U_{ak}^{\delta(x)}$, and $U_{+1}^{k\delta(x)}$ are given by Eqs. (43), (47), and (52).

Let a, γ, and l be appropriate initial states. That is $\gamma(j) = b$ for all $j \geqslant l$. Let a', γ', l' satisfy Eq. (68) with $a' = T^e(a)$, $\gamma' = \gamma^f$ and $l' = l + s$.

Then Eq. (71) in combination with Eqs. (44), (48), and (53) give

$$e^{i\delta H^T(p)} (\psi_{a'}^M \otimes \psi_{\gamma'}^R \otimes \psi_{l'}^{MRh_R}) \begin{cases} = \psi_{a'}^M \otimes (\alpha(\delta)\psi_{\gamma'}^R + \beta(\delta)\psi_{\gamma''}^R) \otimes \psi_{l'}^{h_R} & \text{if } p = 1 \bmod 3 \\ = (\alpha(\delta)\psi_{a'}^M + \beta(\delta)\psi_{T(a')}^M) \otimes \psi_{\gamma'}^R \otimes \psi_{l'}^{h_R} & \text{if } p = 2 \bmod 3 \\ = \psi_{a'}^M \otimes \psi_{\gamma'}^R \otimes (\alpha(\delta)\psi_{l'}^{h_R} + \beta(\delta)\psi_{l'+1}^{h_R}) & \text{if } p = 0 \bmod 3 \end{cases}, \tag{72}$$

where $\gamma''(j) = \gamma'(j)$ if $j \neq l'$ and $\gamma''(l') = a'$ and $\gamma'(l') = b$. The dependence of δ on x has been suppressed to make the notation easier. If a' is such that $a' = T(a')$, then the right-hand side of Eq. (72) for $p = 2 \bmod 3$ is replaced by $\psi_{a'}^M \otimes \psi_{\gamma'}^R \otimes \psi_{l'}^{MRh_R}$. $\alpha(\delta)$ and $\beta(\delta)$ are independent of a', γ', and l' and satisfy the conditions given for Eq. (44). The independence follows from the definitions of H_a, H_{ak}, and H_k used in Eqs. (57)–(59).

As a result, if a, γ, and l are appropriate initial states, one has the result

$$e^{i\delta(x) H^T(p)} \psi_{k_0}^{MRh_R}(p-1) = \alpha(\delta(x)) \psi_{k_0}^{MRh_R}(p-1) + \beta(\delta(x)) \psi_{k_0}^{MRh_R}(p) \tag{73}$$

holds provided that either $p \neq 2 \bmod 3$, or $p = 2 \bmod 3$ and $a' \neq T(a')$. If $p = 2 \bmod 3$ and $a' = T(a')$, the right-hand side of Eq. (73) is replaced by $\psi_{k_0}^{MRh_R}(p-1)$.

This is the desired result because it shows that the part of the overall state $\psi(x,t)$ which represents the partially completed pth step is a linear combination of the states representing the $p-1$th step and the pth step. As the time t [and relevant values of $\delta(x)$] increases, $\alpha(\delta(x))$ decreases and $\beta(\delta(x))$ increases. Thus $\psi_{k_0}^{MRh_R}(p)$ grows in at the expense of $\psi_{k_0}^{MRh_R}(p-1)$.

Thus, one sees that if a, γ, l correspond to a possible initial state, the requirement of locality, discussed in Sec. II is satisfied. Furthermore, it is satisfied for all three types of interactions, record, T-process, and shift. As noted, this is the main reason for defining all operations in terms of elementary operators which exchange just two basis vectors and leave the others alone. This is also the reason why the shift operator, Eq. (56), is defined in terms of P_k^R and U_{+1}^k instead of the bilateral shift operator on $l^2(Z)$.

It should be noted that the actual quantum mechanical model state of the subsystem $M + R + h_R$ is given by the density operator $\rho^{MRh_R}(t)$, where $\rho^{MRh_R}(t) = \mathrm{Tr}_w(\rho^{MRh_R w}(t))$, where $\rho^{MRh_R w}(t)$ is the density operator for the overall system and is pure. From the fact that

$$(x, \rho^{MRh_R w}(t) x) = \psi(x,t)\ (\psi(x,t)$$

one sees that when t is such that the w system is not interacting with a center c_p of C, then the w subsystem state is uncorrelated with the $M + R + h_R$ subsystem state, Eq. (69), and $\rho^{MRh_R}(t)$ is pure. However, when t is such that the states of w and of $M + R + h_R$ are correlated as in Eq. (70), then $\rho^{MRh_R}(t)$ is impure. This happens when at least some of the w wave packet is interacting with a center c_p of C.

It will be recalled that the validity of Eqs. (64)–(66) upon which the usefulness of the description given by Eq. (67)–(73) depends, is conditioned on $\phi(k - k_0)$ being sufficiently narrow. How narrow $\phi(k - k_0)$ must be can be seen from Eq. (39), which gives the condition

$$3n\Delta k K \|H^T\|/k_0 \ll 1.$$

Here Δk is the w wave packet momentum spread and $\|H^T\|$ = maximum of $\|H_1^T\|$, $\|H_2^T\|$, $\|H_3^T\|$.

This condition, which is necessary for the success of the process model description constructed here, gives a relationship among the parameters which must be satisfied. Thus, for given values of k_0, Δk, K, and $\|H^T\|$, this relation gives

an upper limit on the number of process steps which can be correctly described before the model breaks down. Also, the smaller Δk is, the larger Δx is, and by Eq. (29), the larger the separation, d, of the C scattering centers must be.

V. THE DILATION THEOREM OF SZ. NAGY

It is of interest to relate the model extension described in the previous section to the results embodied in the dilation theorem of Sz. Nagy. This theorem[7] says that any contraction W on a Hilbert space $\mathscr{H}$ can be dilated to a unitary operator V on a larger space $\mathscr{K}$ with $\mathscr{H} \subset \mathscr{K}$ such that $W^n P = PV^n P$ for $n = 1,2...$, where $P\mathscr{K} = \mathscr{H}$. Furthermore, there exists a minimal unitary dilation of W which is unique up to isomorphism.

This theorem can be applied to the model process discussed here as follows: Let the process function T be boundedly many-one. That is let T be such that there exists an m such that for all $a\{a'|T(a') = a\}$ has $\leqslant m$ elements. As before T can be into A. Partition A into a family of nonempty subsets $A_j, j = 1,2,...$ such that for each j, A_j is T invariant and irreducible. That is, $TA_j \subseteq A_j$ and there is no nontrivial subset B of A_j such that $TB \subseteq B$ and $T(A_j - B) \subseteq A_j - B$. Depending on T, the partition may contain one, several or denumerably many subsets of A. The proof that such a partition exists and is unique is given in the Appendix.

Consider the operator U_T on $\mathscr{H}_M$ as defined by Eq. (1). For each A_j in the partition, let $\mathscr{H}_{A_j}$ be the subspace of $\mathscr{H}_M$ spanned by $\{\psi_a^M | a \in A_j\}$. Then $\mathscr{H}_M = \oplus_j \mathscr{H}_{A_j}$. It is clear that for each A_j in the partition, $\mathscr{H}_{A_j}$ reduces U_T. Thus, one has $U_T = \oplus_j U_{T_j}$ where U_{T_j} is the restriction of U_T to $\mathscr{H}_{A_j}$ and $U_{T_j}\mathscr{H}_{A_j} \subseteq \mathscr{H}_{A_j}$. Furthermore, each $\mathscr{H}_{A_j}$ is U_{T_j} irreducible since A_j is T irreducible. Since T is boundedly many-one, each U_{T_j} is bounded.

One now defines the operator W_T on $\mathscr{H}_M$ by

$$W_T = \oplus_j (U_{T_j}/\|U_{T_j}\|), \tag{74}$$

where the sum is over all j in the partition: W_T is a contraction operator which satisfies $W_T\psi_a = (\|U_{T_j}\|)^{-1}\psi_{T(a)}$, where j defines the subset A_j in the partition to which a belongs. In what follows, the subscript T on W will be suppressed.

Following Sz. Nagy and Foias,[7] define the operators D_W and $D_{W^\dagger}$ on $\mathscr{H}_M$ by $D_W = (1 - W^\dagger W)^{1/2}$ and $D_{W^\dagger} = (1 - WW^\dagger)^{1/2}$. D_W and $D_{W^\dagger}$ are self-adjoint contractions which satisfy $WD_W = D_{W^\dagger}W$, and $D_W W^\dagger = W^\dagger D_{W^\dagger}$. Also $\|D_W\psi\|^2 = \|\psi\|^2 - \|W\psi\|^2$ for each ψ in $\mathscr{H}_M$. This implies $D_W\psi = 0$ if ψ lies in an $\mathscr{H}_{A_j}$ such that U_{T_j} is an isometry on $\mathscr{H}_{A_j}$ (or $T|A_j$ is one-one). Both $D_W\psi = 0$ and $D_{W^\dagger}\psi = 0$, if $\psi \in \mathscr{H}_{A_j}$ and U_{T_j} is unitary on $\mathscr{H}_{A_j}$ (or $T|A_j$ is one-one onto).

Define $\mathscr{L}'$ to be the Hilbert space $\oplus_{j\in Z}\mathscr{H}_M$ of all functions θ from Z to $\mathscr{H}_M$ such that $\|\theta\|^2 = \Sigma_{j\in Z}\|\theta(j)\|^2 < \infty$. $\mathscr{H}_M$ can be embedded into $\mathscr{L}'$ by the map $\psi \to \theta_\psi$, where $\theta_\psi(j) = 0$ if $j \neq 0$ and $\theta_\psi(0) = \psi$, where ψ is any vector in $\mathscr{H}_M$. Let V be the dilation of W onto $\mathscr{L}'$ defined by $V\theta = \theta'$, where $\theta'(j) = \theta(j+1)$ if $j \neq -1, 0$ and $\theta'(-1) = D_W\theta(0) - W^\dagger\theta(1)$ and $\theta'(0) = W\theta(0) + D_W^\dagger(\theta(1))$. Then $V^{-1}\theta = \theta_1$, where $\theta_1(j) = j - 1$ if $j \neq 0, 1$ and $\theta_1(0) = D_W\theta(-1) + W^\dagger\theta(0)$ and $\theta_1(1) = -W\theta(-1) + D_{W^\dagger}\theta(0)$. Since V is unitary, $V^{-1} = V^\dagger$.

V on $\mathscr{L}'$ is a unitary dilation of W on $\mathscr{H}_M$. However, it is not minimal. To construct a minimal unitary dilation consider for each ψ in $\mathscr{H}_M$ and each $n = 0,1,2...$, the vectors in $\mathscr{L}'$ given by,[7]

$$\begin{array}{llllllll} V^n\theta_\psi = ...0, & D_W\psi, & D_W W\psi, ..., & & D_W W^{n-1}\psi, & W^n\psi, & 0, & ... \\ ... -n-1, & -n, & -n+1 & ..., & -1, & 0, & 1, & ... \end{array} \tag{75}$$

and

$$\begin{array}{llllllll} V^{-n}\theta_\psi = ...0, & (W^\dagger)^n\psi, & D_{W^\dagger}(W^\dagger)^{n-1}\psi, & ..., & D_{W^\dagger}W^\dagger\psi, & D_{W^\dagger}\psi, & 0, & ... \\ ... -1, & 0, & 1, & ..., & n-1, & n, & n+1, & ..., \end{array} \tag{76}$$

where the Z place labels of the corresponding terms are indicated directly underneath the terms. Let $\mathscr{L}$ be the Hilbert space spanned by all the vectors $V^j\theta_\psi$, where j is in Z and ψ is in $\mathscr{H}_M$ and let $V|\mathscr{L}$ be restriction of V to $\mathscr{L}$. V on $\mathscr{L}$ is a minimal unitary dilation of W on $\mathscr{H}_M$.[7] From now on V is considered to be restricted to $\mathscr{L}$.

Examination of $V_T^n\theta_\psi$ (the subscript T is reinstated) with $\psi = \psi_a^M$ shows that $V_T^n\theta_{\psi_a^M}$ contains in its components a history of the first n model steps of the process T. For our purposes, the details are not important. Note, though, that if a is such that the corresponding U_{T_j}, which acts on ψ_a^M, is an isometry, then $D_W W^l\psi_a^M = 0$ for all l. In this case, and only in this case, the history can be recovered from $W_T^n\psi_a^M$ as T is 1-1 on A_j. Also $W_T^n\psi_a^M = \|U_{T_j}\|^{-n}\psi_{T^n(a)}^M$ is the model state corresponding to the completion of n steps of the process T if it starts in state ψ_a^M and j is such that $a \in A_j$.

As a consequence the results of Sec. III can be taken over directly. Let H_T' be any bounded self-adjoint operator which satisfies

$$e^{iKH'_T} = V_T \tag{77}$$

and define by Eq. (27), the Hamiltonian $H_T = H_\theta + \Sigma_{j=1}^n V(x - x_j)H_T'$. Then following the development of Sec. III, one constructs from H_T a quantum-mechanical Hamiltonian model of the minimal unitary dilation of H_T on $\mathscr{H}_M$. The model states all lie in the Hilbert space $\mathscr{H}_w \otimes \mathscr{L}$ where $\mathscr{L}$, defined above replaces $\mathscr{H}'$ defined by Eq. (41). The Schrödinger evolution in the eikonal approximation is given by Eqs. (11) and (31), where in Eq. (31), ψ_a^M is replaced by θ_a. Here and from now on, to conserve on notation $\theta_{\psi_a^M}$ is re-

placed by θ_a.

If, following Eq. (65), one defines

$$\psi^{\mathscr{L}}_{k_0}(x) \simeq \exp\left[(im/\hbar^2 k_0)\int_{-\infty}^{x} V(x' - n(x)d)\,dx' H'_T\right] \times [\exp(iKH'_T)]^{n(x)-1}\theta_a\,, \tag{78}$$

then as discussed in Sec. IV, for all times t in the interval Δ centered on $((m+\frac{1}{2})d - x_0)/v_0$, $\psi^{\mathscr{L}}_{k_0}(x)$ is independent of the relevant values of x and can be replaced in Eq. (64) by $\psi^{\mathscr{L}}_{k_0}(m) = (\exp(iKH'_T)^m\theta_a$. From Eqs. (75) and (77), $\psi^{\mathscr{L}}_{k_0}(m) = V_T^m\theta_a$ where, as noted above, this state contains the history of the first m model steps.

If one compares the three step operators, U_1, U_2, and U_3 on $\mathscr{H}_M \otimes \mathscr{H}_R \otimes \mathscr{H}_{h_R}$ as described in Sec. IV to the extension and expansion of U_T on $\mathscr{H}_M$ as described above, it is clear that the former is not minimal. In particular, the discussion above suggests that the recording system R and the space $\mathscr{H}_R$ are probably sufficient for construction of a unitary extension of U_T on $\mathscr{H}_M$ and that h_R and M as distinct systems with $\mathscr{H}_{h_R}$ and $\mathscr{H}_M$ could be discarded.

However, the extension constructed in Sec. IV has two advantages over the minimal one. One is that for the minimal extension Eq. (77) gives a condition which any interaction operator H'_T must satisfy. It is however, not clear how to express H'_T in terms of simple operators acting on $\mathscr{L}$, or if such an expression is possible at all. In the model described in Sec. IV, the three types of step operators H_1, H_2, and H_3, all have explicit expressions in terms of simple operators acting on $\mathscr{H}'$ [Eqs. (57)–(59)].

Another disadvantage of the minimal extension is that for any H'_T which satisfies Eq. (77), the locality requirement is not satisfied. That is, as discussed in Sec. II, for $0 < \delta < K$ one would like

$$e^{i\delta H'_T}V_T^m\theta_a = \alpha(\delta)V_T^m\theta_a + \beta(\delta)V_T^{m+1}\theta_a \tag{79}$$

to hold for each m and for each a for which $(V_T^m\theta_a, V_T^{m+1}\theta_a) = 0$. The conditions that $\alpha(\delta)$ and $\beta(\delta)$ should satisfy are given for Eq. (44). Note that $(V_T^m\theta_a, V_T^{m+n}\theta_a) = 0$, if and only if $T^n(a) \neq a$.

However, as noted in Sec. II, such an operator is not unitary. This can be proved by noting that from Eq. (79), $(\exp(i\delta H'_T)V_T^m\theta_a, \exp(i\delta H'_T)V_T^{m+1}\theta_a) \neq 0$ unless $T^2 = 1$ which is not the case in general. Thus, the minimal extension does not satisfy the locality requirement, in contrast to the extension described in Sec. IV. Whether or not the expansion of $\mathscr{H}_M$ to $\mathscr{H}_M \otimes \mathscr{H}_R \otimes \mathscr{H}_{h_R}$ and U_T to the three operators U_1, U_2, and U_3 of Eqs. (54)–(56) is minimal if one also requires the locality condition to be satisfied is an open question.

Finally, one notes that the minimal extension is restricted to those maps T which are boundedly many-one. Without this requirement, one can have U_{T_j} unbounded for some A_j in the partition, which invalidates the definition of W_T in Eq. (74). No such restriction is necessary for the extension discussed in Sec. IV.

VI. DISCUSSION

There are several other aspects of the results obtained here which should be noted. One is that Hamiltonian models such as the ones developed here have the property that no energy is dissipated as the overall system evolves—w begins and ends the multiple scattering with energy $\hbar^2 k_0^2/2m$. This holds independently of whether the process T is one-one or many-one and is simply a result of energy conservation for systems evolving under the Schrödinger equation with a self-adjoint Hamiltonian.

For which processes T, if any, the quantum-mechanical Hamiltonian models described here are good approximate descriptions of actual physical systems plus external fields, is difficult to say. The overall states $\psi(x,t)$ are very complex as is the Hamiltonian H^T_{3n}. Furthermore, because of the simplifying assumptions embodied in the choice of H_0, Eq. (7), any system which is described by the models discussed here would have to be operated close to 0°K to minimize spin flips of any of the noninteracting spin systems by unwanted outside stray radiation.

Another aspect of the models constructed here is that the interaction operator $H^T(j)$ given by Eq. (63) is j-dependent. This j-dependence can be removed if desired by letting each of the scattering centers of C be a fixed spin-1 system and defining $H^T(j)$ by

$$H^T(j) = P_{+1j}H_1 + P_{0j}H_2 + P_{-1j}H_3\,,$$

where H_1, H_2, and H_3 are given by Eqs. (57)–(59) and the P_{ij} are projection operators for finding the system c_j in C with spin projection i. In this way, the choice of which of the three types of interaction is to occur depends on a system spin projection rather than a lattice position. Of course, C would then have to be set up so that c_j has spin projection $P(j)$ where $P(j) = +(0)[-]$ if $j = 1(2)[0]$ mod 3.

The models described here have expanded each step of the original process into three steps: record, process, and shift. As far as the mathematics is concerned, one can also collapse these three steps into one step by defining U_T by $U_T = U_3U_2U_1$ or by Eqs. (54)–(56):

$$U_T = \sum_{\substack{a\in A \\ y\in A_n}} \sum_{kk'\in Z} U_y P_a \otimes P^R_{k'} P_{yk} U_{ak} \otimes U^{k'}_{+1} P_k\,.$$

In such a model U_T becomes the single step unitary operator which represents the transformation to be carried out by one w-c scattering.

In order to obtain an interaction operator, for each real number δ define $U_T(\delta)$ by the above equation where U_y, U_{ak}, and $U^{k'}_{+1}$ are replaced by U^{δ}_y, U^{δ}_{ak}, and $U^{k'\delta}_{+1}$ of Eqs. (43), (47), and (52). Since the map $\delta \to U_T(\delta)$ defines a one-parameter group of strongly continuous unitary transformations, there is, by Stone's theorem, a unique self-adjoint operator H_T such that $U_T(\delta) = \exp(i\delta H_T)$.

The problem here is that even though the existence of H_T is guaranteed by Stone's theorem, the existence is implicit. One has no idea how to express H_T explicitly as a simple function of simple operators as was done for H_1, H_2, and H_3. It would seem that H_T should be kept simple if the model is

to be a good description of some actual physical situation.

Finally, it is to be noted that one reason the model of the extended system is so complex is that only one model of $M + R + h_R$ has been given, which is used for all n and all processes T. All model dependence on T is contained in the dependence of the Hamiltonian $H_T(j)$ on T and all dependence on the number of steps $3n$ for which the model is applicable is contained in the overall Hamiltonian, Eq. (62), and in the length of C.

If the model $M + R + h_R$ is allowed to depend on n then its construction can be simplified. For example, R can be modeled by a finite-lattice (from $-n$ to n) of systems r_j rather than an infinite Z lattice, and all states of h_R can be restricted to lie in $l^2(-n,n)$ with appropriate changes in the definition of $U^k{}_{+1}$. If one further wishes to let the model of $M + R + h_R$ depend on T and also restrict the set of initial states in A to a finite set and require that h_R be initially at position 0, then both M and r_j can be modeled by single spin systems rather than infinite lattices. However, in the case of r_j, the spin of the jth system in the finite R lattice must then depend on j.

Thus, by relaxing the construction as defined, one can obtain an n- and T-dependent model of $M + R + h_R$ which is completely finite. This was the type of model constructed in *I*, where extended models of Turing machine computation processes were constructed. The price one pays for this model simplicity is the n- and T-dependence and restricted initial conditions.

ACKNOWLEDGMENT

The author wants to express his appreciation for the support and hospitality of the Institute for the History and Philosophy of Science at the University of Tel Aviv, where part of this work was done, during a semester's stay.

APPENDIX

This goal is to prove that for each map T from A to A where T can be many-one and into there exists a unique partition of A into subsets A_j for $j = 1,2,...$ which are T-invariant and irreducible. That is, $TA_j \subseteq A_j$ and for no nontrivial subset B of A_j does one have $TB \subseteq B$ and $T(A_j - B) \subseteq A_j - B$.

To prove this, first define for each T the *set* function T^{-1} by $T^{-1}(B) = \{a | T(a) \in B\}$ for each subset B of A. Note that $T^{-n}(B) = \{a | T^n(a) \in B\}$ for each $n = 0,1,2,...$. Choose an a_1 in A. For each pair of numbers $n,j = 0,1,2,...$ define A^1_{nj} by $A^1_{nj} = T^{-n}(\{T^j(a_1)\})$. It is clear that $A^1_{nj} = \{a' | T^n(a') = T^j(a_1)\}$. Define A_1 by

$$A_1 = \bigcup_{n,j \in N} A^1_{nj}.$$

Choose an a_2 in $A - A_1$ and repeat the above process to obtain A_2. Continue in this manner selecting a_3 in $A - (A_1 \cup A_2)$ etc. until there are no more elements left to choose.

Let a be in A_l. Then $a \in A^l_{nj}$ for some pair n,j or $T^n(a) = T^j(a_l)$. This implies that $T^{n-1}(T(a)) = T^j(a_l)$ or $T(a) \in A^l_{n-1\,j}$ if $n \geqslant 1$. If $n = 0$, $a = T^j(a_l)$ implies that $T(a) = T^{j+1}(a_l)$ or $T(a) \in A^l_{0j+1}$. In either case, $T(a) \in A_l$ which gives the result that $TA_l \subseteq A_l$. Furthermore, it is clear that A_l is irreducible. To see this, assume there is some nonempty subset B of A_l such that $TB \subseteq B$ and $T(A_l - B) \subseteq A_l - B$ where $A_l - B$ is also nonempty. Suppose $a_l \in B$, let n,j be such that $A^l_{nj} \cap (A_l - B)$ is not empty, and let a' be an element of the intersection. Then $T^n(a') = T^j(a_l)$. But $a_l \in B$ implies $T^j(a_l) \in B$ and $a' \in A_l - B$ implies $T^n(a') \in A_l - B$, which gives a contradiction. If $a_l \in A_l - B$, a similar argument also yields a contradiction. So A_l is T-irreducible.

The uniqueness of the sets A_l follows from the fact that the definition of A_l shows that it is independent of which element a_l in A_l is chosen for the construction. Thus, another choice function generates the same sets but possibly in a different order.

[1] P. A. Benioff, J. Stat. Phys. **22**, 563 (1980).

[2] J. S. Bell, Helv. Phys. Acta. **48**, 93 (1975); K. S. Hepp, Helv. Phys. Acta **45**, 237 (1972).

[3] M. Davis, *Computability and Unsolvability* (McGraw-Hill, New York, 1958).

[4] G.C. Emch, Helv. Phys. Acta **45**, 1049 (1972).

[5] B. Whitten-Wolfe and G. C. Emch, Helv. Phys. Acta **49**, 45 (1976).

[6] L. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1968), pp. 339–41; R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Sec. 18.3.

[7] Bela Sz. Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space* (North-Holland, New York, 1970), Chap. 1. See also Peter Fillmore, *Notes on Operator Theory* (Van Nostrand, New York, 1970), Chap. 7.

[8] Ordinary motion of a quantum mechanical wave packet from x to y is satisfactory in this respect. If a packet is localized around x initially, it does not spread out over all space when it is between x and y and then relocalize around y when the system arrives at y.

[9] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961), pp. 217–20.

[10] A. Messiah, *Quantum Mechanics*, Vol. I (Wiley, New York, 1961), pp. 218–22.

[11] J. Von Neuman, "On Infinite Directed Products," in *Collected Works*, Vol. III, edited by A. H. Taub (Pergamon, New York, 1963), pp. 323–99; M. A. Guichardet, Ann. Sci. Ecol. Norm. Super. **83**, 1 (1966).

The Lipkin model and coherent states

Debajyoti Bhaumik
Physics Department, Bose Institute, 93/1, Acharya Prafulla Chandra Road, Calcutta-700009, India
Ajoy Choudhury
Saha Institute of Nuclear Physics, 92, Acharya Prafulla Chandra Road, Calcutta-700009, India
Mira De and Binayak Dutta Roy
Physics Department, Bethune College, Calcutta, India

(Received 8 January 1980; accepted for publication 29 February 1980)

It is shown that the Bloch or angular momentum coherent states furnish a particularly efficacious basis for a discussion of various aspects of the Lipkin model of the "nucleus." The Hartree–Fock description (as well as its projected version) is elegantly obtained in this framework. It is demonstrated that the "transition probability" between the first excited and ground states is proportional to the square of the number of "nucleons," representing (in contrast to what obtains in the random phase approximation) a cooperativity of the "super-radiant" type. The extension of the model through the introduction of bosons permits, with the use of Bloch and Glauber coherent states, a succinct description of the phenomenon of boson condensation.

PACS numbers: 03.65.Ca, 21.60.Jz, 42.50. + z

I. THE LIPKIN MODEL

The Lipkin model[1] is an exactly soluble model of a system of N fermions occupying two levels (each possessing an N-fold degeneracy) separated by an energy spacing ϵ. Let $a_{p,\sigma}$ be the annihilation operator for a particle in the state labelled by the quantum number p (enumerating the substates 1,...,N in each level or "shell") and σ (adopting values ± 1) is the dichotomic level index. The Hamiltonian of the system is written as

$$H = \tfrac{1}{2}\epsilon \sum_{p,\sigma} \sigma a^{+}_{p,\sigma} a_{p,\sigma} + \tfrac{1}{2} V \sum_{p,p',\sigma} a^{+}_{p,\sigma} a^{+}_{p',\sigma} a_{p',-\sigma} a_{p,-\sigma}\,, \tag{1}$$

where a two-body "monopole–monopole" interaction (of strength V), scattering a pair of particles from one "shell" to the other, without changing the "subshell" quantum number p, is introduced. In terms of "quasispin" operators

$$J_{\pm} \equiv \sum_{p} a^{+}_{p,\pm 1} a_{p,\mp 1}\,, \tag{2a}$$

$$J_z \equiv \tfrac{1}{2} \sum_{p,\sigma} \sigma a^{+}_{p,\sigma} a_{p,\sigma}\,. \tag{2b}$$

Satisfying angular momentum commutation relations, the Hamiltonian may be cast into a particularly elegant form, to wit,

$$H = \epsilon J_z + \tfrac{1}{2} V (J^2_{+} + J^2_{-})\,. \tag{3}$$

It may be remarked that, m, the eigenvalue of J_z is simply half the difference between the number of particles in the upper and the lower states, and consequently its maximum value, namely j, equals $N/2$. It is clear that the unperturbed ($V = 0$) ground state is $|j = N/2, m = -N/2\rangle$ possessing the unperturbed energy $E_0 = -\epsilon N/2$. The interaction mixes states within the same j-multiplet corresponding to different numbers of "hole–particle" pairs.

The Hamiltonian exhibits the following symmetries (a) invariance under a rotation of π about the z axis in quasi-spin space, (b) $H \rightarrow -H$ under a rotation of π about an axis lying in the XY-plane making an angle of $\pi/4$ with respect to the X and Y axes.

II. THE BLOCH STATES

The familiar angular momentum states $|jm\rangle$, with a eigenvalues $j(j+1)$ and m for the operators J^2 and J_z, are superposed to construct a convenient basis, variously called coherent atomic states, Bloch, or Radcliffe states,[2] thus

$$|\alpha, j\rangle = \mathcal{N} \exp(\alpha J_{+}) |j, -j\rangle = (1 + |\alpha|^2)^{-j} \alpha^{j} \sum_{m=-j}^{+j} \alpha^{m} \sqrt{\binom{2j}{j+m}}\, |jm\rangle, \tag{4}$$

where α is, in general, a complex parameter specifying the state, $\mathcal{N}$, the appropriate normalization factor, and $[\binom{2j}{j+m}]$, the binomial combinatorial. This basis has been extensively used[3] in various areas of quantum optics such as laser theory, super-radiance, and resonance propagation.

III. THE LIPKIN MODEL AND BLOCH STATES

Employing the Bloch states as a basis, the expectation values of the Lipkin Hamiltonian become

$$\langle \alpha, j | H | \alpha, j \rangle = -\epsilon j (1 - |\alpha|^2)/(1 + |\alpha|^2) + Vj(2j+1)(\alpha^2 + \alpha^{*2})/(1 + |\alpha|^2)^2, \tag{5}$$

and introducing $\rho \equiv V(2j-1)/\epsilon$, minimization with respect to α yields the solutions

$$\alpha^2 = (\rho + 1)/(\rho - 1) \quad \text{for } \rho < -1, \tag{6a}$$

$$\alpha^2 = -\alpha^2_m = -(\rho - 1)/(\rho + 1) \quad \text{for } \rho > +1, \tag{6b}$$

and

$$\alpha^2 = 0 \quad \text{when } -1 < \rho < +1. \tag{6c}$$

Regarding ρ, which is proportional to the interaction strength and the number of particles, as a control parameter, it is thus seen that for $\rho > +1$ (as also, *mutatis mutandis* for $\rho < -1$) we have arrived at a minimum, which is immediately recognized to be the "deformed" Hartree–Fock ground state as obtained by Agassi,[4] possessing the energy

 022-2488/81/030508-03$1.00

$$E_0 = -(N\epsilon/2)\cos\chi_m - [N(N-1)/4]V\sin^2\chi_m\,, \tag{7a}$$

through the identification

$$\cos\chi_m = (1-\alpha_m^2)/(1+\alpha_m^2) = \epsilon/[V(N-1)]. \tag{7b}$$

The corresponding ground state is best expressed as

$$|\alpha_m, j\rangle = \mathcal{N}\prod_{p=1}^{N} b_p^+ |\text{vacuum}\rangle, \tag{8}$$

where b_p is a "rotated" single particle operator defined by

$$b_p = [\cos(\chi_m/2)]a_{p,-1} - i[\sin(\chi_m/2)]a_{p,+1}\,. \tag{9}$$

The deformed Hartree–Fock description obtains for potential strengths greater than a certain critical value given by $V_c = \epsilon/(N-1)$, while (within the realm of Bloch states for $|V| < V_c$) it is the unperturbed ground state ($\alpha = 0$) which is stabilized, and then the situation is in fact better described by the random phase approximation (RPA). The different branches of the roots of the minimization condition realized in various ranges of the control parameter ρ is depicted through the bifurcation diagram shown in Fig. 1.

Confining our attention to the branch corresponding to $\rho > 1$ wherein $\alpha^2 = -\alpha_m^2$, it may be observed that the states $|i\alpha_m, j\rangle$ and $|-i\alpha_m, j\rangle$ are degenerate. The minimum at $\alpha = 0$ for the region $|\rho| < 1$ splits into these two minima much like what occurs in a typical broken symmetry situation, for example in ϕ^4 field theory, as the coefficient of quadratic "mass" term changes sign. The Lipkin Hamiltonian through the interaction term mixes only an even number of particle–hole pairs, as a consequence of the symmetry of the Hamiltonian. Nevertheless, the states $|i\alpha_m, j\rangle$ and $|-i\alpha_m, j\rangle$ contain arbitrary numbers (even as well as odd) of such pairs, and do not possess the symmetry enjoyed by the Hamiltonian, and the corresponding symmetry is broken. However the gerade (or symmetric) combination of states

$$|\alpha, j; g\rangle \equiv \mathcal{N}_+ [|\alpha, j\rangle + (-1)^{2j}|-\alpha, j\rangle], \tag{10}$$

has a lower energy and represents a better candidate for the ground state. $\mathcal{N}_+$ is the normalization constant. The expectation value of the Hamiltonian is

$$\langle\alpha, j; g\,|\,H\,|\,\alpha, j; g\rangle = \langle\alpha, j|H\,|\alpha, j\rangle[(1+\cos^{2j-2}\chi)/(1+\cos^{2j}\chi)], \tag{11}$$

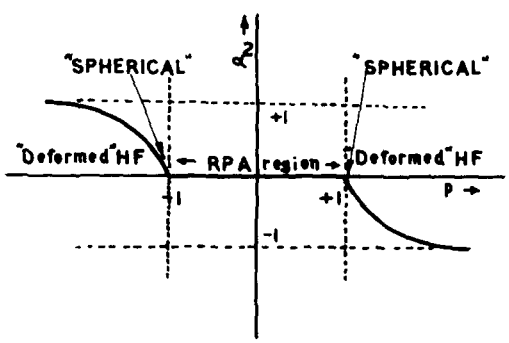


FIG. 1. Bifurcation diagram showing stable branches.

where $\cos\chi$ is defined as $(1-\alpha^2)/(1+\alpha^2)$. Inserting for χ the value χ_m given by Eq. (7b), the resulting expression for the ground state energy is identical to the one obtained by Agassi[4] and called the "projected" Hartree–Fock result. The corresponding state differs from the "unprojected" wave function through the absence of admixtures of odd numbers of particle–hole pairs. For large values of $|\alpha|$ however, $\sin\chi \to 0$ and the projected and unprojected versions become identical.[5]

Just as the ground state was obtained as the symmetric combination of Bloch states, the present approach admits of a simple description of the first excited state, through the orthogonal antisymmetric (or ungerade) combination,

$$|\alpha, j; u\rangle = \mathcal{N}_-[|\alpha, j\rangle - (-1)^{2j}|-\alpha, j\rangle], \tag{12}$$

which has odd numbers of hole–particle pairs unlike the gerade (or ground) state which has even numbers of the same. The nature of correlations present in the projected ground and excited states, manifested through the basis chosen, is further revealed by the study of the transition probability between these states. Within the confines of this model the simplest "transitions" are caused by what may be called the "monopole" transition operators J_x and J_y, so named because they do not change the p-quantum number. The relevant transtion matrix elements are readily calculated to yield

$$\langle\alpha, j; g\,|J_x\,|\alpha, j; u\rangle = j\sin\chi\,[1-(\cos\chi)^{4j}]^{-1/2}, \tag{13a}$$

$$\langle\alpha, j; g\,|J_y\,|\alpha, j; u\rangle = 0, \tag{13b}$$

for the case where α is real. This, and other results, for real α may readily be extended to complex α through the observation

$$\exp(i\theta J_z)|\,\alpha, j\rangle = \exp(i\theta j)|\alpha\exp(i\theta), j\rangle \tag{14}$$

following from the definition, Eq. (4), of Bloch states, where the operation involved is a rotation by an angle θ about the z axis in quasi-spin space. In the limit of large N, as also in the strong-coupling limit (V large), it may be seen from Eq. (7b), that $\cos\chi \to 0$ and the transition matrix element becomes proportional to N. The transition probability is thus proportional to the square of the number of particles, in sharp contrast to what occurs in the random phase approximation where the transition probability is proportional to the number of particles. Indeed, a strong parallelism exists between what has been obtained here and what occurs in the theory of super-radiance in atomic physics,[6] wherein a collection of N atoms initially prepared in some excited state under suitable conditions return to their ground states by emitting electromagnetic radiation, of intensity I, proportional to the square of the number of atoms, in a super-radiant pulse, in contrast to the normal situation where $I \sim N$. This phenomenon has been extensively discussed (in the atomic context) in terms of Bloch states.

It is instructive to consider an extended version of the Lipkin model where the Hamiltonian is

$$H = \epsilon J_z + (V/2)(J_+^2 + J_-^2) + (w/2)(J_+J_- + J_-J_+). \tag{15}$$

Minimization of the expectation value of this Hamiltonian in the Bloch state yields, for the parameter α, the values

$$\alpha^2 = 0, \quad (\rho + \gamma + 1)/(\rho + \gamma - 1)$$

and

$$-(\rho - \gamma - 1)/(\rho - \gamma + 1), \tag{16}$$

for control parameters satisfying $|\rho + \gamma| < 1$, $|\rho + \gamma| < -1$, and $(\rho - \gamma) > 1$ respectively, where $\gamma \equiv (2j-1)w/\epsilon$ is defined analogous to ρ. A perusal of the minimization condition reveals that in the absence of the V term ($\rho = 0$) there exists a degeneracy with respect to the phase of α which is a consequence of the invariance of the Hamiltonian (when $V = 0$) with respect to arbitrary rotations in the quasi-spin space about the Z axis. This symmetry is broken by the very presence of V and the phase of α can then only adopt the values 0 and $\pm \pi/2$.

IV. THE LIPKIN MODEL AND BOSON CONDENSATION

Boson condensation (for pions and scalar bosons) in nuclear matter (and particularly for neutron stars) has received considerable attention in recent years.[7] The Lipkin two-level model may be extended[8] with the inclusion of bosons to provide an exactly-soluble system enabling an instructive discussion of the phenomenon of boson condensation. Introducing the boson annihilation operator b, and implementing the "nucleonic" degrees of freedom through the quasispin formalism, the system is taken to be governed by the Hamiltonian

$$H = \epsilon J_z + (V/2)(J_+^2 + J_-^2) + (w/2)(J_+ J_- + J_- J_+) + \omega b^+ b + G(J_+ b + J_- b^+), \tag{17}$$

where ω stands for the energy of a boson and G is the coupling constant of the boson to the nucleonic degree of freedom. The appropriate basis for the discussion of this system will be taken to be $|\alpha, j; \beta\rangle$ where α and j describe the state of the "nucleons" as before, while β specifies the coherent state, *á lá* Glauber,[9] corresponding to the bosonic degree of freedom, and thus

$$|\alpha, j; \beta\rangle = |\alpha, j\rangle \exp(-\beta^* b + \beta b^+)\,|0\rangle. \tag{18}$$

It is readily verified that

$$\langle \alpha, j; \beta | H | \alpha, j; \beta \rangle = (\epsilon + w) j + 2j\epsilon[\rho(\alpha^2 + \alpha^{*2})/2 + (\gamma - 1)|\alpha|^2 - 1]/(1 + |\alpha|^2)^2 + \omega|\beta|^2 + G\,2j(\beta\alpha + \beta^*\alpha^*)/(1 + |\alpha|^2), \tag{19}$$

and the extremization condition yields

$$\beta = -(Gj\omega)\alpha^*/(1 + |\alpha|^2), \tag{20a}$$

$$\alpha\{(1 + |\alpha|^2) + (\gamma - \xi)(1 - |\alpha|^2)\} + \rho(\alpha^* - \alpha^3) = 0, \tag{20b}$$

where $\xi \equiv (G^2 2j)/\omega\epsilon$. It may be remarked that with $V = 0$ ($\rho = 0$) the Hamiltonian is symmetric under arbitrary rotations about the Z axis in quasispin space accompanied by a corresponding phase change of the boson operators. Thus the ground state then possesses degeneracy on a circle in the complex α and β planes, constrained by the relation given by Eq. (20a), provided $|\gamma - \xi| \not< 1$. The very presence of V however destroys this invariance and forces α to pick a definite phase which is 0 or $\pm \pi/2$ depending on whether $\rho + \gamma - \xi < -1$ or $\rho + \xi - \gamma > 1$. For $|\rho + \gamma - \xi| < 1$, however, the minimum resides at $\alpha = \beta = 0$ which corresponds to the vacuum state of the boson sector. Consequently, the onset of boson condensation occurs when the relevant control parameter ($\xi - \rho - \gamma$) exceeds unity, confining our attention to the range of parameters where real values of α are realized. Correspondingly, noting that $2j = N =$ the number of nucleons, boson condensation takes place for nucleon number above a critical threshold value given by

$$N_{\text{critical}} = [\epsilon - (v + w)]/[G^2/\omega - (v + w)]. \tag{21}$$

It is instructive to note that boson condensation, in this model, is concommitant with the transition from what we have called the RPA region to the deformed Hartree–Fock regime as far as the nucleons are concerned. This is analogous to the modification[7] of the nucleonic Fermi surface in the field of the pion condensate. It may also be noted that the presence of attractive "internucleon" potentials lowers the value of the critical nucleon number necessary for boson condensation. The energy of the system corresponding to the minima in the region of interest is given by

$$E = (\epsilon + w)j - 2j\epsilon(1 + \xi - \rho - \gamma)^2/4(\xi - \rho - \gamma). \tag{22}$$

The variational energy may be driven to a somewhat lower value by invoking the gerade state

$$|\alpha, j; \beta g\rangle = [|\alpha, j; \beta\rangle + (-1)^{2j}|-\alpha, j; -\beta\rangle]/[2\{1 + (\alpha^2 - 1)^{2j}/(\alpha^2 + 1)^{2j} \times \exp(-2\beta^2)\}]^{1/2}. \tag{23}$$

V. CONCLUSION

We have described the Lipkin model as well as its extension to include bosons in the space of Bloch or angular-momentum coherent states and the Glauber coherent states and have thereby obtained the Hartree–Fock approximation and its projected version. The transition amplitude between the excited and ground states exhibits a cooperativity characteristic of the states analogous to super-radiance in atomic physics. The possibility of boson condensation in the model is also readily described in this framework.

ACKNOWLEDGMENTS

We would like to thank Dr. Gautam Ghosh and T.S. Nag for various discussions and also Dr. Brahmananda Dasgupta for suggestions in the style of writing.

[1] H. J. Lipkin, V. Meshkov, and A. J. Glick, Nucl. Phys. **62**, 188, 199, 211 (1965).

[2] F. Bloch, Phys. Rev. **70**, 460 (1946); J. M. Radcliffe, J. Phys. A Gen Phys. **4**, 313 (1971); D. Bhaumik, T. S. Nag, and B. Dutta Roy, J. Phys. Gen. Phys. A **8**, 1868 (1975).

[3] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, in *Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf (Plenum, New York, 1972); F. T. Arecchi in *Co-operative Effects, Progress in Synergetics*, edited by H. Haken (North-Holland, Amsterdam, 1974), p. 33.

[4] D. Agassi, H. J. Lipkin, and N. Meshkov, Nucl. Phys. **86**, 321 (1966).

[5] J. C. Parikh and D. L. Rowe, Phys. Rev. **175**, 1293 (1968).

[6] R. J. Glauber and F. Haake, in *Co-operative Effects, Progress in Synergetics*, edited by H. Haken (North-Holland, Amsterdam, 1974), p. 71.

[7] A. B. Migdal, Rev. Mod. Phys. **50**, 107 (1978).

[8] D. Schütte and J. da Provedencia, Nucl. Phys. A **282**, 518 (1977).

[9] R. J. Glauber, Phys. Rev. **131**, 2766 (1963).

Expansion around instantons in quantum mechanics

J. Zinn-Justin
CEN-Saclay, Boite Postale No. 2, 91190, Gif-Sur-Yvette, France

(Received 27 December 1979; accepted for publication 8 February 1980)

We calculate numerically a few terms of the corrections to the large-order behavior of the ground state energy of the $O(N)$ anharmonic oscillator by analyzing the perturbation series. We then generate 94 terms of the perturbative expansion of the difference between the energies of the two low lying states of the double-well potential and analyze their large-order behavior.

PACS numbers: 03.65.Db, 03.65.Sq

INTRODUCTION

In this article we want to present some results, based mainly on numerical explorations concerning the perturbative expansion around instantons in quantum mechanics. These results are a by-product of the analysis of the large-order behavior of the perturbative expansion of the ground state energy for various potentials.

In Sec. I we discuss the simple integral which counts the number of diagrams of the ϕ^4 theory with the proper weights.

In Sec. II we present the calculation of the coefficients of the perturbative expansion of the ground state energy of the anharmonic potential with $O(N)$ symmetry.

As we know, the large-order behavior of these coefficients is given by an instanton contribution. The numerical analysis of this large-order behavior yields, therefore, an expansion around this instanton. Due to limited numerical accuracy only a few terms are available.

In Sec. III we present a similar calculation for the double-well potential. We recall that the perturbative expansion of the ground state of the double potential has been discovered numerically to be identical up to the sign and a coupling normalization to the similar expansion for the $O(2)$ anharmonic oscillator.

We then explain how a formula found in Landau and Lifschitz allows one to expand systematically the difference between the ground state energy and the first excited state for a symmetric potential. For an analytic potential with two minima, this difference is given by an instanton contribution corresponding to a Euclidean path joining the two minima.

We apply this technique to the double-well potential. We calculate, numerically, 94 terms of the expansion around the instanton, and characterize its large-order behavior. We verify the asymptotic nature of this expansion by comparing the series to a numerical calculation of this energy difference from the Schrödinger equation.

I. THE SIMPLE INTEGRAL

We shall consider as a model for our study of quantum mechanics the simple integral which represents a zero-dimensional ϕ^4 field theory,

$$z(g) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-(x^2 + gx^4)}\, dx. \tag{1}$$

Actually, to be closer to the real problem we shall consider a different quantity $F(g)$,

$$F(g) = \langle x^4 \rangle, \tag{2}$$

$$F(g) = -\frac{\partial}{\partial g} \ln z(g). \tag{3}$$

If the perturbative expansion of $z(g)$ can be explicitly obtained, the expansion of $F(g)$ has a somewhat more complicated form. It can be generated most simply by remarking that $z(g)$ satisfies a differential equation

$$4g^2 z''(g) + (1 + 8g) z'(g) + \tfrac{3}{4} z(g) = 0. \tag{4}$$

Therefore, $F(g)$ satisfies a Ricatti equation

$$4g^2 [F'(g) - F^2(g)] + (1 + 8g) F(g) - \tfrac{3}{4} = 0. \tag{5}$$

From this equation we can derive a recursion formula for the coefficients of the Taylor series expansion of $F(g)$,

$$F(g) = \sum_0^\infty (-1)^k F_k g^k, \tag{6}$$

$$F_k = 4\left((k+1) F_{k-1} + \sum_{0 < n < k-2} F_n F_{k-n-2} \right), \tag{7}$$

$$\text{with } F_0 = \tfrac{3}{4}. \tag{8}$$

An amusing arithmetical remark is that $4F_k$ is an integer while the coefficients z_k of $z(g)$ are rational numbers with increasing denominators.

As we know the large-order behavior of the coefficients F_k and the coefficients z_k is given by the nontrivial saddle point of the integral (1).

More precisely, the saddle point x_c,

$$2x_c + 4gx_c^3 = 0 \rightarrow x_c^2 = -1/2g, \tag{9}$$

yields a small g expansion for the imaginary part $\operatorname{Im} z(g)$ of $z(g)$ for g negative, which gives then an expansion of z_k for k large as

$$z_k = \frac{1}{\pi} \int_{-\infty}^{0} \frac{\operatorname{Im} z(g)}{g^{k+1}}\, dg. \tag{10}$$

As $\operatorname{Im} z(g)$ behaves as $e^{1/4g}$ for g small and negative, the small g behavior of $\operatorname{Im} z(g)$ is directly related to the small g behavior of $\operatorname{Im} F(g)$,

$$\operatorname*{Im}_{g \to 0_-} F(g) = \frac{\operatorname{Re} z'(g) \operatorname{Im} z(g) - \operatorname{Re} z(g) \operatorname{Im} z'(g)}{[\operatorname{Re} z(g)]^2} + O(e^{1/2g}). \tag{11}$$

The real part $\operatorname{Re} z(g)$ of $z(g)$ is just given by the perturbative expansion. At leading order

 022-2488/81/030511-10$1.00

$$\operatorname*{Im}_{g\to 0_-} F(g) \sim -\operatorname*{Im}_{g\to 0_-} z'(g) \sim -\frac{4}{\pi\sqrt{2}}\frac{e^{1/4g}}{g^2}. \tag{12}$$

From the Ricatti equation we can obtain directly an equation for Im $F(g)$

$$4g^2[\operatorname{Im}F(g)' - 2\operatorname{Im}F(g)\operatorname{Re}F(g)] + (1+8g)\operatorname{Im}F(g) = 0, \tag{13}$$

$$\operatorname{Im}F(g) = -\frac{4}{\pi\sqrt{2}}\frac{e^{1/4g}}{g^2}\exp 2\int_0^g \operatorname{Re}F(g')\,dg'. \tag{14}$$

If we neglect exponentially small corrections we can replace $\operatorname{Re}F(g)$ by the perturbative expansion of $F(g)$. Therefore, the small g expansion of Im $F(g)$, which yields an expansion of F_k for k large, is directly related to the perturbative expansion of $F(g)$ itself. In addition this relation is particularly simple on ln Im $F(g)$.

Also, the coefficients of the expansion of $\ln[\operatorname{Im} F(g)]$ have simple arithmetical properties.

II. THE ANHARMONIC OSCILLATOR WITH $O(N)$ INTERNAL SYMMETRY

We shall consider now the ground state energy of the anharmonic oscillator with $O(N)$ internal symmetry whose Hamiltonian H is given by

$$H = \tfrac{1}{2}\mathbf{p}^2 + \tfrac{1}{2}\mathbf{x}^2 + g(\mathbf{x}^2)^2. \tag{15}$$

The wavefunction of the ground state depends only of $|\mathbf{x}|$. It is convenient to write the Schrödinger equation for the logarithmic derivative of the wavefunction $\psi(|\mathbf{x}|)$. Denoting simply by x the radial coordinate $|\mathbf{x}|$,

$$f(x) = -\frac{d}{dx}\ln\psi(x). \tag{16}$$

Then $f(x)$ satisfies the Ricatti equation

$$f'(x) - f^2(x) + [(N-1)/x]f(x) + x^2 + 2gx^4 = 2E, \tag{17}$$

in which $E(g)$ is the ground state energy. To generate a perturbative expansion of $E(g)$ it is more convenient to use this equation rather than the direct Schrödinger equation because the expansion of $f(x,g)$ involves much fewer terms than the expansion of ψ. As a result the calculations are simpler and the accuracy is improved.

Let us expand $f(x,g)$ and $E(g)$

$$f(x,g) = x + \sum_1^\infty (-g)^{k+1} f_k(x), \qquad E(g) = \sum_0^\infty E_k g^k. \tag{18}$$

We get the recursion formula

$$2(-1)^{k+1}E_k = f_k'(x) + [(N-1)/x - 2x]f_k(x) + \sum_{1\le l\le k-1} f_l f_{k-l} + 2x^4\delta_{k1}, \tag{19}$$

for $k > 0$ and with

$$E_0 = N/2. \tag{20}$$

Setting then

$$f_k(x) = \sum_{n\ge 0} C_k^n x^{2n+1}, \tag{21}$$

we obtain a recursion formula for the C_k^n,

$$C_k^{n-1} = \delta_{n2}\delta_{k1} + \left(n + \frac{N}{2}\right)C_k^n + \frac{1}{2}\sum_{\substack{1\le l\le k-1\\ 0\le m\le n-1}} C_l^m C_{k-l}^{n-m-1}, \tag{22}$$

and the coefficients E_k, then, simply by

$$E_k = (N/2)(-1)^{k+1}C_k^0. \tag{23}$$

The recursion formula (22) shows that the C_n^k are all positive rational numbers with as denominators powers of 2 (for N integer, of course). More precisely $2^{2k-n-1}C_k^n$ is an integer for N odd and $2^{k-1}C_k^n$ is an integer for N even.

Notice that C_k^n is nonzero only for $n \leqslant k$. It is possible starting from (22) to derive simple bounds on the C_k^n, which we shall not discuss here.

A systematic WKB expansion of Eq. (17) generates first the coefficients of highest degree at each order in g and then successively the other coefficients. For example,

$$f^{(0)}(x) = \sqrt{x^2 + 2gx^4} = x\left(1 + \sum_1^\infty (-1)^{k+1}(gx^2)^k C_k^k\right), \tag{24}$$

yields

$$C_k^k = \frac{1\cdot 3\cdots(2k-3)}{k!}, \quad \text{for } k \geqslant 2. \tag{25}$$

The next term in the WKB expansion is a generating function for C_k^{k-1}, etc.

We have used these recursion formulas in the past to generate numerically the perturbative expansion of E_k up to order 100, typically, to verify the large-order behavior results[1,2] and to study resummation methods for these divergent series.[3,4] The results we present in this section are just by-products of these studies.

A. Large-order behavior

A steepest descent calculation[2] of the path integral expression for $\operatorname{Tr} e^{-\beta H}$ leads to a small g expansion of the imaginary part of $E(g)$ for g negative. Using then the integral representation

$$E_k = \frac{1}{\pi}\int_{-\infty}^0 \frac{\operatorname{Im}E(g)}{g^{k+1}}\,dg, \quad k > 1, \tag{26}$$

one obtains an expansion, for k large, of the coefficients E_k. At leading order one finds

$$\operatorname*{Im}_{g\to 0_-} E(g) = \frac{1}{\Gamma(N/2)}\left(-\frac{2}{g}\right)^{N/2} e^{1/3g}[1 + O(g)], \tag{27}$$

and, therefore,

$$E_k = E_k^{\text{as}}[1 + O(1/k)],$$

$$E_k^{\text{as}} = (-1)^{k+1}\Gamma(k + N/2)3^{k+N/2}\frac{2^{N/2}}{\pi\Gamma(N/2)}. \tag{28}$$

The first numerical investigations of the successive corrections to expression (27) are due to Bender and Wu.[1] In addition, the $1/k$ term has been calculated analytically by these authors[1] for $N = 1$, and by Seznec[5] for arbitrary N.

We have investigated these corrections numerically for arbitrary N using about 100 terms of the perturbative expan-

sions and a modified Neville procedure. Details about this numerical study are given in Appendix A.

As a result we have obtained a fit of the form

$$E_k = E_k^{\text{as}}\left(1 + \frac{a_1(N)}{k+N/2-1} + \frac{a_2(N)}{(k+N/2-1)(k+N/2-2)} + \cdots\right). \tag{29}$$

This expansion yields then an expansion of Im $E(g)$ in powers of g,

$$\underset{g\to 0}{\text{Im}E(g)} = \frac{1}{\Gamma(N/2)}\left(-\frac{2}{g}\right)^{N/2} e^{1/3g} \times [1 - 3a_1(N)g + 9a_2(N)g^2 + \cdots + (-3)^n a_n(N)g^n + \cdots]. \tag{30}$$

This expansion can, in particular, be obtained by a systematic calculation of the instanton contribution to the path integral for $\text{Tr}e^{-\beta H}$.

The coefficients $a_n(N)$ are given by a set of Feynman diagrams with modified propagators and vertices due to the presence of the instanton field. But the weight factors of the diagrams are not modified, so that the coefficients $a_n(N)$ are polynomials in N of degree $2n$.

For a reason which will become clear later, it is convenient to exponentiate this last expansion and write Im $E(g)$ as

$$\text{Im}E(g) = \frac{1}{\Gamma(N/2)}\left(-\frac{2}{g}\right)^{N/2} \times \exp\left(\frac{1}{3g} + b_1(N)g + b_2(N)g^2 + \cdots\right). \tag{31}$$

It is in this form that we shall present the results. But before doing this, we shall evaluate some coefficients of $b_n(N)$, considered as polynomials in the variable N, analytically.

B. Large N expansion

We can obtain the two coefficients of highest degree of the polynomial $b_n(N)$ from the large-order behavior of the $1/N$ expansion as calculated by Brezin and Hikami[6] and De Vega.[7]

The reason for this is the following: The $1/N$ expansion is obtained in a limit in which the product gN is kept fixed.

As a result, the large-order behavior of the $1/N$ expansion is obtained from the estimate for N large and negative of the imaginary part of $E(g,N)$ for g negative and N positive. As $E(g,N)$ is not only a function of g/N, it has also some singularities at finite values of N which probably yield additional contributions to the imaginary of $E(g,N)$. But these contributions vanish identically for $|N|$ larger than some finite value, and do not affect, therefore, the large N expansion of Im$E(g,N)$ at large order.

One can translate the results of Refs. 6 and 7 in terms of an expansion for Im $E(g,N)$. Setting

$$\lambda = Ng, \tag{32}$$

they can be written

$$\underset{N\to-\infty}{\text{Im}E(\lambda,N)} = \frac{1}{\Gamma(N/2)}\left(-\frac{2N}{\lambda}\right)^{N/2} \times C(\lambda)e^{NA(\lambda)}\left[1 + O\left(\frac{1}{N}\right)\right]. \tag{33}$$

The functions $A(\lambda)$ and $C(\lambda)$ are given explicitly. For λ small they behave like

$$A(\lambda) = 1/3\lambda + (7/8)\lambda + O(\lambda^2), \tag{34}$$

$$C(\lambda) = 1 + (9/4)\lambda + O(\lambda^2). \tag{35}$$

If we expand $NA(\lambda) + \ln C(\lambda)$ in powers of λ and replace λ by Ng, we see that the polynomials $b_n(N)$ are of degree $n+1$ only, in contrast to the polynomials $a_n(N)$.

In addition, we obtain the two coefficients of highest degree b_n^{n+1} and b_n^n of $b_n(N)$.

$$b_n(N) = \sum_{l=0}^{n+1} b_n^l N^l$$

The function $A(\lambda)$ is given by

$$h(\lambda) = (1 + 2\lambda h)^{-1/2} = 1 + O(\lambda), \tag{36}$$

$$A'(\lambda) = -\frac{1}{3\lambda^2}(1 + 3\lambda h)^{-1/2} + \frac{1}{2\lambda} = -\frac{1}{3\lambda^2} + O(1). \tag{37}$$

In terms of the function $h(\lambda)$, the logarithmic derivative of $C(\lambda)$ is

$$\frac{C'(\lambda)}{C(\lambda)} = \frac{1}{\lambda}\frac{(1+2\lambda h)^{1/2}}{(1+3\lambda h)^{3/2}} - \frac{1}{2\lambda}\frac{(1+2\lambda h)}{(1+3\lambda h)} + \frac{21}{4}\frac{(1+2\lambda h)^{1/2}}{(1+3\lambda h)^2} - \frac{1}{2\lambda}. \tag{38}$$

The first ten terms of the expansion of these functions are given in Table I.

From these expressions, one can verify that the coefficients are alternating in sign with the order. In addition, from the arithmetical point of view, the coefficients of $3A'(\lambda)$ and $C'(\lambda)/C(\lambda)$ are rational numbers with only powers of 2 as denominators. More precisely 2^{2n+1} times the coefficients or order n is an integer. The persistence of this arithmetical property, although here it applies to the logarithmic derivative of the imaginary part, is quite remarkable.

C. The case $N = 0$

After this work was essentially completed, we saw an article[8] which also pointed out the relevance of the Ricatti equation and, in addition, noted that the case $N = 0$ could be solved exactly in terms of Airy functions.

We shall present here the part of the argument relevant for our purpose.

Let us take in Eq. (17) x^2 as a variable and set

$$z = x^2, \tag{39}$$

$$f(x) = xu(z). \tag{40}$$

The equation then becomes

$$2zu' + Nu - zu^2 + z + 2gz^2 = 2E. \tag{41}$$

We have already used in the recursion formula the relation

$$E = (N/2)u(0). \tag{42}$$

For $N = 0$ two remarkable facts occur: The ground state energy is exactly known as it vanishes and the equation simplifies drastically,

$$2u' - u^2 + 1 + 2gz = 0. \tag{43}$$

TABLE I. The first three lines give the value of $(-1)^{n+1}3n2^{n+1}/b_n'$. The last line gives the greatest common divisor (GCD) of these numbers.

n	1	2	3	4	5	6	7	8	9	10
$l=n+1$	21	246	3453	53226	868962	14749164	257473629	4591179714	83241320694	1529546073588
$l=n$	27	618	13572	292782	6255252	132848268	2810155032	59274603054	1247635616868	26217999638148
$l=0$	5	60	1105	27120	828250	30220800	1282031525	61999046400	3366961243750	202903221120000
GCD	1	6	1	6	2	12	1	6	2	12

Indeed, now $y(z,g)$ instead of being of function of z and g separately, is a function of only one variable. If we set

$$y = (1+2gz)/g^{2/3}, \tag{44}$$

and

$$u = g^{1/3}v(y), \tag{45}$$

then $v(y)$ satisfies the differential equation

$$4v' - v^2 + y = 0. \tag{46}$$

The function $v(y)$ is the logarithmic derivative of an Airy function. As we are interested in an expansion for g small, we have to expand $v(y)$ for y large

$$v(y) = \sqrt{y}\left(1 + \sum_1^\infty (-1)^{k+1} v_k y^{-3k/2}\right), \tag{47}$$

with the recursion formula

$$v_k = (3k-4)v_{k-1} + \frac{1}{2}\sum_{1\leqslant l\leqslant k-1} v_l v_{k-l}, \tag{48}$$

$k \geqslant 2$ with $v_1 = 1$.

Now, from Eq. (42), we get an interesting result,

$$\lim_{N\to 0} 2E(N)/N = \epsilon(g) = u(z=0, N=0, g), \tag{49}$$

with

$$u(0) = g^{1/3}v(1/g^{2/3}) = \epsilon(g), \tag{50}$$

so that

$$\epsilon(g) = 1 + \sum_1^\infty (-1)^{k+1} v_k g^k. \tag{51}$$

In this limit the coefficients of the perturbative expansion of the ground state energy are given by a recursion formula with only one index as in the case of the simple integral of Sec. I.

Furthermore, we can now study easily the large-order behavior of $u(0)$ and, therefore, obtain information on our polynomials $b_n(N)$ for $N = 0$.

To do this we shall write the differential equation for $\epsilon(g)$,

$$6g^2\epsilon' - 2g\epsilon + \epsilon^2 - 1 = 0. \tag{52}$$

For g negative, $\epsilon(g)$ is complex

$$\epsilon(g) = \alpha(g) + i\beta(g). \tag{53}$$

Taking the imaginary part of the equation we obtain

$$3g^2\beta'(g) - g\beta(g) + \alpha(g)\beta(g) = 0. \tag{54}$$

So that we can calculate $\beta(g)$ in terms of $\alpha(g)$,

$$\frac{\beta'(g)}{\beta(g)} = \frac{1}{3g} - \frac{\alpha(g)}{3g^2}, \tag{55}$$

or

$$\beta(g) = C\exp\frac{1}{3g} - \frac{1}{3}\int_0^g [\alpha(g') - 1 - g']\frac{dg'}{g'^2}. \tag{56}$$

For g small and negative, one can replace $\alpha(g)$ by the perturbative expansion of $\epsilon(g)$ up to exponentially small corrections. As a result, the logarithmic derivative of the imaginary part of $\epsilon(g)$ has an expansion directly proportional to the expansion of $\epsilon(g)$ itself.

TABLE II. Values of $(-1)^{n+1}3n2^{n+l}b_n^l$ for $n \leqslant 6$.

l \ n	1	2	3	4	5	6
0	5	60	1105	27120	828250	30220880
1	27	438	9720	270552	9038790	352590000
2	21	618	18663	629352	23947884	1025295480
3		246	13572	626742	29050770	1423155756
4			3453	292782	18669510	1118597184
5				53226	6255252	517369608
6					868962	132848268
7						14749164

From the recursion formula (48) we see that $2^{k-1}v_k$ is an integer. Therefore, $3n2^nB_n(0)$ is also an integer. The first ten coefficients of $B_n(0)$ are listed in Table I.

One could then imagine expanding in powers of N around $N=0$, we haven't done it, but an encouraging empirical remark can be made. From the comparison of the perturbative expansion of the ground state energy and the coefficients $b_n(N)$ it follows that

$$\frac{\beta_0'(g)+N\beta_1'(g)}{\beta_0(g)+N\beta_1(g)} = \frac{1}{3g} - \frac{\alpha_0(g)}{3g^2} - \frac{3\alpha_1(g)N}{3g^2} + O(N^2), \quad (57)$$

where we have defined

$$\frac{2E(g,N)}{N} = \alpha_0(g) + N\alpha_1(g) + i\beta_0(g) + iN\beta_1(g) + O(N^2, e^{2/3g}), \quad (58)$$

for $g = -|g| + i\epsilon$.

As a general comment, it should be emphasized that, as $\epsilon(g)$ satisfies a Ricatti equation, it has not the analytic structure of $E_N(g)$ for general N with an infinite number of branch points accumulating at the origin in the second sheet.[9]

This explains why this particular case can be solved so explicitly.

A last point: For all negative even integer values of N, one can find $E_N(g)$ as a solution of an algebraic equation. This can be most easily seen be differentiating systematically Eq. (41).

For example, if we differentiate once, we get

$$2zu'' + (N+2)u' - u^2 - 2zuu' + 1 + 4gz = 0. \quad (59)$$

Setting $z=0$ yields

$$(N+2)u'(0) - u^2(0) + 1 = 0. \quad (60)$$

So, for $N=-2$, we find $E(g)$,

$$E_{-2}(g) = (N/2)u(0) = -1. \quad (61)$$

One could, therefore, think to calculate the derivative of $E_N(g)$ with respect to N for $N=-2$, by solving the equation for $u(z)$ and finding $u'(0)$. Unfortunately, the function $u(z,g)$ does not become this time a function of only one variable and some further investigation is needed.

D. The coefficients $b_n(N)$

The first coefficient $b_1(N)$ which is a second degree polynomial in N is now over determined from the results for $N=1$,[4] $N=0$ and the large N limit and both our numerical calculations and the analytical result of Seznec[4] agree with this value.

For higher-order terms we have to rely more and more on numerical calculations. We have made the following ansatz: The quantity $3n2^{2n+1}b_n(N)$ is an integer for N integer. In addition for N even, $3n2^nb_n(N)$ is an integer. This means that in $b_n(N)$, the coefficient b_n^n of N^n should be divisible by 2, the coefficient of N^{n-1} by 4, etc.

This ansatz, which is of course compatible with the results coming from the large N limit, and the $N=0$ case, has been tested with decreasing accuracy, but beyond any doubts for n varying form 2 to 5.

For $n=6$, we have used it to try to determine all the coefficients of $b_6(N)$. The result we give has a very good chance of being correct and in any case is an excellent fit of the numbers.

Our results for the six polynomials $b_1(N)$ to $b_6(N)$ are given in Table II. One notices that the coefficients of a given polynomial are all of the same sign and that for N positive $b_n(N)$ alternate in sign. As g is negative, the series is, therefore, not Borel summable. In addition, it suggests that the large-order behavior of the b_n's is also dominated by an instanton of action $1/3g$. We know at least one case for which this conjecture is true, it is the case $N=0$ as Eq. (56) shows.

The prediction then is that, in general, $b_n(N)$ should behave, for n large, as

$$b_n(N) \sim Cn^a(-3)^{n+1}n!$$

We have tried to calculate for $N=1$ and 2 as many terms as possible. They are given in Table III. It can be

TABLE III. The values of $b_n(N)(-1)^{n+1}$ for $N=1$ and 2.

n	$N=1$	$N=2$
1	3.95833···	8.83333···
2	19.3437500···	56.7500···
3	174.2092014	646.01388···
4	2177.286133	9894.656250
5	34045.54329	184707.1208
6	632817.0536	4005925.573
7	1.357206×10^7	9.82462×10^9
8	3.2942×10^8	2.679×10^9
9	8.92×10^9	8.03×10^{10}
10	2.65×10^{11}	2.61×10^{12}

checked that the behavior of these terms is compatible with the prediction. In any case, as the coefficients of $(-1)^{n+1} \times b_n(N)$ are all positive, $b_n(N)$ is an increasing function of N for N positive, and $N=0$ provides a lower bound. In addition, the coefficients of $b_n(N)$ are higher than or equal to (for the term of degree zero) to the corresponding term of the perturbative expansion (in the sense explained in the $N=0$ case). From this remark we can also obtain a probable lower bound on the behavior of the $b_n(N)$.

III. THE DOUBLE-WELL POTENTIAL

We have also explored the ground state energy and the energy difference between the ground state and the first excited state of the double-well potential whose Hamiltonian H is

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}x^2(1 - x\sqrt{g})^2. \tag{62}$$

The analysis of the perturbative expansion of the ground state energy $E(g)$ is reduced to the following observation.[3] The numerical calculation shows the remarkable relation

$$E(g) = \tfrac{1}{2}E_{O(2)}(-g), \tag{63}$$

where $E_{O(2)}(g)$ is the ground state energy of the $O(2)$ anharmonic oscillator studied in Sec. II. This relation is obviously only valid in the sence of a series expansion, as $E_{O(2)}(-g)$ is complex for g positive. In Ref. 3 it was even verified that in the sense of functions

$$E(g) \neq \tfrac{1}{2}\mathrm{Re}(E_{O(2)}(-g)) \quad \text{for } g>0. \tag{64}$$

The difference decreases exponentially when g goes to zero as $\exp(-1/3g)$.

The relation (63) has not yet been proved analytically although this should not be too difficult.

As a result the structure of the large-order behavior of $E(g)$ can be found entirely in Sec. II by specializing all formulas to $N=2$ and by changing g to $-g$.

We shall now study a more interesting problem, the perturbative expansion of the difference $\Delta E(g)$ between the energy of the first excited state and the ground state. This energy difference can be calculated from the path integral expression for $\exp(-\beta H)$, by expanding the integrand around the instanton solution which goes from one minimum of the potential at 0 to the other at $1/\sqrt{g}$.

At leading order,

$$\Delta E(g) \sim \frac{2}{\sqrt{\pi g}} e^{-1/6g}, \tag{65}$$

because the classical action corresponding to the instanton is $1/6g$. The corrections to this formula take the form of a power series in g which we shall calculate,

$$\Delta E(g) = \frac{2}{\sqrt{\pi g}} e^{-1/6g}\left(1 + \sum_{1}^{\infty} \epsilon_k g^k\right). \tag{66}$$

But before presenting this calculation, we shall present the recursion formula we have used to expand $E(g)$.

A. The ground state energy

It is easy in the case of the double-well potential to perform a systematic WKB expansion of the wavefunction. This leads to a faster and more accurate calculation of the coefficients of the expansion of the ground state energy. The recursion formula can also be more easily used to prove some bounds on the coefficients.

Let us write again the Ricatti equation for the logarithmic derivative of the wavefunction $\psi(x)$,

$$f(x) = -\frac{d}{dx}\ln\psi(x), \tag{67}$$

$$f'(x) - f^2(x) + x^2(1 - x\sqrt{g})^2 = 2E(g). \tag{68}$$

The WKB expansion is an expansion in powers of g, at $x\sqrt{g}$ fixed,

$$f(x) = x(1 - x\sqrt{g}) - \frac{\sqrt{g}}{1 - x\sqrt{g}} - \sum_{k=2}^{\infty} g^{k-1/2} \sum_{n=2}^{2k-1} \frac{C_k^n}{(1 - x\sqrt{g})^n}. \tag{69}$$

The coefficients C_k^n satisfy then the recursion equation,

$$C_{k+1}^{n+2} = C_{k+1}^{n+3} + \tfrac{1}{2}(n+2)C_k^n + \tfrac{1}{2}\sum_{\substack{2\leqslant l\leqslant k-1 \\ 2\leqslant m\leqslant n-1}} C_l^m C_{k+1-l}^{n-m+1}, \tag{70}$$

with $C_2^2 = C_2^3 = 1$ for $k \geqslant 2$, and with $C_k^n = 0$ for $n > 2k-1$. If we now define

$$E(g) = \tfrac{1}{2} + \sum_{k=1}^{\infty} E_k g^k, \tag{71}$$

we obtain

$$E_k = -C_{k+1}^2. \tag{72}$$

The recursion formula shows that all C_k^n are positive, again have simple airthmetical properties, but in addition leads to simple evaluations.

Let us give an example. We can first find a generating function for C_k^{2k-1} and, therefore, evaluate C_k^{2k-1} for k large, as C_k^{2k-1} is the coefficient of the most singular contribution order by order in g,

$$h(z) = z - \frac{1}{z} - \sum_{k=2}^{\infty} C_k^{2k-1} z^{-2k+1}. \tag{73}$$

The function $h(z)$ satisfies

$$h' + h^2 - z^2 + 1 = 0. \tag{74}$$

It has a trivial solution,

$$h_0(z) = -z. \tag{75}$$

We can, therefore, find the other solution by setting

$$h(z) = -z + 1/u. \tag{76}$$

The function $u(z)$ is solution of a linear equation,

$$u'(z) + 2zu(z) - 1 = 0. \tag{77}$$

The solution is

$$u(z) = e^{-z^2}\int_a^z e^{t^2}\,dt. \tag{78}$$

The quantity a is an arbitrary and, for z large, irrelevant constant.

The large-order behavior of the coefficients of the ex-

pansion of $u(z)$ for z large can then be obtained, and, therefore, the behavior of C_k^{2k-1} for k large,

$$k\to\infty \quad C_k^{2k-1}\sim(2/\sqrt{\pi})\Gamma(k+\tfrac{1}{2}). \tag{79}$$

The equation for $f(x,g)$ can then be linearized around the sum of the most singular contributions, and an evaluation of C_k^n can be obtained,

$$C_k^n\sim\frac{2}{\sqrt{\pi}}\frac{\Gamma(3k-\frac{1}{2}-n)}{\Gamma(2k-n)}\left(\frac{2}{3}\right)^{2k-1-n},\quad k\to\infty, \tag{80}$$

which is in agreement with the large-order behavior of E_k[10] obtained for $n=2$.

The $1/k$ correction can also be calculated, but the calculation is very tedious.

B. The energy shift between the ground state energy and the first excited state

We shall expand in a power series the difference $\Delta E(g)$ between the two lowest eigenstates of H. To do this we shall use a very simple argument exposed in Landau and Lifshitz.[11] It is based essentially upon the symmetry of the potential. In d dimensions the symmetry of the potential with respect to a $(d-1)$-dimensional hyperplane is required.

Here the symmetry corresponds to the interchange $x\sqrt{g}$ in $1-x\sqrt{g}$.

C. The method

Let us assume that we have an analytic potential symmetric under the change $x\to -x$,

$$V(x)=V(-x),$$

and possessing two minima at $x=\pm R$.

The ground state energy can be calculated by expanding the potential at one of its minima, and making a perturbative calculation starting from the harmonic approximation. This yields the ground state energy up to corrections exponentially small for large R. The leading correction can be calculated and corresponds to the half-difference $\Delta E/2$ between the ground state energy and the first excited state.

Let us assume that $\varphi_1(x)$ is the wavefunction obtained by expanding g around $x=R$ and E_0 is the corresponding energy,

$$H=\tfrac{1}{2}p^2+V(x), \tag{81}$$

$$-\tfrac{1}{2}\varphi_1''(x)+V(x)\varphi_1(x)=E_0\varphi_1(x). \tag{82}$$

The solution $\varphi_2(x)$ obtained by expanding around $x=-R$ is, of course, related to $\varphi_1(x)$,

$$\varphi_2(x)=\varphi_1(-x). \tag{83}$$

The solution $\varphi_1(x)$ is valid in the region

$$x>-R+\eta,\quad \eta>0. \tag{84}$$

The symmetric situation is true for $\varphi_2(x)$. In addition $\varphi_1(x)$ decreases exponentially when $|x-R|$ increases. As a result, in order to calculate the leading exponential correction to E_0, we can use the variational principle with a trial function $\phi_\epsilon(x)$,

$$\phi_\epsilon(x)\begin{cases} x>a>0, & \varphi_1(x)\left(1+\epsilon\dfrac{\varphi_2(a)}{\varphi_1(a)}\right),\\ -a<x<a, & \varphi_1(x)+\epsilon\varphi_2(x),\\ x<-a, & \varphi_2(x)\left[\epsilon+\dfrac{\varphi_1(-a)}{\varphi_2(-a)}\right]. \end{cases} \tag{85}$$

The trial function $\phi_\epsilon(x)$ is even or odd if ϵ is $+1$ or -1, respectively, and will yield, therefore, the ground state energy or the energy of the first excited state.

We now calculate

$$E_\epsilon=\frac{\langle\phi_\epsilon|H|\phi_\epsilon\rangle}{\langle\phi_\epsilon|\phi_\epsilon\rangle}. \tag{86}$$

If the derivative of ϕ_ϵ at $\pm a$ would be continuous, E_ϵ would just be equal to E_0, as $\varphi_1(x)$ and $\varphi_2(x)$ satisfy locally the Schrödinger equation. The additional contributions come from the discontinuity of $\phi_\epsilon'(x)$ at $x=\pm a$. The contributions from $+a$ and $-a$ are, of course, equal. At the same time the norm of ϕ_ϵ receives two identical contributions coming from the neighborhood of $\pm R$. Integrating by part,

$$\begin{aligned}\int_{-\infty}^{+\infty}&[\tfrac{1}{2}\phi_\epsilon'(x)^2+V(x)\phi_\epsilon^2(x)]\,dx\\ &=\int_{-\infty}^{+\infty}[-\tfrac{1}{2}\phi_\epsilon(x)\phi_\epsilon''(x)+V(x)\phi_\epsilon^2(x)]dx\\ &\quad+\phi_\epsilon(a)[\phi_\epsilon'(a_-)-\phi_\epsilon'(a_+)].\end{aligned} \tag{87}$$

We have, therefore, for E_ϵ, up to higher-order exponentially small corrections,

$$\begin{aligned}E_\epsilon=E_0+\frac{1}{2}&\frac{[\varphi_1(a)+\epsilon\varphi_2(a)]}{\int\varphi_1^2(x)\,dx}\\ &\times\left[\varphi_1'(a)+\epsilon\varphi_2'(a)-\varphi_1'(a)\left(1+\epsilon\frac{\varphi_2(a)}{\varphi_1(a)}\right)\right].\end{aligned} \tag{88}$$

The energy difference ΔE is thus

$$\Delta E=E_+-E_-=\frac{[\varphi_1(a)\varphi_2'(a)-\varphi_2(a)\varphi_1'(a)]}{\int\varphi_1^2(x)\,dx}. \tag{89}$$

As $\varphi_1(x)$ and $\varphi_2(x)$ are two solutions at the same energy, the Wronskian which appears in the numerator is independent of a, as it should be. In higher dimensions, the Wronskian is replaced by the charge associated with the conserved current $J_\mu(x)$,

$$J_\mu(x)=\varphi_1(x)\overleftrightarrow{\partial\mu}\varphi_2(x). \tag{90}$$

In our case, in order to be able to generate a systematic expansion of $\Delta E(g)$ in powers of g we need a WKB expansion of $\varphi_1(x)$, as we have derived it above. In order to calculate the norm of $\varphi_1(x)$ we need only the simple perturbative expansion of $\varphi_1(x)$.

As we know, the logarithmic derivative of the wavefunction $f(x)$, the Wronskian $W(a)$ can be written

$$W(a)=\psi(a)\psi\left(\frac{1}{\sqrt{g}}-a\right)\left[f(a)+f\left(\frac{1}{\sqrt{g}}-a\right)\right]. \tag{91}$$

Remember that in our problem the symmetry corresponds to x going to $(1/\sqrt{g})-x$.

We still have the choice of the normalization of the wavefunction $\psi(x)$. We choose

$$\ln\psi(x) = -x^2/2 + (x^3/3)\sqrt{g} - \ln(1 - x\sqrt{g}) + \sum_{k=2}^{\infty} g^{k-1} \sum_{n=2}^{2k-1} \frac{C_k^n}{(n-1)(1-x\sqrt{g})^{n-1}}. \quad (92)$$

As $W(a)$ is independent of a, we can calculate it for a large,

$$f(a) = a(1 - a\sqrt{g}) + O(1/a), \quad (93)$$

$$\ln\psi(a) = -a^2/2 + a^3\sqrt{g}/3 - \ln(1 - a\sqrt{g}) + O(1/a).$$

Thus

$$W(a) = \left(\frac{1}{ag(1-a\sqrt{g})} e^{-1/6g}\right)\left[2a(1-a\sqrt{g})\right], \quad (94)$$

$$W(a) = (2/\sqrt{g})e^{-1/6g}. \quad (95)$$

We have completely reduced the problem to the calculation of the norm of $\psi(x)$ with the special normalization given above.

At leading order

$$\int \psi^2(x)\,dx = \int \frac{e^{-x^2 + (2/3)x^3\sqrt{g}}}{(1-x\sqrt{g})^2}\,dx + O(g) = \sqrt{\pi} + O(g). \quad (96)$$

This integral has to be understood as a power series in g. The fact that the integral does not converge as such, shows that the series will not be Borel summable as was the case for the series of $E_0(g)$.

At leading order we have recovered the instanton result,

$$\Delta E(g) = (2/\sqrt{\pi g})e^{-1/6g}[1 + O(g)]. \quad (97)$$

Higher-order calculations will involve the perturbative calculation of integrals of the form

$$I_n(g) = \int \frac{e^{-x^2 + (2/3)x^3\sqrt{g}}}{(1-x\sqrt{g})^n}\,dx. \quad (98)$$

It is easy to write recursion formulas for these integrals. These recursion formulas and other properties of $I_n(g)$ can be found in Appendix B. It is easy to see that the contributions coming from these integrals all have a positive sign, and that the expansion of $\psi^2(x)$ is a linear combination of the $I_n(g)$ with positive coefficients. It is, therefore, possible to write various lower bounds for the coefficients of g^k in the expansion of $\int\psi^2(x)\,dx$.

The simplest bound comes from the expansion of the integral $I_2(g)$. At large order $I_{2,k}$ behaves as

$$I_2(g) = \sum_0^{\infty} I_{2,k} g^k, \quad (99)$$

$$I_{2,k} \sim (3/\sqrt{\pi})3^k k!, \quad k \to \infty. \quad (100)$$

It is possible to calculate by hand a few terms of the expansion. Higher-order calculations can be done on a computer. We give in Table IV the first ten terms of the expansion of the quantity $L(g)$,

TABLE IV. Values of $-6n2^n L_n$. It is easy to verify that these numbers are divisible by the GCD given in Table I.

$n=$ 1	71
2	1890
3	65953
4	2733150
5	128867746
6	6758057124
7	388879707749
8	24338697845358
9	1646577122826766
10	119850766679371980

$$L(g) = \ln\left(\frac{\sqrt{\pi g}}{2} e^{1/6g}\Delta E(g)\right). \quad (101)$$

Notice that the coefficients are rational numbers with denominators identical to those of the expansion of the ground state energy given by the $O(2)$ anharmonic oscillator. We derive in Appendix B this property for the coefficients of $\ln I_2(g)$.

A check of our calculation is that we have[4] postulated the first three terms of this expansion from a numerical analysis of the energy difference based on a numerical solution of the Schrödinger equation. They agree with the present calculation. In addition we have verified by the same procedure the asymptotic nature of the expansion, using the additional terms calculated here (10^{-15} accuracy).

We have calculated 94 terms of the expansion, and analyzed its large-order behavior, which is related to the triple instanton contribution to the path integral which gives $\exp(-\beta H)$. Some details of this analysis are reported in Appendix A.

The result is

$$L(g) = \sum_0^{\infty} L_k g^k, \quad (102)$$

$$L_k = \frac{3^{k+1}}{\pi} k!\left[3(\ln 6k + \gamma) - \frac{1}{k}\left(\frac{35}{6}(\ln 6k + \gamma) + 9\right) + O\left(\frac{\ln k}{k^2}\right)\right], \quad k \to \infty \quad (103)$$

in which γ is the Euler constant, $\gamma = 0.5772156\cdots$. The remarkable new fact is the appearance of $\ln k$ factors. This indicates that the small g expansion of the triple instanton contribution which probably governs the large orders of this expansion has $\ln g$ factors too.

D. First-order correction to the energy difference between the two lowest states of a symmetric potential

Using the method described in the special case of the double-well potential, it is straightforward to calculate the first-order correction to $\Delta E(g)$ for an arbitrary symmetric analytic potential having two degenerate minima. We take as Hamiltonian H

$$H = \tfrac{1}{2}p^2 + (1/g)V(x\sqrt{g}). \quad (104)$$

The potential is symmetric under the change

$$V(x) = V(-x),$$

and has a small x expansion,

$$V(x) = \frac{1}{2}x^2 + \frac{\beta x^3}{6} + \frac{\gamma x^4}{24} + O(x^5). \tag{105}$$

The method used above is applicable without any modification and we just quote the result:

$$\Delta E(g) = \frac{2a}{\sqrt{\pi g}} \exp\left\{ -\frac{1}{g}\int_0^a \sqrt{V(x)}\,dx + \frac{1}{2}\int_0^a dx\left(\frac{1}{\sqrt{V}} - \frac{1}{x(1-x/a)}\right) + g\int_0^a dx\left[\frac{E_1}{\sqrt{V}} + \frac{1}{8} \times \left(V^{-3/2} - \frac{V'^2}{4V^{5/2}} + \frac{2\beta}{3x^2} + \frac{2\beta}{3(a-x)^2}\right)\right] + g\left(\frac{\beta}{6a} + \frac{7\gamma}{64} - \frac{299\beta^2}{(12)^3}\right)\right\} + O(g^2). \tag{106}$$

In this expression E_1 is the coefficient of g in the expansion of the ground state energy,

$$E_1 = \frac{\gamma}{32} - \frac{11\beta^2}{288}. \tag{107}$$

APPENDIX A: EXTRAPOLATION METHODS

A. Ground state energy

We explain here the method we have used to extract the corrections to the larger-order behavior from the coefficients of the series expansion of the ground state energy.

Let us consider a sequence $\{S_n\}$ which admits an asymptotic expansion of the form

$$s_n = \alpha_0 + \frac{\alpha_1}{n} + \frac{\alpha_2}{n^2} + \cdots, \tag{A1}$$

for n large. We want to improve the convergence of s_n towards α_0. The standard method is to construct a Neville table, using the following iterative scheme: One defines a sequence of sequences $\{s_n^{(p)}\}$ by the following recursion formula,

$$s_n^{(p+1)} = s_{n+1}^{(p)} + \frac{n}{p}(s_{n+1}^{(p)} - s_n^{(p)}). \tag{A2}$$

It is easy to verify that

$$s_n^{(p)} = \alpha_0 + O(1/n^p). \tag{A3}$$

In our case it was possible to again improve the convergence of this procedure by extrapolating the sequence $s_n^{(p)}$ at fixed n by the means of Padé approximants. However, it appeared simpler and more efficient to use a slightly different procedure.[12]

Starting from the sequence $\{s_n\}$ we can construct

$$s_n^{(p+1)} = s_{n+1}^{(p)} - \frac{2p}{(2p-1)} \frac{[s_{n+2}^{(p)} - s_{n+1}^{(p)}][s_{n+1}^{(p)} - s_n^{(p)}]}{s_{n+2}^{(p)} + s_n^{(p)} - 2s_{n+1}^{(p)}}, \tag{A4}$$

$$s_n^{(1)} = s_n.$$

One step of this procedure is equivalent to two steps of the Neville method in the sense that two successive powers of n are eliminated at each step. This a consequence of the main feature of this method: The variable n does not appear explicitly in the recursion formula, so that everything is translationally invariant in n. The following problem is, therefore, solved approximatively: In the Neville table we have eliminated successive powers of $1/n$, but it might have been more efficient to eliminate, instead, powers of $1/(n+a)$, with a constant. The new procedure optimizes locally the choice of the variable. If the only correction term would be of the form $b/(n+a)$, it would be eliminated in one step only.

In our example this method worked extremely well, and no improvement coming from Padé approximants was needed anymore.

B. Energy difference of the lowest lying states of the double-well potential

In this case the analysis revealed that the structure of the series divided by $3^n n!$ was more complicated and did not seem to converge. We guessed that it had the form

$$s_n = a_n \ln n + b_n, \tag{A5}$$

where a_n and b_n have expansions in powers of $1/n$,

$$a_n = \alpha_0 + \frac{\alpha_1}{n} + \frac{\alpha_2}{n^2} + \cdots,$$
$$b_n = \beta_0 + \frac{\beta_1}{n} + \frac{\beta_2}{n^2} + \cdots. \tag{A6}$$

We had, therefore, to modify the Neville procedure in the following way: We first calculated

$$s'_n = n(s_{n+1} - s_n). \tag{A7}$$

As a result the new series had the form

$$s'_n = a'_n \ln n + b'_n,$$
$$a'_n = \frac{\alpha'_1}{n} + \frac{\alpha'_2}{n^2} + \frac{\alpha'_3}{n^3} + \cdots, \tag{A8}$$
$$b'_n = \beta'_0 + \frac{\beta'_1}{n} + \frac{\beta'_2}{n^2} + \cdots.$$

Now the calculation of β'_0 involves the use of the Neville procedure, performing each step twice. At the first step a term $\ln n/n^p$ is eliminated. At the second step the term proportional to $1/n^p$ is cancelled.

In general, if we were to encounter a series with structure

$$s_n = a_n^0 (\ln n)^q + a_n^1 (\ln n)^{q-1} + \cdots + a_n^q, \tag{A9}$$

we would repeat, $q-1$ times, the operation

$$n(s_{n+1} - s_n),$$

and then each Neville step q times, in order to find the limit of a_n^0.

APPENDIX B: SERIES EXPANSION OF THE INTEGRALS $I_n(g)$

We shall briefly explain some properties of the integrals $I_n(g)$ which appear in the calculation of the coefficients of the expansion around the simple instanton for the double-well potential,

$$I_n(g) = \int_{-\infty}^{+\infty} \frac{e^{-x^2 + (2/3)x^3\sqrt{g}}}{(1 - x\sqrt{g})^n}\,dx. \tag{B1}$$

We emphasize again that these integrals are just generating functions of a power series in g. We shall consider first $I_2(g)$ because it is the leading contribution to the norm of $\psi(x)$, and it seems also to govern the arithmetical properties of the whole expansion. Integrating by part, its is possible to rewrite $I_2(g)$ as

$$I_2(g) = \frac{2}{\sqrt{g}} \int x e^{-x^2 + (2/3)x^3\sqrt{g}}\, dx. \tag{B2}$$

The coefficients of the expansion can, therefore, be obtained explicitly,

$$I_2(g) = \sum_{k=0}^{\infty} I_2^k g^k,$$
$$I_2^k = \frac{4}{3}\left(\frac{4}{9}\right)^k \frac{\Gamma(3k + (5/2))}{(2k+1)!}, \tag{B3}$$

which can be rewritten

$$I_2^k = \frac{\sqrt{\pi}}{(12)^k k!} \prod_{n=1}^{k} (6n-1)(6n+1). \tag{B4}$$

The large-order behavior of I_2^k can easily be found,

$$I_2^k \underset{k\to\infty}{=} \frac{1}{\sqrt{\pi}} 3^{k+1} k! [1 + O(1/k)], \tag{B5}$$

which is, up to a factor $\sqrt{\pi}$, the large-order behavior of the ground state energy, and is a lower bound for the behavior of the expansion around the instanton.

A simple calculation shows also that $I_2(g)$ satisfies a differential equation

$$36g^2 I_2''(g) + (108g - 12) I_2'(g) + 35 I_2(g) = 0. \tag{B6}$$

This translates in a Ricatti equation for the logarithmic derivative

$$F(g) = \frac{I_2'(g)}{I_2(g)},$$
$$36g^2(F' + F^2) + (108g - 12)F(g) + 35 = 0. \tag{B7}$$

If we now expand $F(g)$ in a power series

$$F(g) = \sum F_k g^k, \tag{B8}$$

we get the recursion formula

$$F_k = 3(k+2)F_{k-1} + 3 \sum_{0 \le l \le k-2} F_l F_{k-2-l}, \quad k \ge 2 \tag{B9}$$

$F_0 = 35/12.$

It follows from this equation that F_k is a rational number with a denominator of the form $3(2^{k+2})$, to be compared with the denominator of I_2^k.

We shall now explain how we calculate the other integrals $I_n(g)$. It is straightforward, using an integration by part, to derive the relation

$$I_n(g) - I_{n-1}(g) = [(n+1)/2]\, g I_{n+2}(g), \tag{B10}$$

which yields a recursion formula for the coefficients I_n^k,

$$I_n^k - I_{n-1}^k = [(n+1)/2] I_{n+2}^{k-1}. \tag{B11}$$

Since we know I_2^k, these relations allow us to calculate all I_n^k.

[1] C. Bender and T.T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Lett. **27**, 461 (1971); Phys. Rev. D **7**, 1620 (1973).
[2] E. Brezin, J.C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **15**, 1544, 1558 (1977). See also, J. Zinn-Justin, Salamanca 1977, *Lecture Notes in Physics* **77**, *Number* 126, edited by J.A. de Azcerraga (Springer-Verlag, Berlin, 1977); Zinn-Justin, *Cargèse Lectures* 1977 (Plenum, New York, 1977).
[3] Le Guillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977); and Dph-T Preprint Dph-T/79/94.
[4] R. Seznec and J. Zinn-Justin, J. Math. Phys. **20**, 1398 (1979).
[5] R. Seznec, Thesis, Paris University, 1979.
[6] E. Brezin and S. Hikami, J. Phys. A **12**, 759 (1979).
[7] H. de Vega, Dph-T preprint, Dph-T/79/105.
[8] A.D. Dolgov and V.S. Popov, Phys. Lett. B **86**, 185 (1979).
[9] B. Simon, Ann. Phys. (N.Y.) **58**, 79 (1970); S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970); and Ref. 1.
[10] E. Brezin, G. Parisi, and J. Zinn-Justin, Phys. Rev. D **16**, 408 (1977).
[11] L.D. Landau and E.M. Lifschitz, *Quantum Mechanics*, 2nd ed. (Pergamon, Oxford, 1965), p. 292.
[12] J. Zinn-Justin, J. Phys. **40**, 969 (1979).

On decoupling of finite singularities in the scattering theory for the Schrödinger operator with a magnetic field

M. A. Perelmuter and Ya. A. Semenov
Department of Mathematics, Kiev Polytechnic Institute, 252056 Kiev 56, USSR

(Received 5 November 1979; accepted for publication 15 February 1980)

In this paper, the authors extend the theory of the effect of local singularities on scattering theory to include magnetic fields.

PACS numbers: 03.65.Nk, 03.80. + r

1. INTRODUCTION

The problem of the potential scattering theory for the Schrödinger operator $H = H_0 + q$ acting in the Hilbert space $\mathscr{H} = L^2(\mathbb{R}^l)$ is to study a spectrum of the operator H. According to the spectral theory for a self-adjoint operator $H:\mathscr{H}\to\mathscr{H}$ a unique decomposition for $\mathscr{H}$ is

$$\mathscr{H} = \mathscr{H}_{ac} \oplus \mathscr{H}_{\text{sing}} \oplus \mathscr{H}_p$$

and H leaves each $\mathscr{H}_{\dots}$ invariant and, moreover, $H \upharpoonright \mathscr{H}_{ac}$ has an absolutely continuous spectrum, $H \upharpoonright \mathscr{H}_{\text{sing}}$ has a singular continuous spectrum and $H \upharpoonright \mathscr{H}_p$ has a complete set of eigenvectors.

The existence of the wave operators

$$\Omega_{\pm}(H,H_0) = \operatorname*{s-lim}_{t\to\pm\infty} e^{itH}e^{-itH_0}$$

guarantees the validity of the following conclusions

$$\operatorname{Ran}\Omega_{\pm} \subset \mathscr{H}_{ac} \subset \mathscr{H}_p^{\perp}.$$

The demonstration of the existence of the wave operators is one of the main problems in the potential scattering theory. In early days of the scattering theory J. M. Cook[1] proved the following simple and strong criterion of the existence of the wave operators:

Theorem: Suppose that there is a subset $\mathscr{E} \subset \mathscr{D}(H_0)\cap\mathscr{D}(q)$ dense in $\mathscr{H}$ so that for any $\varphi\in\mathscr{E}$

$$\int_{-\infty}^{\infty} \|qe^{itH_0}\varphi\|\, dt < \infty.$$

Then $\Omega_{\pm}(H,H_0)$ exist.

M. Schechter[2,3] has recently proved the following useful generalization of the Cook theorem:

Theorem: Let H, H_0 be self-adjoint operators in $\mathscr{H}$. Assume that

(1) $\mathscr{D}(B)\supset\mathscr{D}(H)$ and $\|Bu\|\leqslant\text{const}(\|Hu\| + \|u\|)$ for all $u\in\mathscr{D}(H)$;

(2) $\mathscr{D}(H_0)\subset\mathscr{D}(A)$;

(3) For all $\varphi\in\mathscr{D}(H)$, $\psi\in\mathscr{D}(H_0)$,

$$\langle H\varphi,\psi\rangle = \langle\varphi,H_0\psi\rangle + \langle B\varphi,A\psi\rangle$$

($\langle\cdot,\cdot\rangle$ is the inner product).

(4) There is a dense subset $\mathscr{E}\subset\mathscr{H}_{ac}(H_0)$ so that for any $\varphi\in\mathscr{E}$ there is $T_0 < \infty$ with $\exp(itH_0)\varphi\in\mathscr{D}(A)$ for $|t| > T_0$ and

$$\int_{|t|>T_0} \|Ae^{itH_0}\varphi\|\, dt < \infty.$$

Then $\Omega_{\pm}(H,H_0)$ exist.

Remark: B. Simon[4] has found a proof of this theorem based on some modification of Cook's arguments.

On the basis of physical intuition it is clear that only a behavior of the potential $q(x)$ at infinity is critical for the wave operators $\Omega_{\pm}(H,H_0)$ to exist. The most general result on the existence of $\Omega_{\pm}(H,H_0)$ refelcts this heuristic fact.

Theorem[5]: Let H be a self-adjoint operator so that for any $\varphi\in\mathscr{S}(\mathbb{R}^l)$ ($\mathscr{S}$ is the Schwartz space) with a support of φ outside some ball $\{x:|x|\leqslant R\}$ with the fixed radius R, $H\varphi = H_0\varphi + q\varphi$. Let χ_R be a characteristic function of the exterior of the $\{x:|x|\leqslant R\}$, and suppose that

$$\int_{-\infty}^{\infty} \|(1-\chi_R)q\, e^{itH_0}\varphi\|\, dt < \infty, \quad \forall\varphi\in\mathscr{S}(\mathbb{R}^l).$$

Then $\Omega_{\pm}(H,H_0)$ exist.

Next problem of the scattering theory is the investigation of the completeness of the wave operators $\Omega_{\pm}(H,H_0)$ [or the equality $\operatorname{Ran}\Omega_{\pm}(H,H_0) = \mathscr{H}_{ac}(H)$]. At present there are two methods, at any rate, in the time-dependent scattering theory. One method lies in the direct examining of the existence of $\Omega_{\pm}(H,H_0)$, where the following result was obtained:

Theorem 1.1[6,7]: Suppose that

$$\sup_{x\in\mathbb{R}^l}\left[(1+|x|)^{2+2\epsilon}\int_{|x-y|\leqslant 1}|x-y|^{-l+4-4\theta}q^2(y)\,dy\right] < \infty,$$

for some $\epsilon > 0$ and $0 < \theta < \frac{1}{2}$. Then $\Omega_{\pm}(H_0,H_0+q)$ exist and are complete.

Remark: By this theorem we can consider potentials whose behavior at infinity is $|x|^{-1-\epsilon}$, $\forall\epsilon > 0$.

The other method to prove the completeness is based on the fundamental Birman theorem[8]:

Theorem: Let $H\geqslant 0$ be a self-adjoint operator. Suppose that $\varphi(H) - \varphi(H_0)$ is the trace class (see Sec. 4 below) for the C^2 function φ on $[0,\infty)$ with strictly negative derivative. Then $\Omega_{\pm}(H,H_0)$ exist and are complete.

This theorem is valid in a more general sitaution for the generalized wave operators

$$W_{\pm}(H,H_0) = \operatorname*{s-lim}_{t\to\pm\infty} e^{itH}e^{-itH_0}E_{ac}(H_0),$$

where $E_{ac}(H_0)$ is the projection onto the $\mathscr{H}_{ac}(H_0)$. Note also the following fact:

Chain rule: Suppose that $W_{\pm}(H_2,H_1)$ and $W_{\pm}(H_3,H_2)$ exist (are complete). Then $W_{\pm}(H_3,H_1)$ exist (are complete).

The results[9-11] which are an analog of the Kupsh–Sandhas theorem for the completeness have been recently obtained.

Theorem[9-12]: Let $V\in L^1_{\text{loc}}(\mathbb{R}^l\setminus S)$, where S is a closed set

 022-2488/81/030521-13$1.00

of Lebesque measure zero: $V^-(x) \equiv \max\{0, -V\}$ and V^- is a H_0-form bounded operator with the bound < 1. Suppose that χ_R is a characteristic function of the ball $\{x:|x| \leqslant R\}$. Then $\Omega_{\pm}(H_0 + V, H_0)$, exist and are complete if and only if $\Omega_{\pm}(H_0 + V_{ex}, H_0)$, where $V_{ex} = (1 - \chi_R)V$, exist and are complete. (The operators $H_0 + V$ and $H_0 + V_{ex}$ are defined by the form method.)

The object of this article is to demonstrate the analogous principle for Schrödinger operators with a magnetic field $(i\nabla + \mathbf{a})^2 + V$ with general $\mathbf{a}$, V. Moreover, using the commutator method developed below the sufficient conditions for the existence and completeness of the $W_{\pm}[H_0, (i\nabla + (1 - \chi_R)\mathbf{a})^2 + (1 - \chi_R)V]$ will be given.

2. FORMULATION AND DISCUSSION OF THE BASIC RESULTS

We shall now consider the Schrödinger operator with a magnetic field $(i\nabla + \mathbf{a})^2 + V$, where

$$\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_l}\right)$$

is the distributional gradiant and $\mathbf{Q} \equiv (a^{(1)}, \ldots, a^{(l)})$ is the real valued vector function. Denote $\mathbf{a} \in L^p(\mathbb{R}^l)$ if and only if $a^{(j)} \in L^p(\mathbb{R}^l)$, $j = 1,2,\ldots,l$. Let us also assume S is a closed set of Lebesque measure zero. We define the sesquilinear form $\tilde{t}$ as the closure of the form

$$t[u,v] = \langle i(i\nabla + \mathbf{a})u, (i\nabla + \mathbf{a})v\rangle, \quad \mathscr{D}(t) = \mathscr{D}(\nabla) \cap \mathscr{D}(\mathbf{a}).$$

Let $H(\mathbf{a})$ be an operator associating with the form $\tilde{t}$. We shall also introduce the following designations: $V^{\pm}(x) = \max\{0, \pm V\}$, so that $V = V^+ - V^-$, $V^{\pm} \geqslant 0$ for any real valued function V. $A + B$ is the form sum of the self-adjoint operators A,B. (Subsequently we shall usually write $A + B$ instead of $A \dotplus B$.)

Our main result is

Theorem 2.1: Let $\mathbf{a} \in L^2_{loc}(\mathbb{R}^l \setminus S)$, $V \in L^1_{loc}(\mathbb{R}^l \setminus S)$. Suppose that

(1) $p_0 V^- \leqslant H_0 + d$, for some $p_0 > 1$, $d \geqslant 0$.

(2) There exists some $0 < R < \infty$ so that the generalized wave operators $W_{\pm}[H_0, (i\nabla + (1 - \chi_R)\mathbf{a})^2 + (1 - \chi_R)V]$ exist and are complete, where χ_R is the characteristic function of the ball $\{x:|x| \leqslant R\}$.

Then $W_{\pm}(H_0, H(\mathbf{a}) + V)$ exist and are complete.

Remarks:

(1) The necessity of the condition (1) Theorem 2.1 follows from the Pearson example.[13]

(2) Theorem 2.1 is new even in the case $\mathbf{a} = 0$. (Refs. 9 and 10).

(3) In Theorem 2.4 below we shall give conditions which guarantee the validity of the assumption (2) Theorem 2.1 (see also Ref. 14). In the case $\mathbf{a} = 0$ we can use, for example, Theorem 1.1.

Let us pick the two principal results out of Theorem 2.1:

Corollary 2.2: Suppose that $0 \leqslant V \in L^1_{loc}(\mathbb{R}^l \setminus S)$ and supp(V) is compact. Then $W_{\pm}(H_0, H_0 + V)$ exist and are complete.

Corollary 2.3: Let $\mathbf{a} \in L^2_{loc}(\mathbb{R}^l \setminus S)$ and supp$(\mathbf{a})$ is compact. Then $W_{\pm}(H_0, H(\mathbf{a}))$ exist and are complete.

Remarks:

(1) Corollary 2.2 was obtained previously by P. Dieft and B. Simon,[10] Yu. A. Semenov,[11] M. Combescure and J. Ginibre.[9] (See also D. B. Pearson,[12] E. B. Davies and B. Simon.[15])

(2) In the case of the operator $-\Delta + V$ the above Kupsch–Sandhas existence theorem precedes the completeness Theorem 2.1. However, Corollary 2.3 is a new result both for the existence problem and the completeness problem for the operator $(i\nabla + \mathbf{a})^2$.

The following theorem gives sufficient conditions for the validity of the assumption (2) Theorem 2.1.

Theorem 2.4: Suppose that $\mathbf{b} \in L^2_{loc}(\mathbb{R}^l)$, $Q \in L^1_{loc}(\mathbb{R}^l)$. Let $p(x) = (1 + |x|^2)^{1/2}$. Suppose that $\mathbf{b} = p^{-\nu}\mathbf{b}_0$, $Q = p^{-\nu}Q_0$ for some $\nu > l$ and the functions $\mathbf{b}_0$, Q_0 obey the following conditions

$$p_0 \cdot 4\mathbf{b}_0^2 \leqslant H_0 + d, \qquad p_0 \cdot Q_0 \leqslant H_0 + d, \quad p_0 > 1, \ d \geqslant 0. \tag{2.1}$$

Then $W_{\pm}(H(\mathbf{b}) + Q + H_0)$ exist and are complete.

Remarks:

(1) In the recent paper by J. Avron, I. Herbst and B. Simon[16] the completeness of the wave operators $W_{\pm}(H(\mathbf{a}) + V, H(\mathbf{a}))$, where $\mathbf{a} \in L^2_{loc}(\mathbb{R}^l)$ and V belongs to the Rollnik space if $l = 3$ (see, for example, Ref. 17, p. 3) or V belongs to the Birman–Solomjak space (see Ref. 18) if $l \geqslant 4$, was demonstrated. In connection with the Avron *et al.* Results we may note that our methods allow proving Theorem 2.4 if we replace the assumption (2.1) by the following assumption: Q belongs to the Birman–Solomjak space.

(2) Our methods allow us to combine the results obtained by different ways. For example, the following is valid:

Theorem 2.5: Suppose that q and Q obey the assumptions of Theorem 1.1 and Theorem 2.4 (with $\mathbf{b} = 0$), respectively. Then $W_{\pm}(H_0 + q + Q, H_0)$ exist and are complete.

3. CONSTRUCTION OF SELF-ADJOINT EXTENSTIONS

(i) We shall first review some of the main ideas in the theory of quadratic forms on a Hilbert space. (There is a very complete discussion in Refs. 19 and 20.

Let $\mathscr{H}$ be a complex Hilbert space, t a densely defined symmetric closed nonnegative form on $\mathscr{H}$. Then there exists a self-adjoint operator $T \geqslant 0$ such that

$$t[u,v] = \langle T^{1/2}u, T^{1/2}v\rangle, \qquad \mathscr{D}(t) = \mathscr{D}(T^{1/2}).$$

Let us also consider a symmetric quadratic form α. The form α is called t-bounded if and only if $\mathscr{D}(\alpha) \supset \mathscr{D}(t)$ and

$$|\alpha[u,u]| \leqslant ct[u,u] + d\|u\|^2, \quad \forall u \in \mathscr{D}(t)$$

for some constants $c,d \geqslant 0$. (Subsequently we shall put $\alpha[u,u] = \alpha[u]$.)

The infimum of all possible values of the constants is called a t-form bound of the form α.

If α is the t-bounded form, we can represent α by the operator $A \in \mathscr{L}(\mathscr{H}_+, \mathscr{H}_-)$ ($\mathscr{L}$ is the space of the bounded operators.):

$$\alpha[u,v] = \langle u, Av\rangle, \quad u,v \in \mathscr{H}_+ = D(A),$$

where $\mathscr{H}_+ = \{\mathscr{D}(t), \|\cdot\|_+ = \|(T+1)^{1/2}\cdot\|\}$ and $\mathscr{H}_-$ is a space

dual to $\mathscr{H}_+$ with the norm $\|\cdot\|_- = \|(T+1)^{-1/2}\cdot\|$, so $\mathscr{H}_+ \subset \mathscr{H} \subset \mathscr{H}_-$ and $\|\cdot\|_- \leqslant \|\cdot\| \leqslant \|\cdot\|_+$.

As $(T+\lambda)^{-1/2} \in \mathscr{L}(\mathscr{H},\mathscr{H}_+) \cap \mathscr{L}(\mathscr{H}_-,\mathscr{H})$, $\lambda > 0$, the oprator $(T+\lambda)^{-1/2} A (T+\lambda)^{-1/2}$ is a bounded map from $\mathscr{H}$ to $\mathscr{H}$. It is clear that

$$\|(T+\lambda)^{-1/2} A (T+\lambda)^{-1/2}\| < c + \gamma$$

for any $\gamma > 0$ and all sufficiently large $\lambda > 0$, where c is a t-bound of the form α.

Let α be a t-bounded form with the t-bound less than one. Then the form $t + \alpha$, $\mathscr{D}(t+\alpha) = \mathscr{D}(t)$ is closed, bounded from below and, consequently, it associates the self-adjoint operator S. Moreover, we have the representation

$$\begin{aligned}(S+\lambda)^{-1} &= (T+\lambda)^{-1/2}[1 + (T+\lambda)^{-1/2} A (T+\lambda)^{-1/2}]^{-1} \\ &\quad \times (T+\lambda)^{-1/2}, \end{aligned} \tag{3.1}$$

for all sufficiently large $\lambda > 0$.

Remark: A complete discussion and numerous applications of the representation (3.1) are given in Simon's book.[17]

Let t and r be closed symmetric nonnegative forms on $\mathscr{H}$ and let $\mathscr{D}(t) \cap \mathscr{D}(r)$ be dense in $\mathscr{H}$. Then the form $s = t + r$, $\mathscr{D}(s) = \mathscr{D}(t) \cap \mathscr{D}(r)$ is also closed. Let S, T, R be operators associated by the forms s, t, r, respectively. The operator $S = T \dot{+} R$ is called the form sum of the operators T, R. Notice, it is possible $\mathscr{D}(T) \cap \mathscr{D}(R) = \{0\}$. (Further we shall write $+$ instead of $\dot{+}$.) We shall also denote $A \in PK(T)$ ($A \in PK_0(T)$) if and only if the form of the operator A is bounded with respect to the form of the operator T with a bound < 1 (equals zero).

(ii) Subsequently we shall need some approximation theorems about the convergence of the sequence of quadratic forms.

Theorem 3.1 *Ref.* 19, *Chapter* VIII, *Theorem* 3.11: Let $\{t_n\}$ be a nonincreasing sequence of densely defined closed symmetric nonnegative forms on $\mathscr{H}$. If T_n is a self-adjoint operator associated with t_n, then the sequence $\{T_n\}$ convergences in a strong resolvent sense to some self-adjoint operator $T_\infty \geqslant 0$. Thus if $n \to \infty$ we have $T_n \overset{R}{\to} T_\infty$, so that

$$(T_n + \lambda)^{-1} \overset{s}{\to} (T+\lambda)^{-1}, \quad \mathrm{Re}\lambda > 0.$$

In particular, if the symmetric form $t_\infty[u] = \lim_{n\to\infty} t_n[u]$, $\mathscr{D}(t_\infty) = \cup_{n\geqslant 1} \mathscr{D}(t_n)$ is closable, then T_∞ is a self-adjoint operator associated with the closure $\tilde{t}_\infty$ of the form t_∞.

Before formulating one more approximation theorem we shall introduce some definitions.

Let $\mathscr{H} = L^2(\mathbb{R}^l)$ and $\nabla \equiv (\partial/\partial x_1,\ldots,\partial/\partial x_l)$ be the distributional gradient. Consider the form

$$\begin{aligned} h_0[u,v] &= \sum_{k=1}^{l} \left\langle i\frac{\partial u}{\partial x_k}, \frac{\partial v}{\partial x_k} \right\rangle \\ &\equiv \langle i\nabla u, i\nabla v\rangle, \quad \mathscr{D}(h_0) = \mathscr{D}(\nabla) \end{aligned}$$

and the Laplace operator H_0 associated with the form. If $0 \leqslant q \in L^1_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, where S is some closed set of the measure zero, then $\mathscr{D}(\nabla) \cap \mathscr{D}(q^{1/2})$ is dense in $L^2(\mathbb{R}^l)$. So we can define the form sum $H_0 + q$.

Let W be a multiplicative operator associated with the measurable real-valued function $W(x)$. Let $W[u\,v] = \int uW\bar{v}dx$ be a h_0-bounded form with the h_0-bound less than one, so $W \in PK(H_0)$. Then the form sum $H_0 + q + W$, which is a bounded from below self-adjoint operator associated with the form $h_0 + W + q$ is well defined.

Theorem 3.2[20–24] Let H_0, q, W be above-defined.

(1) Let q_n and W_n be truncated operators corresponding to the q, W, respectively, so

$$q_n(x) = \begin{cases} q(x) & \text{if } q(x) \leqslant n, \\ 0 & \text{if } q(x) > n, \end{cases} \quad n = 1,2,\cdots$$

Then

$$\text{s-}\lim_{n\to\infty} (H_0 + q_n + W_n + \lambda)^{-1} = (H_0 + q + W + \lambda)^{-1},$$

for all sufficiently large $\lambda > 0$.

(2) There exist the functions $Z_n \in C_0^\infty(\mathbb{R}^l)$ $(n = 1,2,\cdots)$ such that the operators $\{H_0 + Z_n\}$ are bounded from below uniformly on $n = 1,2,\cdots$ and

$$\text{s-}\lim_{n\to\infty} (H_0 + Z_n + W_n + \lambda)^{-1} = (H_0 + q + W + \lambda)^{-1},$$

for all sufficiently large $\lambda > 0$ and, consequently, by the Trotter–Kato theorem

$$\text{s-}\lim_{n\to\infty} e^{-t(H_0 + Z_n)} = e^{-t(H_0 + q + W)}.$$

(iii) Definition of the Schrodinger Operator with Magnetic Field

Let $\mathbf{a} \in L^2_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, so $\mathbf{a} \equiv (a^{(1)},\ldots,a^{(l)})$, $a^{(j)} \in L^2_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, $j = 1, 2,\ldots,l$. We shall suppose $\mathbf{a}$ is a real-valued vector function.

As $\mathbf{a} \in L^2_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$ and $\mathbf{a}^2 \geqslant 0$ we can define the form sum $H_\epsilon(\mathbf{a}) = H_0 + (1+\epsilon)\mathbf{a}^2$, $(\epsilon \geqslant 0)$ associated with the form

$$\begin{aligned} &t_\epsilon[u,v] = h_0[u,v] + (1+\epsilon)\langle \mathbf{a}u, \mathbf{a}v\rangle, \\ &\mathscr{D}(t_\epsilon) = \mathscr{D}(\nabla) \cap \mathscr{D}(\mathbf{a}) = \mathscr{D}(\nabla) \cap \mathscr{D}(|\mathbf{a}|) \end{aligned}$$

Let us introduce the symmetric form α by the following way:

$$\begin{aligned} &\alpha[u,v] = \langle i\nabla u, \mathbf{a}v\rangle + \langle \mathbf{a}u, i\nabla v\rangle, \\ &\mathscr{D}(\alpha) = \mathscr{D}(\nabla) \cap \mathscr{D}(\mathbf{a}) = \mathscr{D}(t_\epsilon). \end{aligned}$$

Lemma 3.3: The form α is t_ϵ-bounded with the t_ϵ-bound $\leqslant (1+\epsilon)^{-1/2}$.

Proof: Let $u \in \mathscr{D}(t_\epsilon) = \mathscr{D}(\alpha)$, then

$$\begin{aligned} &t_\epsilon[u] + \sqrt{1+\epsilon}\alpha[u] \\ &\quad = \langle (i\nabla + \sqrt{1+\epsilon}\mathbf{a})u,\ (i\nabla + \sqrt{1+\epsilon}\mathbf{a})u\rangle \geqslant 0, \\ &t_\epsilon[u] - \sqrt{1+\epsilon}\alpha[u] \\ &\quad = \langle (i\nabla - \sqrt{1+\epsilon}\mathbf{a})u,\ (i\nabla - \sqrt{1+\epsilon}\mathbf{a})u\rangle \geqslant 0. \end{aligned}$$

Then $t_\epsilon[u] \geqslant \sqrt{1+\epsilon}|\alpha[u]|$. ∎

Introduce the form

$$t_a^{(\epsilon)} = t_\epsilon + \alpha, \quad \mathscr{D}(t_a^{(\epsilon)}) = \mathscr{D}(t_\epsilon), \ \epsilon > 0.$$

The symmetric nonnegative form $t_a^{(\epsilon)}$ is closed by Lemma 3.3, so the operator $H_\epsilon(\mathbf{a})$ associated with the form is self-adjoint nonnegative and

$$\begin{aligned} &t_a^{(\epsilon)}[u,v] = \langle u, H_\epsilon(\mathbf{a})v\rangle, \\ &u \in \mathscr{D}(t_a^{(\epsilon)}),\ v \in \mathscr{D}(H_\epsilon(\mathbf{a})), \\ &t_a^{(\epsilon)}[u,v] = \langle H_\epsilon(\mathbf{a})^{1/2} u, H_\epsilon(\mathbf{a})^{1/2} v\rangle, \end{aligned}$$

$$u,v\in\mathscr{D}(H_\epsilon(\mathbf{a})^{1/2})=\mathscr{D}(t_a^{(\epsilon)}),$$
$$(H_\epsilon(\mathbf{a})+\lambda)^{-1}=(H_+(\mathbf{a})+\lambda)^{-1/2}\times[1+(H_+(\mathbf{a})+\lambda)^{-1/2}\Lambda(\mathbf{a})(H_+(\mathbf{a})+\lambda)^{-1/2}]^{-1}(H_+(\mathbf{a})+\lambda)^{-1/2},$$

where $\Lambda(\mathbf{a})$ is the operator associated with the form α. [It follows from Lemma 3.3 that $\alpha[u,v]=\langle u,\Lambda(\mathbf{a})v\rangle$, $u,v\in\mathscr{D}(\Lambda(\mathbf{a}))$.]

Notice that the sets $\mathscr{D}(\Lambda(\mathbf{a}))$ and $\mathscr{D}(\alpha)=\mathscr{D}(t_a^{(\epsilon)})$ coincide.

It is clear that $0<\epsilon_1\leqslant\epsilon_2$ implies $0\leqslant H_{\epsilon_1}(\mathbf{a})\leqslant H_{\epsilon_2}(\mathbf{a})$. Hence it follows from Theorem 3.1, that if $\epsilon\downarrow 0$, the sequence $\{H_\epsilon(\mathbf{a})\}$ converges in the sense of a strong resolvent convergent to the self-adjoint operator $H(\mathbf{a})\geqslant 0$, so

$$(H_\epsilon(\mathbf{a})+\lambda)^{-1}\xrightarrow[\epsilon\downarrow 0]{s}(H(\mathbf{a})+\lambda)^{-1},\quad \mathrm{Re}\,\lambda>0.$$

Definition: The operator $H(\mathbf{a})=\text{R-lim}_{\epsilon\downarrow 0}\,H_\epsilon(\mathbf{a})$ is called the self-adjoint Schrödinger operator with a magnetic field.

Define the form t_a:

$$t_a[u,v]=\langle(i\nabla+\mathbf{a})u,\ (i\nabla+\mathbf{a})v\rangle,$$
$$\mathscr{D}(t_a)=\mathscr{D}(\nabla)\cap\mathscr{D}(\mathbf{a}).$$

As operator $i\nabla+\mathbf{a}$ with $\mathscr{D}(i\nabla+\mathbf{a})=\mathscr{D}(\nabla)\cap\mathscr{D}(\mathbf{a})$, is densely defined and symmetric, so it is closable. Thus the form t_a is closable (see Ref. 19, Chapter VI), nonnegative and consequently its closure $\tilde{t}_a$ associates the self-adjoint operator $H'(\mathbf{a})$. Theorem 3.1 implies $H(\mathbf{a})=H'(\mathbf{a})$.

(iv) Approximation Theorems for the Schrödinger Operator with a Magnetic Field

The object of this section is to prove the following fact: the operator $H(\mathbf{a})$ may be approximated (in the sense of a strong resolvent convergent) be operators $H(\mathbf{a}_n)$ with smooth $\mathbf{a}_n$. All results given in the section can be found in Ref. 25. Therefore we shall confine only the formulation of the results and sketch the proofs.

Theorem 3.4: Let $\mathbf{a},\mathbf{a}_n\in L^2_{\mathrm{loc}}(\mathbb{R}^l\setminus S)$. Suppose that (1) $\|(\mathbf{a}-\mathbf{a}_n)\varphi\|\to 0$, $(n\to\infty)$ for any $\varphi\in C_0^\infty(\mathbb{R}^l\setminus S)$; (2) $H_+(\mathbf{a}_n)\xrightarrow[L^2]{R}H_+(\mathbf{a})$, $(n\to\infty)$, where $H_+(\mathbf{a})=H_0+(1+\epsilon)\mathbf{a}^2$.

Then

$$H_\epsilon(\mathbf{a}_n)\xrightarrow[L^2]{R}H_\epsilon(\mathbf{a}),\quad \epsilon>0.$$

Remark: The importance of the theorem seems to lie in that the approximation of the operator $H(\mathbf{a})$ can be reduced to the approximation of an operator of the (H_0+V)-type, which is a more habitual problem. So the approximation of the operator $H(\mathbf{a})$ by $H(\mathbf{a}_n)$, where $\mathbf{a}_n$ are smooth is possible and follows now from Theorem 3.2.

Proof: Let us use the representations:

$$g=g_+^{1/2}(1+g_+^{1/2}+\Lambda(\mathbf{a})g_+^{1/2})^{-1}g_+^{1/2},$$
$$g_n=g_{+,n}^{1/2}(1+g_{+,n}^{1/2}\Lambda(\mathbf{a}_n)g_{+,n}^{1/2})^{-1}g_{+,n}^{1/2},$$

where

$$g_n=(H_\epsilon(\mathbf{a}_n)+\lambda)^{-1},\ g_{+,n}=(H_+(\mathbf{a}_n+\lambda)^{-1},$$
$$g=(H_\epsilon(\mathbf{a})+\lambda)^{-1},\ g_+=(H_+(\mathbf{a})+\lambda)^{-1}.$$

As $g_{+,n}\xrightarrow[L^2]{}g_+$ and Lemma 3.3 implies the estimate

$$\sup_{n\geqslant 1}\|(1+g_{+,n}^{1/2}\Lambda(\mathbf{a}_n)g_{+,n}^{1/2})^{-1}\leqslant\frac{\sqrt{1+\epsilon}}{\sqrt{1+\epsilon}-1}$$

our problem reduces to the proof of the following:

$$g_{+,n}^{1/2}\Lambda(\mathbf{a}_n)g_{+,n}^{1/2}\xrightarrow[L^2]{s}g_+^{1/2}\Lambda(\mathbf{a})g_+^{1/2}.$$

Hence it is sufficient to prove that

$$\nabla g_{+,n}^{1/2}\xrightarrow[L^2]{s}\nabla g_+^{1/2}\tag{3.2}$$

and

$$\mathbf{a}_n g_{+,n}^{1/2}\xrightarrow[L^2]{s}\mathbf{a}g_+^{1/2}.\tag{3.3}$$

Let $f\in L^2(\mathbb{R}^l)$, $u_n=g_{+,n}^{1/2}f$. Then (2) implies

$$u_n\xrightarrow[L^2]{s}u=g_+^{1/2}f.$$

Now we shall prove

$$i\nabla u_n\xrightarrow[L^2]{w}i\nabla u,$$
$$\mathbf{a}_n u_n\xrightarrow[L^2]{w}\mathbf{a}u\quad(\to\text{ is weak convergence}).\tag{3.4}$$

The estimates

$$\|\nabla u_n\|\leqslant\|f\|,\quad\|\mathbf{a}_n u_n\|\leqslant\|f\|$$

imply that for the proof of (3.4) it is sufficient to show the following:

$$\langle i\nabla u_n,\varphi\rangle\to\langle i\nabla u,\varphi\rangle;$$
$$\langle\mathbf{a}_n u_n,\varphi\rangle\to\langle\mathbf{a}u,\varphi\rangle$$

for any $\varphi\in\epsilon$, where ϵ is dense in $L^2(\mathbb{R}^l)$.

Let $\epsilon=C_0^\infty(\mathbb{R}^l\setminus S)$, $\varphi\in\epsilon$. Using $u\in\mathscr{D}(\nabla)\cap\mathscr{D}(\mathbf{a})$ we obtain

$$\lim_{n\to\infty}\langle i\nabla u_n,\varphi\rangle=\lim_{n\to\infty}\langle u_n,i\nabla\varphi\rangle=\langle u,i\nabla\varphi\rangle=\langle i\nabla u,\varphi\rangle,$$

$$\lim_{n\to\infty}\langle\mathbf{a}_n u_n,\varphi\rangle=\lim_{n\to\infty}\langle u_n,\mathbf{a}_n\varphi\rangle=\langle u,\mathbf{a}\varphi\rangle=\langle\mathbf{a}u,\varphi\rangle.$$

We now prove (3.2) and (3.3). Using the identity

$$\|u_n\|^2+\|\nabla u_n\|^2+(1+\epsilon)\|\mathbf{a}_n u_n\|^2=\|f\|^2=\|u\|^2+\|\nabla u\|^2+(1+\epsilon)\|\mathbf{a}u\|^2,$$

we have

$$\begin{aligned}&\|\nabla(u_n-u)\|^2+(1+\epsilon)\|\mathbf{a}_n u_n-\mathbf{a}u\|^2\\&=\|\nabla u_n\|^2+\|\nabla u\|^2+(1+\epsilon)\|\mathbf{a}_n u_n\|^2+(1+\epsilon)\|\mathbf{a}u\|^2\\&\quad-2\,\mathrm{Re}\langle\nabla u,\nabla u_n\rangle-2(1+\epsilon)\mathrm{Re}\langle\mathbf{a}_n u_n,\mathbf{a}u\rangle\\&=2(\|R\|^2-\tfrac{1}{2}\|u_n\|^2-\tfrac{1}{2}\|u\|^2-\mathrm{Re}\langle\nabla u,\nabla u_n\rangle\\&\quad-(1+\epsilon)\mathrm{Re}\langle\mathbf{a}_n u_n,\mathbf{a}u\rangle)\\&\to 2(\|R\|^2-\|u\|^2-\|\nabla u\|^2-(1+\epsilon)\|\mathbf{a}u\|^2)\equiv 0.\end{aligned}$$

Theorem 3.4 is proved. ■

Theorem 3.4 and the definition of the operator $H(\mathbf{a})$ imply

Theorem 3.5: Suppose that all the assumptions of Theorem 3.4 are valid. Then

$$H(\mathbf{a})=\text{R-lim}_{\epsilon\downarrow 0}\lim_{n\to\infty}H_\epsilon(\mathbf{a}_n).$$

Corollary 3.6: Let $\mathbf{a}\in L^2_{\rm loc}(\mathbb{R}^l\setminus S)$. Then

$$|e^{-tH(\mathbf{a})}f|\leqslant e^{-tH_0}|f|, \tag{3.5}$$

for all $f\in L^2(\mathbb{R}^l)$.

Proof: Let $\mathbf{a}$ be smooth. Then the proof of (3.5) based on the Feynman–Kac–Ito formula is given in Ref. 26 (see also Sec. 4). Thus corollary follows from Theorem 3.5. ∎

Subsequently we shall consider the operator $(i\nabla+\mathbf{a})^2+V$, which is more general than $(i\nabla+\mathbf{a})^2$, where V is a multiplication operator. For a correct definition of the form sum we need the following simple

Propositon 3.7: Let V be a multiplicative operator such that $V\in PK(H_0)$. Then $V\in PK(H(\mathbf{a}))$.

Proof. Using the representation

$$(H(\mathbf{a})+\lambda)^{-1/2}=\frac{1}{\Gamma(\frac{1}{2})}\int_0^\infty e^{-tH(\mathbf{a})}e^{-\lambda t}t^{-1/2}\,dt,$$

and Corollary 3.6 we obtain

$$|(H(\mathbf{a})+\lambda)^{-1/2}u|\leqslant(H_0+\lambda)^{-1/2}|u|,\quad \forall u\in L^2(\mathbb{R}^l),\ \lambda>0.$$

Hence

$$\|(H(\mathbf{a})+\lambda)^{-1/2}V(H(\mathbf{a})+\lambda)^{-1/2}\|\leqslant\|(H_0+\lambda)^{-1/2}V(H_0+\lambda)^{-1/2}<1.$$ ∎

Suppose that $0\leqslant V^\pm\in L^1_{\rm loc}(\mathbb{R}^l\setminus S)$ and $V^-\in PK(H_0)$. Using Proposition 3.7 it is not difficult to see that the definition of the operator $H=H(\mathbf{a})+V^+-V^-$ as a form sum is coorect. Moreover, we have

Theorem 3.8: Let $\mathbf{a}\in L^2_{\rm loc}(\mathbb{R}^l\setminus S)$, $0\leqslant V^\pm\in L^1_{\rm loc}(\mathbb{R}^l\setminus S)$, $V^-\in PK(H_0)$. Suppose that $V_n^\pm$ $(n=1,2,\cdots)$ are corresponding truncated operators and the consequence $\{\mathbf{a}_n\}$ obeys all assumptions of Theorem 3.4. Then

$$H(\mathbf{a})+V^+-V^-=\text{R-}\lim_{\epsilon\downarrow 0}\lim_{k\to\infty}\lim_{n\to\infty}(H_\epsilon(\mathbf{a}_k)+V_n^+-V_n^-)$$

Proof: It is a direct corollary of the above mentioned approximation theorems.

Remark: It was shown by T. Kato[27] (see also Refs. 28 and 26) that in the case $S=\phi$ it is possible to omit the $\lim_{\epsilon\downarrow 0}$.

4. SOME PRELIMINARY INFORMATION.

A. σ_r-spaces

Further we shall mainly use some properties of the σ_r-spaces. For the completeness we shall give here some definitions and facts.

An operator X is called the trace (or $X\in\sigma_1$) if and only if there is an orthonormal basis $\{\psi_m\}_{m=1}^\infty$ such that

$$\text{tr}\,X\equiv\sum_{m=1}^\infty\langle|X|\psi_m,\psi_m\rangle<\infty.$$

We shall denote $X\in\sigma_r$ $(1\leqslant r<\infty)$ if and only if $X^r\in\delta_1$. A norm in the space δ_r is introduced by the following way: $|||X|||_r=(\text{tr}X^r)^{1/2}$.

Each σ_r is double ideal in the space $\mathscr{L}(L^2)$ and, moreover, the Hölder inequality

$$|||X_1\cdots X_n|||_r\leqslant|||X_1|||_{r_1}\cdots|||X_n|||_{r_n}, \tag{4.1}$$

where $r^{-1}=\sum_{j=1}^n r_j^{-1}$ (for $r=\infty$ we shall put $|||\cdot|||_r=\|\cdot\|$), is valid. The following simple proposition is valid:

Proposition 4.1: (See, for example, Ref. 29). Let $X_n\in\sigma_r$, $1\leqslant r<\infty$, $n=1,2,\ldots$, $X_n\overset{s}{\to}X$ and $\sup_{n\geqslant 1}|||X_n|||_r<\infty$. Then $X\in\sigma_r$.

We shall also formulate a theorem which was essentially proved in Refs. 10 and 22:

Theorem 4.2: Suppose that $\mathbf{a}\in L^2_{\rm loc}(\mathbb{R}^l\setminus S)$, $V\in L^1_{\rm loc}(\mathbb{R}^l\setminus S)$ and $V^-\in PK(H_0)$. Let $s>0$, $\delta\in\mathbb{R}^1$. Then

$$(1+|x|^2)^{\delta-s}(H(\mathbf{a})+V+\lambda)^{-s}(1+|x|^2)^{-s}\in\sigma_{r_s},$$

where $r_s>\frac{l}{2s}$.

Remark: In Refs. 10 and 22 this theorem was formulated for the case $\mathbf{a}=0$, $V\in PK(H_0)$, but it is easy to extend the theorem for our assumptions.

Remark: As $e^{-t(H(\mathbf{a})+V)}$ is a holomorphic semigroup in $L^2(\mathbb{R}^l)$, $(H(\mathbf{a})+V+\lambda)^s e^{-t(H(\mathbf{a})+V)}\in\mathscr{L}(L^2)$. Thus Theorem 4.2 and the representation

$$\begin{aligned}(1+|x|^2)^{-s}e^{-t(H(\mathbf{a})+V)}&=[(1+|x|^2)^{-s}(H(\mathbf{a})+V+\lambda)^s]\\&=[(H(\mathbf{a})+V+\lambda)^s e^{-t(H(\mathbf{a})+V)}]\end{aligned}$$

imply $(1+|x|^2)^{-s}e^{-t(H(\mathbf{a})+V)}\in\sigma_2$ for any $s>l/4$.

In particular, it follows that $e^{-t(H(\mathbf{a})+V)}$ is the integral operator.

Below we shall also use the following well-known

Proposition 4.3: Suppose that operators F, G in $L^2(\mathbb{R}^l)$ obey the assumption

$$|Gf|\leqslant F|f|\ \text{for any}\ f\in L^2(\mathbb{R}^l)$$

and $F\in\sigma_{2k}$ for some integer $k\geqslant 1$. Then $G\in\sigma_{2k}$.

Remark: For the noninteger K the validity of the proposition is open and for $\frac{1}{2}\leqslant K<1$ the proposition is false. There is an interesting and complete discussion in Ref. 30.

B. Integral properties

Using Corollary 3.6 it is clear that

$$|e^{-t(H(\mathbf{a})+V)}f|\leqslant e^{-tH_0}|f|$$

for $\forall f\in L^2(\mathbb{R}^l)$, where $\mathbf{a}\in L^2_{\rm loc}(\mathbb{R}^l\setminus S)$, $0\leqslant V\in L^1_{\rm loc}(\mathbb{R}^l\setminus S)$. This estimate and the properties of e^{-tH_0} imply that $e^{-t(H(\mathbf{a})+V)}$ is an integral operator whose integral kernel $e^{-t(H(\mathbf{a})+V)}(x,y)$ obeys the estimate

$$\sup_{(x,y)\in\mathbb{R}^l\times\mathbb{R}^l}|e^{-t(H(\mathbf{a})+V)}(x,y)|<\infty.$$

C. The Feynman–Kac-Ito formula

In this section we shall consider the Feynman–Kac–Ito formula (FKI) which has a decisive role in our proof of the completeness of $W_\pm(H_0,H(\mathbf{a})+V)$.

Let μ_0 be a Wiener measure on the path space Ω_0, which consists of all paths ω such that $[0,\infty]\ni s\to\omega_s\in\mathbb{R}^l$ with $\omega(0)=0$.

Theorem 4.4: Let $\mathbf{a}\in L^2_{\rm loc}(\mathbb{R}^l\setminus S)$, $V\in L^1_{\rm loc}(\mathbb{R}^l\setminus S)$, $V^-\in PK(H_0)$, $f\in L^2(\mathbb{R}^l)$. Introduce the sets:

$$\Omega_0(\mathbf{a}^2,x)=\{\omega\in\Omega_0:\int_0^t\mathbf{a}^2(\omega_s+x)\,ds=\infty\},$$

$$\Omega_0^c(\mathbf{a}^2,x)=\Omega_0\setminus\Omega_0(\mathbf{a}^2,x),$$

$$\Omega_0(V,\mathbf{a}^2,x)=\{x\in\Omega_0^c(\mathbf{a}^2,x):\int_0^t V^+(\omega_s+x)\,ds=\infty\}.$$

Then we have (pointwise a.e.)

$$(e^{-t(H_0+V)}f)(x) = \int_{\Omega_0^c(V^+,x)} f(\omega_t + x) \times \exp\left[-\int_0^t V(\omega_s + x)\,ds\right] d\mu_0(\omega), \quad \text{(FK)}$$

$$(e^{-tH(\mathbf{a})}f)(x) = \int_{\Omega_0^c(\mathbf{a}^2,x)} f(\omega_t + x) U(\mathbf{a},\omega,t,x)\,d\mu_0(\omega), \quad (\text{FKI}_1)$$

$$(e^{-t(H(\mathbf{a})+V)}f)(x) = \int_{\Omega_0^c(\mathbf{a}^2,x)\setminus\Omega_0(V,\mathbf{a}^2,x)} f(\omega_t + x) \times \exp\left[-\int_0^t V(\omega_s + x)\,ds\right] U(\mathbf{a},\omega,t,x)\,d\mu_0(\omega), \quad (\text{FKI}_2)$$

where $U(\mathbf{a},\omega,t,x)\in L^\infty(\Omega_0,d\mu_0)$ for a.e. $x\in\mathbb{R}^l$,

$$\|U(\mathbf{a},\omega,t,x)\|_{L^\infty(\Omega_0,d\mu 0)} \leqslant 1$$

and formally (if $\mathbf{a}$ is smooth it is in fact)

$$U(\mathbf{a},\omega,t,x) = \exp(-B),$$

$$B = \frac{i}{2}\int_0^t \operatorname{div}\mathbf{a}(\omega_s + x)\,ds + i\int_0^t \mathbf{a}(\omega_s + x)\,d\boldsymbol{\omega}.$$

(Here $\int_0^t \mathbf{a}(\omega_s + x)\,d\boldsymbol{\omega}$ is the Ito stochastic integral.)

We shall give a sketch of the proof. If $\mathbf{a}$, V is smooth, FK and FKI are well known. So our problem is the investigation of the singular $\mathbf{a}$, V. Note that for $l = 3$, $V\in L^2(\mathbb{R}^l) + L^\infty(\mathbb{R}^l)$ the proof of FK is contained in Reed-Simon,[31] but this proof is also valid in our case.

Let us prove the FKI_1. Consider first the case $\mathbf{a}\in L^\infty(\mathbb{R}^l)$. Let $\{\mathbf{a}_n\}$ be a sequence of the C_0^∞ vector-valued functions such that $\|(\mathbf{a}-\mathbf{a}_n)\varphi\|\to 0$ for any $\varphi\in C_0^\infty(\mathbb{R}^l)$, then

$$H_0 + \mathbf{a}_n^2 \xrightarrow[L^2]{R} H_0 + \mathbf{a}^2, H_\epsilon(\mathbf{a}) = R - \lim_{n\to\infty} H_\epsilon(\mathbf{a}_n) \quad \text{and}$$

$$\varphi_n(\omega) = f(\omega_t + x)\exp\left[-\epsilon\int_0^t \mathbf{a}_n^2(\omega_s + x)\,ds\right] \to \varphi(\omega)$$
$$= f(\omega_t + x)\exp\left[-\epsilon\int_0^t \mathbf{a}^2(\omega_s + x)\,ds\right]$$

for μ_0 − a.e. $\omega\in\Omega_0$ and a.e. $x\in\mathbb{R}^l$. Choosing a suitable subsequence we can consider $(e^{-tH_\epsilon(\mathbf{a}_n)}f)(x)\to(e^{-tH_\epsilon(\mathbf{a})}f)(x)$ pointwise a.e. For $H_\epsilon(\mathbf{a}_n)$ we have the usual Ito formula

$$(e^{-tH_\epsilon(\mathbf{a}_n)}f)(x) = \int_{\Omega_0} f(\omega_t + x)e^{-\epsilon\int_0^t \mathbf{a}_n^2(\omega_s + x)\,ds} U(\mathbf{a}_n,\omega,t,x)\,d\mu_0(\omega)$$

for every $f\in L^2(\mathbb{R}^l)$, a.e. $x\in\mathbb{R}^l$, where $U(\mathbf{a}_n,\omega,t,x) = e^{-B_n}$

$$B_n = \frac{i}{2}\int_0^t \operatorname{div}\mathbf{a}_n(\omega_s + x)\,ds + i\int_0^t \mathbf{a}_n(\omega_s + x)\,d\boldsymbol{\omega}.$$

Let us introduce the Banach spaces $L^p(\Omega_0,d\mu 0)$, $1\leqslant p<\infty$. Denote a value of the functional $\beta\in L^p(\Omega_0,d\mu_0)$ on the vector $v\in L^{p'}(\Omega_0,d\mu_0)$, $(1/p)+(1/p')=1$ by $\langle v,\beta\rangle$. Let $t>0$ be fixed. An element $U_n(\omega)\equiv U(\mathbf{a}_n,\omega,t,x)$ belongs to any $L^q(\Omega_0,d\mu_0)$ and, clearly, $\|U_n\|_{L^\infty(\Omega_0,d\mu_0)} = 1$. The space $L^p(\Omega_0,d\mu_0)$ is weakly compact, so that there is a subsequence $\{U_{n'}\}$ which has a weak limit in $L^p(\Omega_0,d\mu_0)$:

$$\langle v,U_{n'}\rangle\to\langle v,u\rangle,$$
$$u\in L^p(\Omega_0,d\mu_0),\ \forall v\in L^{p'}(\Omega_0,d\mu_0).$$

By choosing $v=\varphi_n(\omega)$ and using $\phi_n \xrightarrow[L^2]{s} \phi$ we obtain

$$\langle\varphi_{n'},U_{n'}\rangle\to\langle V,u\rangle$$

and

$$\langle\varphi_{n'},U_{n'}\rangle\equiv\exp(-tH_\epsilon(\mathbf{a}_{n'}))f\to\exp(-tH_\epsilon(\mathbf{a}))f.$$

As weak $\lim_{n'\to\infty}\langle\varphi,U_{n'}\rangle = \lim_{n'\to\infty}\langle\varphi_{n'},u_{n'}\rangle = \langle\varphi,U\rangle$ does not depend on the subsequence $\{n'\}$, then

$$\langle\varphi_n,U_n\rangle\to\langle\varphi,u\rangle.$$

Thus

$$\langle\varphi,u\rangle = [\exp(-tH_\epsilon(\mathbf{a}))f](x) = \int_{\Omega_0} f(\omega_t + x)e^{-\epsilon\int_0^t \mathbf{a}^2(\omega_s + x)\,ds} U(\mathbf{a},\omega,t,x)\,d\mu_0(\omega)$$

for μ_0 − a.e. $\omega\in\Omega_0$ and a.e. $x\in\mathbb{R}^l$. It is clear that

$$\|U\|_{L^\infty(\Omega_0,d\mu_0)} = 1.$$

Let now $\mathbf{a}\in L^2_{\text{loc}}(\mathbb{R}^l\setminus S)$ and $\{\mathbf{a}_n\}$, a sequence of the L^∞ functions, obey assumptions of Theorem 3.4. (For example, $\mathbf{a}_n$ is a truncated operator corresponding to $\mathbf{a}$.) Now we can repeat the above mentioned procedure and construct $U(\mathbf{a},\omega,t,x)\in L^\infty(\Omega_0,d\ \mu_0)$, $\|U\|_{L^\infty(\Omega_0,d\mu_0)} = 1$, such that

$$(e^{-tH_\epsilon(\mathbf{a})}f)(x) = \int_{\Omega_0^c(\mathbf{a}^2,x)} f(\omega_t + x)e^{-\epsilon\int_0^t \mathbf{a}^2(\omega_s + x)\,ds} U(\mathbf{a},\omega,t,x)\,d\mu_0(\omega)$$

for μ_0-a.e., ω, a.e. $x\in\mathbb{R}^l$ and every $f\in L^2(\mathbb{R}^l)$. To see this we have used the following:

$$\varphi(\omega) = \begin{cases} f(\omega_t + x)\exp(-\epsilon\int_0^t \mathbf{a}^2(\omega_s + x)\,ds) & \text{if } \omega\in\Omega_0^c(\mathbf{a}^2,x), \\ 0 & \text{if } \omega\in\Omega_0(\mathbf{a}^2,x). \end{cases}$$

It is also clear that

$$f(\omega_t + x)\exp(-\epsilon\int_0^t \mathbf{a}^2(\omega_s + x)\,ds) \xrightarrow{\epsilon\downarrow 0} f(\omega_t + x)$$

for μ_0-a.e. $\omega\in\Omega_0^c(\mathbf{a}^2,x)$. and a.e. $x\in\mathbb{R}^l$.

It follows from Theorem 3.4 that $e^{-tH_\epsilon(\mathbf{a})}\to e^{-tH(\mathbf{a})}$ so using the dominated convergence theorem we obtain FKI_1.

Proof of FKI_2: The proof of FKI_1 shows that

$$(e^{-t(H_\epsilon(\mathbf{a})+V_k)}f)(x) = \int_{\Omega_0^c(\mathbf{a}^2,x)} f(\omega_t + x)u(\mathbf{a},\omega,t,x)e^{-\int_0^t (V_k+\epsilon\mathbf{a}^2)(\omega_s+x)\,ds}\,d\mu_0(\omega)$$

for any $V_k\in L^\infty(\mathbb{R}^l)$, $f\in L^2(\mathbb{R}^l)$ and a.e. $x\in\mathbb{R}^l$.

So if $\epsilon\downarrow 0$, we obtain

$$(e^{-t(H(\mathbf{a})+V_k)}f)(x) = \int_{\Omega_0^c(\mathbf{a}^2,x)} f(\omega_t + x)e^{-\int_0^t (V_k+(\omega_s+x)\,ds} U(\mathbf{a},\omega,tx)\,d\mu_0(\omega).$$

The limit approaching when $K\to\infty$ completes the proof. ■

Remark: Using arguments based on the approximation

theorem 3.5 it is easily to see that

$$\mathrm{supp}(\mathbf{a})\cap\mathrm{supp}(\mathbf{b})=\phi$$

implies

$$U(\mathbf{a}+\mathbf{b},\omega,t,x)=U(\mathbf{a},\omega,t,x)U(\mathbf{b},\omega,t,x)$$

for μ_0-a.e. $\omega\in\Omega_0$ and a.e. $x\in\mathbb{R}^l$.

Moreover, if X_R is a characteristic function of the ball $\{x:|x|\leqslant R\}$, we have

$$U(\mathbf{a},\omega,t,x)=U(X_R\mathbf{a},\omega,t,x)U((1-X_R)\mathbf{a},\omega,t,x).$$

5. PRINCIPLE OF THE DECOUPLING OF FINITE SINGULARITIES

Our proof of the completeness of wave operators is based on the following asymptotic estimate:

Lemma 5.1: Let $\mathbf{a}\in L^2_{\mathrm{loc}}(\mathbb{R}^l\setminus S)$, $0\leqslant V\in L^1_{\mathrm{loc}}(\mathbb{R}^l\setminus S)$ and $H=H(\mathbf{a})+V$. Suppose that

$$\mathrm{supp}(\mathbf{a})\cup\mathrm{supp}(V)\subset\{x:|x|\leqslant R\}$$

for some finite $R>0$. Then

$$\left|\int_{\mathbb{R}^l}[e^{-tH_0}(x,y)-e^{-tH}(x,y)]\,dy\right|\leqslant ce^{-\mathscr{K}|x|^2} \tag{5.1}$$

for every $t>0$ and a.e. $x\in\mathbb{R}^l$, where C and $\mathscr{K}>0$ are the constants depending on t and R only.

Proof: Denote

$$\Omega^1=\Omega_0^c(\mathbf{a}^2,x)\setminus\Omega_0(V,\mathbf{a}^2,x).$$

Then the following equality is valid for a.e. $x\in\mathbb{R}^l$:

$$\int_{\mathbb{R}^l}[e^{-tH_0}(x,y)-e^{-tH}(x,y)]\,dy=1-\int_{\Omega^1}e^{-\int_0^t V(\omega_s+x)\,ds}U(\mathbf{a},\omega,t,x)\,d\mu_0(\omega).$$

In fact, let X_n be a characteristic function of the ball $\{x:|x|\leqslant n\}$. By dominated convergence theorem and (FKI_2) we have

$$\begin{aligned}\int_{\mathbb{R}^l}e^{-tH}(x,y)\,dy&=\lim_{n\to\infty}\int_{\mathbb{R}^l}e^{-tH}(x,y)X_n(y)\,dy\\&=\lim_{n\to\infty}\int_{\Omega^1}U(\mathbf{a},\omega,t,x)\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]\\&\quad\times X_n(\omega_t+x)\,d\mu 0(\omega)\\&=\int_{\Omega^1}U(\mathbf{a},\omega,t,x)\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]d\mu 0(\omega).\end{aligned}$$

Let $|x|>2R$. Let us decompose the path space Ω_0 on two subsets $\Omega_0^{(1)}$ and $\Omega_0^{(2)}$ in the following way

$$\Omega_0^{(1)}=\{\omega\in\Omega_0:\inf_{0\leqslant s\leqslant t}|\omega_s+x|\geqslant R\},$$

$$\Omega_0^{(2)}=\Omega_0\setminus\Omega_0^{(1)}=\{\omega\in\Omega_0:\inf_{0\leqslant s\leqslant t}|\omega_s+x|\leqslant R\},$$

and introduce the set

$$\Omega_0^{(3)}=\{\omega\in\Omega_0:\sup_{0\leqslant s\leqslant t}|\omega_s|\leqslant|x|/2\}.$$

It follows from the inequalities

$$|\omega_s+x|\geqslant\||x|-|\omega_s|\|,$$

$$|x|-R\geqslant|x|/2,$$

that

$$\Omega_0^{(2)}\subset\{\omega\in\Omega_0:\inf_{0\leqslant s\leqslant t}\||x|-|\omega_s|\|\leqslant R\}\subset\Omega_0^{(3)}.$$

It is clear that if $\omega\in\Omega_0^{(1)}$ and $0\leqslant s\leqslant t$, then $\mathbf{a}(\omega_s+x)=0=V(\omega_s+x)$. So we have $\Omega_0^{(1)}\subset\Omega_0'$ and

$$\begin{aligned}\int_{\Omega_0'}U(\mathbf{a},\omega,t,x)\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]d\mu_0(\omega)&=\mu_0(\Omega_0^{(1)})+\int_{\Omega_0'\setminus\Omega_0^{(1)}}U(\mathbf{a},\omega,t,x)\\&\quad\times\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]d\mu_0(\omega).\end{aligned}$$

Applying the imbeddings $\Omega'\setminus\Omega_0^{(1)}\subset\Omega_0^{(2)}\subset\Omega_0^{(3)}$ and the inequality

$$\left|U(\mathbf{a},\omega,t,x)\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]\right|\leqslant 1$$

we obtain

$$\begin{aligned}&\left|1-\int_{\Omega_0'}U(\mathbf{a},\omega,t,x)\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]d\mu_0(\omega)\right|\\&\quad\leqslant|1-\mu_0(\Omega_0^{(1)})|+\left|\int_{\Omega_0'\setminus\Omega_0^{(2)}}U(\mathbf{a},\omega,t,x)\right.\\&\quad\left.\times\exp\left[-\int_0^t V(\omega_s+x)\,ds\right]d\mu_0(\omega)\right|\leqslant 2\mu_0(\Omega_0^{(3)}).\end{aligned}$$

The well-known estimate of the right part of the last inequality gives (see, for example, Ref. 32):

$$\mu_0(\Omega_0^{(3)})\leqslant C_1e^{-\mathscr{K}|x|^2},$$

$$C_1=C_1(R,t),\quad\mathscr{K}=\mathscr{K}(R,t).$$

Consider now the case $|x|\leqslant 2R$. Introduce the notation $C_2=e^{4\mathscr{K}R^2}$. Next, we have

$$\left|\int_{\mathbb{R}^l}[e^{-tH_0}(x,y)-e^{-tH}(x,y)]\,dy\right|\leqslant 2\leqslant C_2e^{-\mathscr{K}|x|^2}.$$

So by putting $C=\max\{C_1,C_2\}$ we obtain the estimate (5.1). ∎

Lemma 5.2: Suppose that all the assumptions of Lemma 5.1 are valid. Then the operator $\rho^{\mu}[e^{-tH_0}-e^{-tH}]$ (and thus $[e^{-tH_0}-e^{-tH}]\rho^{\mu}$), where $\rho(x)=(1+|x|^2)$, belongs to the Hilbert–Schmidt class σ_2 for every $t>0$ and $\mu\in\mathbb{R}^1$.

Proof: It follows from the estimates (5.1) and

$$\operatorname*{ess\,sup}_{(x,y)\in\mathbb{R}^l\times\mathbb{R}^l}|e^{-t(H(\mathbf{a})+V)}(x,y)|<\infty$$

that the integral kernel of this operator belongs to $L^2(\mathbb{R}^l\times\mathbb{R}^l)$. ∎

Theorem 5.3: Let $\mathbf{a}\in L^2_{\mathrm{loc}}(\mathbb{R}^l\setminus S)$, $0\leqslant V\in L^1_{\mathrm{loc}}(\mathbb{R}^l\setminus S)$. Suppose that $\mathbf{a}$ and V have compact supports. Then

$$e^{-tH}-e^{-t(H(\mathbf{a})+V)}\in\sigma_1$$

for every $t>0$. In particular, $W_{\pm}(H(\mathbf{a})+V,H_0)$ exist and are complete.

Proof: Denote $G=\rho^{-\mu}e^{-t(H(\mathbf{a})+V)}$, $F=\rho^{-\mu}e^{-tH_0}$

where $\mu > l/2$. Since $|Gf| \leqslant F|f|$ for every $f \in L^2(\mathbb{R}^l)$ (pointwise a.e.), so by Poposition 4.3 $G \in \sigma_2$ because $F \in \sigma_2$. By analogy $e^{-t(H(\mathbf{a})+V)}\rho^{-\mu} \in \sigma_2$, $(\mu > l/2)$. In view of the identity

$$e^{-tH_0} - e^{-tH} = [(e^{-(t/2)H_0} - e^{-(t/2)H})\rho^{\mu}][\rho^{-\mu}e^{-(t/2)H_0}] + [e^{-(t/2)H}\rho^{-\mu}][\rho^{\mu}(e^{-(t/2)H_0} - e^{-(t/2)H})] \quad [H \equiv H(\mathbf{a}) + V],$$

we see that every operator in square brackets is the Hilbert–Schmidt operator, and the product of the two Hilbert–Schmidt operators is a trace operator. A completeness of the $W_{\pm}(H, H_0)$ follows from the Birman theorem. ■

Theorem 5.4: Let $\mathbf{a} \in L^2_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, $0 \leqslant V \in L^1_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$ and $V(x)$ has a compact support. Then

$$e^{-tH(\mathbf{a})} - e^{-t(H(\mathbf{a})+V)} \in \sigma, \tag{5.2}$$

for every $t > 0$ and hence $W_{\pm}(H(\mathbf{a}) + V, H(\mathbf{a}))$ exist and are complete.

Proof: It follows from the (FKI_2) that

$$|[e^{-tH(\mathbf{a})} - e^{-t(H(\mathbf{a})+V)}]f| \leqslant [e^{-tH_0} - e^{-t(H_0+V)}]|f|$$

for every $f \in L^2(\mathbb{R}^l)$ (pointwise a.e.). In particular,

$$\rho^{\mu}[e^{-t(H(\mathbf{a})+V)} - e^{-tH(\mathbf{a})}] \in \sigma_2$$

for every $\leqslant \in \mathbb{R}^1$, because by Lemma 5.2

$$\rho^{\mu}[e^{-tH_0} - e^{-t(H_0+V)}] \in \sigma_2.$$

Moreover, combining the inequality

$$|\rho^{-\mu}e^{-t(H(\mathbf{a})+V)}f| \leqslant \rho^{-\mu}e^{-tH_0}|f|$$

and Proposition 4.3 we have

$$\rho^{-\mu}e^{-tH(\mathbf{a})} \in \sigma_2,$$

$$\rho^{-\mu}e^{-t(H(\mathbf{a})+V)} \in \sigma_2, \quad \mu > l/2.$$

Now, by analogy with the proof of Theorem 5.3 we obtain (5.2). ■

Theorem 5.5: Let $V \in L^1_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, $V^- \in PK(H_0)$ and $\mathrm{supp}(V^+) \subset \{x : |x| \leqslant R\}$ for some finite $R > 0$. Then

$$e^{-tH^-} - e^{-tH} \in \sigma_1$$

for every $t > 0$, where $H^- = H_0 - V^-$, $H = H_0 + V$.

Proof: It follows from the proof of Theorem 5.3 that it is sufficient to show the following inclusions

$$\rho^{-\mu}e^{-tH} \in \sigma_2, \quad \mu > l/2,$$

$$\rho^{\mu}[e^{-tH^-} - e^{-tH}] \in \sigma_2.$$

The first inclusion is the consequence of Theorem 4.2.

Let us consider the operator $A = \rho^{-\mu}[e^{-tH^-} - e^{-tH}]$, $\mu \in \mathbb{R}^1$. Let $\chi_v^c(\cdot) = (1 - \chi_v)(\cdot)$, where χ_v is a characteristic function of the set

$$\{\omega \in \Omega_0 : \int_0^t V^+(\omega_s + x)\, ds = \infty\}.$$

Using (FK) and the Hölder inequality we have for every $f \in L^2(\mathbb{R}^l)$

$$\begin{aligned}
&|(e^{-tH^-} - e^{-tH})f(x)| \\
&= \left| \int_{\Omega_0} e^{\int_0^t V^-(\omega_s + x)\, ds} \{1 - \chi_v^c(\omega, x) e^{-\int_0^t V^+(\omega_s + x)\, ds}\} \right. \\
&\quad \left. \times f(\omega_t + x)\, d\mu_0(\omega) \right| \\
&\leqslant \left(\int_{\Omega_0} |f(\omega_t + x)|^{\alpha p} e^{\int_0^t pV^-(\omega_s + x)\, ds} d\mu_0(\omega) \right)^{1/p} \\
&\quad \times \left(\int_{\Omega_0} |f(\omega_t + x)|^{(1-\alpha)p'} \right. \\
&\quad \left. \times \{1 - \chi_v^c(\omega, x) e^{-\int_0^t V^+(\omega_s + x)\, ds}\}^{p'} d\mu_0(\omega) \right)^{1/p'} \\
&\leqslant (e^{-t(H_0 - pV^-)}|f|^{\alpha p})^{1/p} (e^{-tH_0} - e^{-t(H_0+V^-)}) \\
&\quad \times |f|^{(1-\alpha)p'})^{1/p'}, \quad \frac{1}{p} + \frac{1}{p'} = 1,
\end{aligned}$$

where $0 < \alpha < 1$ and $p \in (1, \infty)$ is fixed by the condition $pV^- \in PK(H_0)$ (note that it is always possible). Moreover, we have used the obvious inequality

$$0 \leqslant 1 - \exp\left(- \int_0^t V^+(\omega_s + x)\, ds \right) \chi_v^c(\omega, x) \leqslant 1$$

for μ_0 – a.e. ω, a.e. $x \in \mathbb{R}^l$.

Putting $\alpha = 1/p$ and denoting

$$B = \rho^{-\mu}e^{-t(H_0 - V^-)}, \quad C = \rho^{\nu}(e^{-tH_0} - e^{-t(H_0+V^-)}),$$

$\nu = \mu(p+1)/(p-1)$, we have from the last inequality

$$|Af| \leqslant (B|f|)^{1/p}(C|f|)^{1/p'}, \frac{1}{p} + \frac{1}{p'} = 1$$

for every $f \in L^2(\mathbb{R}^l)$ (pointwise a.e.).

We must show that $A \in \sigma_2$. In fact,

$$\begin{aligned}
\sum_i \|Ae_i\|^2 &= \sum_i \langle Ae_i, Ae_i \rangle \\
&\leqslant \sum_i \langle (Be_i)^{2/p}, (Ce_i)^{2/p'} \rangle \\
&\leqslant \sum_i \|Be_i\|^{2/p}, \|Ce_i\|^{2/p'} \\
&\leqslant \left(\sum_i \|Be_i\|^2 \right)^{1/p} \left(\sum_i \|Ce_i\|^2 \right)^{1/p'}
\end{aligned}$$

for every finite sequence of characteristic functions of the in pairs disjoint sets. If L is the subspace of $L^2(\mathbb{R}^l)$ generated by $\{e_i\}$ and P_L is a corresponding orthoproyector, then

$$\begin{aligned}
|||A|||_2^2 &= \mathrm{tr}\,(AA^*) \\
&= \sup_L (P_L AA^*) = \sup_L \sum_i \|Ae_i\|^2 \\
&\leqslant \left(\sup_L \sum_i \|Be_i\|^2 \right)^{2/p} \left(\sup_L \sum_i \|Ce_i\|^2 \right)^{2/p'} \\
&= |||B|||_2^{2/p} |||C|||_2^{2/p'}.
\end{aligned}$$

we have only to remark $B, C \in \sigma_2$.

Theorem 5.6: Let $V \in L^1_{\mathrm{loc}}(\mathbb{R}^l \setminus S)$, $V^- \in PK(H_0)$ and χ_R be a characteristic function of the ball $\{x : |x| \leqslant R\}$. Then $W_{\pm}(H_0 + V, H_0)$ exist and are complete if and only if $W_{\pm}(H_0 + (1 - \chi_R)V, H_0)$ exist and are complete for arbitrary $R < \infty$.

Proof: Denote $V_{\mathrm{ex}} = (1 - \chi_R)V$, $V_R = \chi_R V$. Suppose that $W_{\pm}(H_0 + V_{\mathrm{ex}}, H_0)$ exist and are complete for some $R < \infty$. The existence and completeness of $W_{\pm}(H_0 + V, H_0)$ will be implied by the chain rule if we demonstrate the existence and the completeness of $W_{\pm}(H_0 + V_{\mathrm{ex}} - V_R^-, H_0 + V_{\mathrm{ex}})$, $W_{\pm}(H_0 + V, H_0 + V_{\mathrm{ex}} - V_R^-)$.

It follows from (FK) that

$$|e^{-t(H_0+V_{ex}-V_R)}f - e^{-t(H_0+V_{ex})}f)(x)|$$
$$= \left|\int_{\Omega_0} f(\omega_t+x)\exp\left[-\int_0^t (V_{ex}-V_R^-)(\omega_s+x)\,ds\right]\right.$$
$$\left.\{1+\exp\left[-\int_0^t V_R^-(\omega_s+x)\,ds\right]\}\right|$$
$$\leqslant \int_{\Omega_0} |f(\omega_t+x)|\exp\left[-\int_0^t V^-(\omega_s+x)\,ds\right]$$
$$\{1-\exp\left[-\int_0^t V_R^-(\omega_s+x)\,ds\right]\}$$
$$= [e^{-t(H_0-V^-)} - e^{-t(H_0-V^-+V_R^-)}]|f|.$$

Hence

$$|\rho^{\mu}[e^{-t(H_0+V_{ex}-V_R)} - e^{-t(H_0+V_{ex})}]f|$$
$$\leqslant \rho^{\mu}[e^{-t(H_0-V^-)} - (e^{-t(H_0-V^-+V_R^-)}]|f|,$$

and, obviously,

$$|\rho^{-\mu}e^{-t(H_0+V_{ex}-V_R)}f|$$
$$\leqslant \rho^{-\mu}e^{-t(H_0-V^-)}|f|, \quad \mu > l/2.$$

The operators in the right parts of the last two inequalities belong to σ_2, so that the operators in the left parts of these inequalities also belongs to σ_2. Thus the above-mentioned arguments give

$$e^{-t(H_0+V_{ex}-V_R)} - e^{-t(H_0+V_{ex})} \in \sigma_1.$$

The trace of the operator $e^{-t(H_0+V)} - e^{-t(H_0+V_{ex}-V_R)}$ follows from the inequalities:

$$|\rho^{\mu}[e^{-t(H_0+V_{ex}-V_R)} - e^{-t(H_0+V)}]f|$$
$$\leqslant \rho^{\mu}[e^{-t(H_0-V^-)} - (e^{-t(H_0-V^-+V_R^-)}]|f|;$$
$$|\rho^{-\mu}e^{-t(H_0+V)}f| \leqslant \rho^{-\mu}e^{-t(H_0-V^-)}|f|,$$

which are the consequence of the (FK) and the above-mentioned argurments.

We have only to apply the chain rule and Birman theorem to obtain the existence and the completeness of $W_{\pm}(H_0+V, H_0)$.

The proof of the inverse assertion is analogous. ■

Theorem 5.7 (*The general principle of the decoupling of finite singularities*): Let $\mathbf{a}\in L^2_{loc}(\mathbb{R}^l\setminus S)$, $V\in L^1_{loc}(\mathbb{R}^l\setminus S)$, $V^-\in PK(H_0)$. Let χ_R be a characteristic function of the ball $\{x:|x|\leqslant R\}$. Then $W_{\pm}(H(\mathbf{a})+V)$ exist and are complete if and only if $W_{\pm}[H((1-\chi_R)\mathbf{a}) + (1-\chi_R)V, H_0]$ exist and are complete for arbitrary $R<\infty$.

Proof: Introduce the notations

$$\mathbf{a}_{ex} = (1-\chi_R)\mathbf{a}, \quad \mathbf{a}_R = \chi_R\mathbf{a},$$
$$V_{ex} = (1-\chi_R)V, \quad V_R = \chi_R V.$$

Using the chain rule we obtain

$$W_{\pm}(H_0, H(\mathbf{a}_{ex}) + V_{ex})$$
$$= W_{\pm}(H_0, H(\mathbf{a})+V)W_{\pm}(H(\mathbf{a})+V, H(\mathbf{a}_{ex})+V)$$
$$\times W_{\pm}(H(\mathbf{a}_{ex})+V, H(\mathbf{a}_{ex})+V_{ex}),$$
$$W_{\pm}(H_0, H(\mathbf{a})+V)$$
$$= W_{\pm}(H_0, H(\mathbf{a}_{ex})+V_{ex})W_{\pm}(H(\mathbf{a}_{ex})$$
$$+ V_{ex}, H(\mathbf{a}_{ex})+V)W_{\pm}(H(\mathbf{a}_{ex})+V, H(\mathbf{a})+V),$$

So we need to show the existence and the completeness of the operators $W_{\pm}(H(\mathbf{a})+V, H(\mathbf{a}_{ex})+V)$, $W_{\pm}(H(\mathbf{a}_{ex})+V, H(\mathbf{a}_{ex})+V_{ex})$. By the Birman theorem and the above arguments it is sufficient to show that

$$\rho^{-\mu}\exp[-t(H(\mathbf{a})+V])\in\sigma_2, \quad \mu > l/2$$
(it is the consequence of Theorem 4.2)

and

$$A \equiv \rho^{\mu}[\exp[-t(H(\mathbf{a})+V)]]$$
$$-\exp[-t(H(\mathbf{a}_{ex})+V))]\in\delta_2,$$
$$A_1 \equiv \rho^{\mu}[\exp[-t(H(\mathbf{a})+V)]]$$
$$-\exp[-t(H(\mathbf{a})+V_{ex}))]\in\delta_2.$$

Lemma 5.8: $A\in\sigma_2$ and $A_1\in\sigma_2$.

Proof: Denote $\chi^c_{\mathbf{a}}(\omega) = (1-\chi_{\mathbf{a}})(\omega)$, where $\chi_{\mathbf{a}}(\omega)$ is the characteristic function of the set $\Sigma\omega\in\Omega_0: \int_0^t \mathbf{a}^2(\omega_s+x)\,ds = \infty\}$. Let $p>1$ and $pV^-\in PK(H_0)$. Using (FKI_2), the Hölder inequality, the property

$$U(\mathbf{a},\omega,t,x) = U(\mathbf{a}_{ex},\omega,t,x)U(\mathbf{a}_R,\omega,t,x)$$

and the inequalities $|U(\mathbf{a}_{ex},\omega,t,x)|\leqslant 1$, $|1-U(\mathbf{a}_R,\omega,t,x)|\leqslant 2$, we have

$$|(e^{-t(H(\mathbf{a})+V)} - e^{-t(H(\mathbf{a}_{ex})+V)})f(x)|$$
$$\leqslant \int_{\Omega_0} |f(\omega_t+x)|\exp\left(\int_0^t V^-(\omega_s+x)\,ds\right)$$
$$\times|1-U(\mathbf{a}_R,\omega,t,x)\chi^c_{\mathbf{a}_R}(\omega)|\,d\mu_0(\omega)$$
$$\leqslant 2^{1/p}(e^{-t(H_0 pV^-)}|f|)^{1/p}$$
$$\cdot\left(\int_{\Omega_0} |f(\omega_t+x)||1-U(\mathbf{a}_R,\omega,t,x)\chi^c_{\mathbf{a}_R}(\omega)|\,d\mu_0(\omega)\right)^{1/p'}$$

for any $f\in L^2(\mathbb{R}^l)$ (pointwise a.e.).

Consider the operator $N: L^q\to L^q$ defined by the equality

$$(Nf)(x) = \int_{\Omega_0} f(\omega_t+x)|$$
$$|1-U(\mathbf{a}_R,\omega,t,x)\chi^c_{\mathbf{a}_R}(\omega)|\,d\mu_0(\omega), \quad f\in L^q(\mathbb{R}^l).$$

It is clear that $N|f|\geqslant 0$ and $|Nf|\leqslant 2^{-tH_0}|f|$. The latter implies that N is an integral operator and its integral kernel obeys the estimate

$$\operatorname*{ess\,sup}_{(x,y)\in\mathbb{R}^l\times\mathbb{R}^l} |N(x,y)| < \infty.$$

Moreover, by analogy with the proof of Lemma 5.1 we obtain the estimate

$$0\leqslant\int_{\mathbb{R}^l} N(x,y)\,dy \leqslant Ce^{-\mathscr{K}|x|^2} \quad \text{for a.e. } x\in\mathbb{R}^l,$$

where $C, \mathscr{K} > 0$ depend on R, T only.

Thus $\rho^{\mu}N, N\rho^{\mu}\in\sigma_2$ for every $\mu\in\mathscr{R}^1$. Putting

$$A = \rho^{\mu}[\exp[-t(H(\mathbf{a})+V)]]$$
$$-\exp[-t(H(\mathbf{a}_{ex})+V))],$$
$$B = 2\rho^{-\mu}\exp(-t(H_0-V^-)), \quad C = \rho^{-\nu}N,$$
$$\nu = \mu\frac{p+1}{p-1}, \quad \mu > l/2,$$

and using (5.3) it is not difficult to see that $|Af|\leqslant(B|f|)^{1/p}(C|f|)^{1/p'}$ and $B, C\in\sigma_2$ so that $A\in\sigma_2$. Let us prove the inclusion $A_1\in\sigma_2$. It follows from (FKI_2) that

$$|e^{-t(H(\mathbf{a})+V_{ex})} - e^{-t(H(\mathbf{a})+V)})f|$$

$$\leqslant (e^{-t(H_0 - V^-)} - e^{-t(H_0 - V^- + V_R)})|f|$$
$$+ (e^{-t(H_0 - V^-)} - e^{-t(H_0 - V^- + V^- - R)})|f|.$$

Combining the last inequality and Theorem 5.6 we obtain $A_1 \in \sigma_2$. Thus Lemma 5.8 and Theorem 5.7 are proved. ■

6. ONE SUFFICIENT CRITERION OF THE COMPLETENESS OF THE WAVE OPERATORS $W_\pm(H(\mathbf{a}) + V, H_0)$

In Sec. 4 we solved the problem of decoupling of finite singularities for the potential V and magnetic field $\mathbf{a}$. Thus the probelm of the completeness of the $W_\pm(H(\mathbf{a}) + V, H_0)$ is equivalent to the definition of conditions which guarantee the validity of the assumption (2) of Theorem 2.1. In this direction we shall prove the following theorem:

Theorem 6.1: Let $\rho(x) = (1 + |x|^2)^{1/2}$. Suppose that $\mathbf{b} = \rho^{-\nu}\mathbf{b}_0$, $Q = \rho^{-\nu}Q_0$ for some $\nu > l$ and the functions $\mathbf{b}_0$, Q_0 obey the assumptions

$$4\mathbf{b}_0^2 \in PK(H_0),$$
$$Q_0 \in PK(H_0).$$

Then

$$(H(\mathbf{b}) + Q + \lambda)^{-k} - (H_0 + \lambda)^{-k} \in \sigma_1$$

for all integers $k > l/2$ and all sufficiently large $\lambda > 0$. In particular, $W_\pm(H(\mathbf{b}) + Q, H_0)$ exist and are complete.

Proof: Introduce the operators $H_\epsilon(\mathbf{b})$, $H_+ \equiv H_0 + (1 + \epsilon)\mathbf{b}^2$, $\epsilon > 0$, where ϵ is fixed by the condition $(1 + \epsilon)\mathbf{b}_0^2 \in PK(H_0)$. Then it is sufficient to show that

$$(H(\mathbf{b}) + Q + \lambda)^{-k} - (H(\mathbf{b}) + \lambda)^{-k} \in \sigma_1, \tag{6.1}$$
$$(H(\mathbf{b}) + \lambda)^{-k} - (H_\epsilon(\mathbf{b}) + \lambda)^{-k} \in \sigma_1, \tag{6.2}$$
$$(H_\epsilon(\mathbf{b}) + \lambda)^{-k} - (H_+ + \lambda)^{-k} \in \sigma_1, \tag{6.3}$$
$$(H + \lambda)^{-k} - (H_0 + \lambda)^{-k} \in \sigma_1. \tag{6.4}$$

We shall consider (6.3) only, since other inclusions are analogous to (6.3) and their proofs are more simple. Let $\mathbf{b}_n \in \rho^{-\nu}\mathbf{b}_{0,n}$, $\mathbf{b}_{0,n} \in C_0^\infty(\mathbb{R}^l)$ and

$$(H_\epsilon(\mathbf{b}_n) + \lambda)^{-k} - (H_{+n} + \lambda)^{-k} \xrightarrow[L^2]{s} (H_\epsilon(\mathbf{b}) + \lambda)^{-k} - (H_+ + \lambda)^{-k},$$

where $H_{+,n} = H_0 + (1 + \epsilon)\mathbf{b}_n^2$. (It is possible by Theorems 3.4 and 3.2.)

Thus by Proposition 4.1 it is sufficient to show that

$$Z_n \equiv (H_\epsilon(\mathbf{b}_n) + \lambda)^{-k} - (H_{+,n} + \lambda)^{-k} \in \sigma_1$$

and $\sup_n |||Z_n|||_1 < \infty$.

Denote $g = (H_\epsilon(\mathbf{b}_n + \lambda)^{-1}$, $g_+ = (H_{+,n} + \lambda)^{-1}$. Using the representation $g = g_+^{1/2}(1 + g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2})^{-1}g_+^{1/2}$ we have

$$g^k - g_+^k = \sum_{j=0}^{k-1} g^{k-1-j}(g - g_+)g_+^j$$
$$= \sum_{j=0}^{k-1} g^{k-1-j}g_+^{1/2}\{[1 + g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2}]^{-1} - 1\} \times g_+^{j+1/2}$$
$$= \sum_{j=0}^{k-1} g^{k-1-j}g_+^{1/2}\{\sum_{m=1}^{\infty}(-1)^m [g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2}]^m\} g_+^{j+1/2}.$$

Denote

$$\alpha_j = \frac{2k - (2j+1)}{2k}\nu, \quad \beta_j = \frac{2j+1}{2k}\nu.$$

We have

$$g^k - g_+^k = \sum_{j=0}^{k-1} g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j} \times \{\sum_{m=1}^{\infty}(-1)^m [\rho^{\alpha_j}g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2}\rho^{\beta_j}]^m\} \times \rho^{-\beta_j}g_+^{j+1/2}.$$

It follows from Lemma A1 (see Appendix) that

$$\|\rho^{\alpha_j}g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2}\rho^{\beta_j}\| \leqslant C < 1$$

for all sufficiently large $\lambda > 0$, where C is n independent. Hence, the series

$$\sum_{m=1}^{\infty}(-1)^m [\rho^{\alpha_j}g_+^{1/2}\Lambda(\mathbf{b}_n)g_+^{1/2} + \rho^{\beta_j}]^m$$

converges to some bounded operator T_j. Note that $\|T_j\|$ are dominant by the n independent constants. Thus

$$g^k - g_+^k = \sum_{j=0}^{k-1} g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j}T_j\rho^{-\beta_j}g_+^{j+1/2}.$$

It follows from Theorem 4.2 that

$$\rho^{-\alpha_j}g_+^{1/2} \in \sigma_{r_{j,1}}, \tag{6.5}$$
$$g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j} \in \sigma_{r_{j,2}}, \tag{6.6}$$

where

$$r_{j,1}^{-1} < \frac{1}{l}\min\{\beta_j, 2j+1\},$$
$$r_{j,2}^{-1} < \frac{1}{l}\min\{\alpha_j, 2k - 2j - 1\}.$$

[(6.5) is a direct consequence of Theorem 4.2.] Equation (6.6) follows from the identity

$$g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j} \equiv [g^{k-1-j}\rho^{-\gamma(M-1)}][\rho^{\gamma(M-1)}g_+^{1/2}\rho^{-\gamma M}],$$
$$M = 2k - 2j - 1, \quad \gamma = \alpha_j/M,$$

Theorem 4.2, and the Hölder inequality (4.1). Moreover,

$$|||g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j}|||_{r_{j,1}},$$
$$|||\rho^{-\beta_j}g_+^{j+(1/2)}|||_{r_{j,2}},$$

are dominated by the n independent constants.

Using the Hölder inequality for σ_r – spaces we get

$$|||t|||_r \equiv |||g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j}T_j\rho^{-\beta_j}g_+^{j+1/2}|||_r$$
$$\leqslant |||g^{k-1-j}g_+^{1/2}\rho^{-\alpha_j}|||_{r_{j,2}}\|T_j\| \; |||\rho^{-\beta_j}g_+^{j+1/2}|||_{r_{j,1}},$$
$$\frac{1}{r} = \frac{1}{r_{j,1}} + \frac{1}{r_{j,2}} < \min\left\langle \frac{2k}{l}, \frac{\nu}{l} \right\rangle.$$

Using the inequalities $k > l/2$ and $\nu > l$, we obtain $t_j \in \sigma_1$ and, hence, $g^k - g_+^k \in \sigma_1$ uniformly in $n = 1, 2, \cdots$. ■

Remarks:

(1) If $l \geqslant 3$, we can take $b_0^{(j)} \in L_w^l(\mathbb{R}^l), j = 1,\ldots,l$, with a suitable norm in L_w^l, where $L_w^p(\mathbb{R}^l)$ consists of all measurable functions which obey the estimate

$$\|f\|_{p,w} \equiv \sup_{t>0} t\,[\operatorname{meas}\{x \mid f(x)| > t\}]^{1/p} < \infty.$$

Note that $L^p(\mathbb{R}^l) \subset L_w^l(\mathbb{R}^l)$ (see Ref. 31) and the inequalities (2.1) follows from the Strichartz inequality

$$\|fg\|_p \leqslant \text{const}\|f\|_{l,w}\|\nabla g\|_p, \quad 1 < p < l.$$

There is a discussion of the best constant in this inequality in Refs. 33 and 34.

(2) Suppose that the functions $\mathbf{e}, q$ obey the assumption: $W_\pm(H(\mathbf{e}) + q, H_0)$ exist and are complete and, moreover, $\mathbf{e}^2 \in PK_0(H_0)$, $q \in PK_0(H_0)$. (For example, $\mathbf{e} = 0$, and q obeys the assumption of Theorem 1.1.) Using (FKI) and the representation of the powers of resolvent by the semigroup, we obtain the inequality $|\rho^{-1}(H(\mathbf{e}) + q + \lambda)^{-1/2} f| \leqslant \rho^{-1}(H_0 + q + \lambda)^{-1/2}|f|$. This says that $\rho^{-1}(H(\mathbf{e}) + q + \lambda)^{-1/2} \in \delta_{2k}$ for all integer $k > l/2$ (see Theorem 4.2 and the Remark 1 below Proposition 4.3). Hence we can replace the operator H_0 in Theorem 6.1 by the operator $H(\mathbf{e}) + q$ and to prove thereby the existence and the completeness of the $W_\pm(H(\mathbf{e} + \mathbf{b}) + q + Q, H_0)$.

Note that the proof of Theorem 6.1 is based on the arguments which are very general. So we can extend them on the case of elliptic operators of higher order. We shall confine to a formulation of the theorem:

Theorem 6.2: Let H_0 be an elliptic operator of the order $2m$ with C^∞ coefficients. Suppose that A is an operator associated with the symmetric form $\Sigma_{|\alpha|<m,|\beta|<m}\langle a_{\alpha\beta} D^\alpha u, D^\beta v\rangle$ $|a| + |\beta| < 2m$ and, moreover, $A = \rho^{-\nu} A_0$, $\nu > l$, $A_0 \in PK(H_0)$. Then $W_\pm(H_0 + A, H_0)$ exist and are complete.

Remark: Evidently, Theorem 6.2 is the most general result (without consideration of the behavior of coefficients at the infinity) about the completeness of $W_\pm(H_0 + A, H_0)$ for the elliptic operator H_0 of higher order (see, for example, Refs. 35–38.

APPENDIX: COMMUTATOR ESTIMATES

Let us introduce the following notations:

$$\rho = \rho(x) = (1 + |x|^2)^{1/2}, \quad \Lambda(\mathbf{b}) = \mathbf{b}i\nabla + i\nabla\mathbf{b}, \quad g_+ = (H_0 + \mathbf{b}^2 + \lambda)^{-1}, \quad \lambda > 0.$$

Suppose that the vector function $\mathbf{b}$ and the function Q obey the assumptions of Theorem 6.1 and, moreover, $\mathbf{b} \in C_0^\infty(\mathbb{R}^l)$, ($\mathbf{b} \equiv \mathbf{a}_n$!).

We shall consider in this section the estimates of the norm of the operator

$$\rho^\alpha g_+^{1/2} \Lambda(\mathbf{b}) g_+^{1/2} \rho^\beta, \quad \alpha, \beta \geqslant 0, \ \alpha + \beta = \nu.$$

Put $[A, B] = AB - BA$. It is easy to see that the following equalities are valid (on the Schwartz space $\mathcal{S}$)

$$[\nabla_k, \rho_\alpha] = \alpha x_k \rho^{\alpha-2} \quad \left(\nabla_k \equiv \frac{\partial}{\partial x_k}\right), \quad [\Delta, \rho^\alpha] = 2\alpha\rho^{\alpha-2}\mathbf{x}\nabla + \alpha\rho^{\alpha-2}\left[l + (\alpha - 2)\frac{|x|^2}{+1|x|^2}\right],$$

$$[g_+, \rho^\alpha] = g_+[\Delta, \rho^\alpha]g_+.$$

Let $\eta = i(\beta - \alpha)\rho^{\nu-2}\mathbf{x}\mathbf{b}_0$, then, clearly,

$$\Lambda(\mathbf{b}) = i\nabla\rho^{-\alpha}\mathbf{b}_0\rho^{-\beta} + \rho^{-\alpha}\mathbf{b}_0\rho^{-\beta}i\nabla = \rho^{-\alpha}\Lambda(\mathbf{b}_0)\rho^{-\beta} + \eta$$

and

$$\begin{aligned}\rho^\alpha g_+^{1/2}\Lambda(\mathbf{b})g_+^{1/2}\rho^\beta = {} & [\rho^\alpha, g_+^{1/2}]\rho^{-\alpha}\Lambda(\mathbf{b}_0)\rho^{-\beta}[g_+^{1/2}, \rho^\beta] + [\rho^\alpha, g_+^{1/2}]\rho^{-\alpha}\Lambda(\mathbf{b}_0)g_+^{1/2} + g_+^{1/2}\Lambda(\mathbf{b}_0)\rho^{-\beta}[g_+^{1/2}, \rho^\beta] \\ & + g_+^{1/2}\Lambda(\mathbf{b}_0)g_+^{1/2} + \rho^\alpha g_+^{1/2}\eta g_+^{1/2}\rho^\beta. \end{aligned} \tag{A1}$$

Our assumptions about $\mathbf{b}_0$ imply that

$$\sup_{n>1}\|g_+^{1/2}\Lambda(\mathbf{b}_0)g_+^{1/2}\| < 1 \quad (\text{recall that } \mathbf{b} = \mathbf{a}_n)$$

for all sufficiently large $\lambda > 0$.

We shall now consider all other terms in the equality (A1). We shall prove that if $\lambda \to \infty$, their norms tend to zero uniformly on $\mathbf{b}_0$.

Consider the commutator $[\rho^\alpha, g_+^{1/2}]$. Note, at first, that

$$g_+^{1/2} \equiv (H_+ + \lambda)^{-1/2} = \frac{1}{\pi}\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\,d\xi$$

(see, for example, Ref. 19, Chapter V.3), so that

$$\begin{aligned}[\rho^{\alpha-2}, g_+^{1/2}] &= \frac{1}{\pi}\int_0^\infty \xi^{-1/2}[\rho^\alpha, (H_+ + \lambda + \xi)^{-1}]\,d\xi = \frac{1}{\pi}\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}[\rho^\alpha, \Delta](H_+ + \lambda + \xi)^{-1}\,d\xi \\ &= -\frac{2\alpha}{\pi}\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\rho^{\alpha-2}\mathbf{x}\nabla(H_+ + \lambda + \xi)^{-1}\,d\xi\end{aligned}$$

$$-\frac{\alpha}{\pi}\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\Big(l + (\alpha-2)\frac{|x|^2}{1+|x|^2}\Big)(H_+ + \lambda + \xi)^{-1}\,d\xi.$$

Next we have

$$\begin{aligned}\rho^{\alpha-2}(H_+ + \delta)^{-1} &= (H_+ + \delta)^{-1}\rho^{\alpha-2} + (H_+ + \delta)^{-1}[\rho^{\alpha-2}, \Delta\,](H_+ + \delta)^{-1}\\ &= (H_+ + \delta)^{-1}\rho^{\alpha-2} + C_1'(H_+ + \delta)^{-1}\rho^{\alpha-4}(H_+ + \delta)^{-1} + C_2'(H_+ + \delta)^{-1}\varphi_0\rho^{\alpha-4}(H_+ + \delta)^{-1}\\ &\quad + C_3'(H_+ + \delta)^{-1}\rho^{\alpha-4}\mathbf{x}\nabla(H_+ + \delta)^{-1},\end{aligned} \tag{A2}$$

where $\varphi_0 = (|x|^2(1+|x|^2)^{-1} \in L^\infty$, C_1', C_2', C_3' are constants depending on α and l only.

The last item in (A2) is transposed in the following way (omitting the constant C_3')

$$\begin{aligned}&(H_+ + \delta)^{-1}\nabla\mathbf{x}\,\rho^{\alpha-4}(H_+ + \delta)^{-1} - (H_+ + \delta)^{-1}(\nabla\mathbf{x}\,\rho^{\alpha-4})(H_+ + \delta)^{-1}\\ &= (H_+ + \delta)^{-1}\nabla\mathbf{x}\,\rho^{\alpha-4}(H_+ + \delta)^{-1} - (H_+ + \delta)^{-1}(l + (\alpha-4)\varphi_0)\,\rho^{\alpha-4}(H_+ + \delta)^{-1}.\end{aligned}$$

It is obvious that the representation for $\rho^{\alpha-2}(H_+ + \delta)^{-1}$ as a finite sum consisting of three kinds of terms shown below may be obtained by the above mentioned commutation process.

(1) $C_j'(H_+ + \delta)^{-1}\varphi_{j,1}(H_+ + \delta)^{-1}\varphi_{j,2}\cdots\varphi_{j,m_j}(H_+ + \delta)^{-1}\rho^{\alpha-k_j}$; $k_j \geqslant 0$,

(2) $d_j(H_+ + \delta)^{-1}\varphi_{j,1}(H_+ + \delta)^{-1}\varphi_{j,2}\cdots\varphi_{j,n_j}\rho^{\alpha-t_j}(H_+ + \delta)^{-1}$,

(3) $e_j(H_+ + \delta)^{-1}\theta_{j,1}(H_+ + \delta)^{-1}\theta_{j,2}\cdots\theta_{j,r_j}(H_+ + \delta)^{-1}\mathbf{x}\,\rho^{\alpha-\mathscr{K}_j}\nabla(H_+ + \delta)^{-1}$ $\quad (j = 1, 2, \ldots, N)$,

where all t_j obey the condition $\alpha - t_j \leqslant 0$, all obey the condition $\alpha - \mathscr{K}_j + 1 \leqslant 0$ and the function $\varphi_{j,m}$, $\varphi_{j,n}$, $\varphi_{j,r}$ are elements of $L^\infty(\mathbb{R}^l)$, moreover, their -norms depend on α and l only.

Thus we have

$$\begin{aligned}&\Big\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\rho^{\alpha-2}(H_+ + \lambda + \xi)^{-1}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\Big\|\\ &\leqslant \sum_{j=1}^N \Big(C_j'\Big\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\varphi_{j,1}(H_+ + \lambda + \xi)^{-1}\varphi_{j,2}\cdots\varphi_{j,m_j}(H_+ + \lambda + \xi)^{-1}\rho^{-k_j}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\Big\|\\ &\quad + d_j\Big\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\varphi_{j,1}(H_+ + \lambda + \xi)^{-1}\varphi_{j,2}\cdots\varphi_{j,n_j}\rho^{\alpha-t_j}(H_+ + \lambda + \xi)^{-1}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\Big\|\\ &\quad + e_j\Big\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\theta_{j,1}(H_+ + \lambda + \xi)^{-1}\theta_{j,2}\cdots\theta_{j,r_j}(H_+ + \lambda + \xi)^{-1}\mathbf{x}\,\rho^{\alpha-\mathscr{K}_j}\nabla\\ &\quad \times(H_+ + \lambda + \xi)^{-1}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\Big\|\Big)\\ &= \sum_{j=1}^N (C_j'I_{j,1} + d_jI_{j,2} + e_jI_{j,3}),\end{aligned}$$

$$I_{j,1} \leqslant C_1(\alpha,l,j)\sup_{\xi>0}\|(H_+ + \lambda + \xi)^{-1/2}\rho^{-k_j}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\|\int_0^\infty \xi^{-1/2}(\lambda + \xi)^{-m_j - 1/2}\,d\xi,$$

$$I_{j,2} \leqslant C_2(\alpha,l,j)\sup_{\xi>0}\|(H_+ + \lambda + \xi)^{-1/2}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\|\int_0^\infty \xi^{-1/2}(\lambda + \xi)^{-n_j - 1/2}\,d\xi,$$

$$I_{j,3} \leqslant C_3(\alpha,l,j)\sum_{k=1}^l \sup_{\xi>0}(\|\nabla_k(H_+ + \lambda + \xi)^{-1/2}\| \times \|(H_+ + \lambda + \xi)^{-1/2}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\|)\cdot\int_0^\infty \xi^{-1/2}(\xi + \lambda)^{-r_j - 1}\,d\xi.$$

The following relations were used in the last estimate:

$$\rho^{\alpha-t_j} \leqslant 1,\ x_k\,\rho^{\alpha-\mathscr{K}_j} \leqslant 1, \quad \|\nabla_k(H_+ + \lambda + \xi)^{-1/2}\| \leqslant 1, \quad \varphi_{j,m}, \psi_{j,m}, \theta_{j,m} \in L^\infty(\mathbb{R}^l),$$

$$\sup_{\xi>0}\|(H_+ + \lambda + \xi)^{-1/2}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\| < 1.$$

The latter is the direct consequence of the equality

$$\rho^{-\alpha}\Lambda(\mathbf{b}_0) = \Lambda(\rho^{-\alpha}\mathbf{b}_0) + i\alpha\,\rho^{\alpha-2}\mathbf{x}\mathbf{b}_0$$

and the following estimates

$$\|g_+^{1/2}\Lambda(\mathbf{b}_0)g_+^{1/2}\| < 1, \quad \|\rho^{-\alpha-2}\mathbf{b}_0 g_+^{1/2}\| < 1, \quad \sup_{\xi>0}\|(H_+ + \lambda + \xi)^{-1/2}(H_+ + \lambda)^{1/2}\| \leqslant 1.$$

Note that each estimate for the $I_{j,1}$, $I_{j,2}$, $I_{j,3}$ contain the integral $\gamma(\lambda) = \int_0^\infty \xi^{-1/2}(\lambda + \xi)^{-k-1}\,d\xi$ — type, $k > 0$. Since $\gamma(\lambda) \to 0$, $(\lambda \to \infty)$, we have

$$\Big\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\rho^{\alpha-2}(H_+ + \lambda + \xi)^{-1}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\Big\| = \mathscr{K}_1(\lambda),$$

where $\mathscr{X}_1(\lambda)\to 0(\lambda\to\infty)$. By analogy we have

$$\|\int_0^\infty \xi^{-1/2}(H_+ + \lambda + \xi)^{-1}\rho^{\alpha-2}\mathbf{x}\nabla(H_+ + \lambda + \xi)^{-1}\rho^{-\alpha}\Lambda(\mathbf{b}_0)(H_+ + \lambda)^{-1/2}\,d\xi\|\to 0 \quad (\lambda\to\infty).$$

Thus we have

$$\|[\rho^\alpha, g_+^{1/2}]\rho^{-\alpha}\Lambda(\mathbf{b}_0)g_+^{1/2}\| \leqslant \mathscr{X}(\lambda)$$

for any $\lambda > 0$, where $\lim_{\lambda\to\infty}\mathscr{X}(\lambda) = 0$.

The following can be proved similarly

$$\|g_+^{1/2}\Lambda(\mathbf{b}_0)\rho^{-\beta}[g_+^{1/2}, \rho^\beta]\|\to 0, \quad \|[\rho^\alpha, g_+^{1/2}]\rho^{-\alpha}\Lambda(\mathbf{b}_0)\rho^{-\beta}[g_+^{1/2}, \rho^\beta]\|\to 0 \quad (\lambda\to\infty).$$

So we need only to consider the term

$$\|\rho^\alpha g_+^{1/2}\eta g_+^{1/2}\rho^\beta\| = \|\rho^\alpha g_+^{1/2}\rho^{-\alpha}(\beta-\alpha)\rho^{-2}\mathbf{x}\mathbf{b}_0\rho^{-\beta}g_+^{1/2}\rho^\beta\|.$$

As was proved in Ref. 22,

$$(\rho^{\alpha-2}(H_+ + \lambda)^{-1/2}\rho^{-\alpha})(x,y) \leqslant (H_0 + \lambda_\alpha)^{-1/2}(x,y),$$

where $\lambda_\alpha\to\infty$ if $\lambda\to\infty$, so that using our assumptions about $\mathbf{b}_0$ and the boundess of the $x_k\,\rho^{-2}(k = 1,2,\ldots,l)$ we obtain:

$$\|\rho^\alpha g_+^{1/2}\eta g_+^{1/2}\rho^\beta\| \leqslant \|(H_0 + \lambda_\alpha)^{-1/2}(\beta-\alpha)|\mathbf{x}\mathbf{b}_0|\,\rho^{-2}(H_0 + \lambda_\beta{}^{-1/2}\|\to 0 \quad (\lambda\to\infty).$$

Thus we have

Lemma A.1: Suppose that $\mathbf{b}$ obey all the assumptions of Theorem 6.1 and, moreover, $\mathbf{b}\in C_0^\infty(\mathbb{R}^l)$. Let $\alpha, \beta \geqslant 0, \alpha + \beta = \nu$. Then

$$\|\rho^\alpha g_+^{1/2}\Lambda(\mathbf{b})g_+^{1/2}\rho^\beta\| < 1$$

uniformly in $\mathbf{b}_0$ for all sufficiently large $\lambda > 0$.

[1]J. Cook, J. Math. Phys. Cambridge, Mass. **36**, 82 (1957).
[2]M. Schechter, Bull. Amer. Math. Soc. **83**, 381 (1977).
[3]M. Schechter, Duke Math. J. **44**, 863 (1977).
[4]B. Simon, Commun. Math. Phys. **53**, 151 (1977).
[5]J. Kupsh and W. Sandas, Commun. Math. Phys. **2**, 147 (1966).
[6]U. Greifenegger, K. Jorgens, J. Weismann, and M. Winkler, "Streutheorie fur Schrödinger-Operatoren," preprint.
[7]S. Agmon, Ann. Scuola Norm. Super. Pisa. Classe Sci. Ser. 4 **2**, 151 (1975).
[8]M. S. Birman, Dokl. Akad. Nauk SSSR, **147**, 1008 (1962).
[9]M. Combescure and J. Ginibre, J. Funct. Anal. **29**, 54 (1978).
[10]P. Deift and B. Simon, J. Funct. Anal. **23**, 218 (1976).
[11]Yu. A. Semenov, Lett. Math. Phys. **1**, 457 (1977).
[12]D. B. Pearson, Helv. Phys. Acta **48**, 639 (1975).
[13]D. B. Pearson, Commun, Math. Phys. **40**, 125 (1975).
[14]T. Ikebe and T. Tayoshi, Publ. Res. Inst. Math. Sci. Kyoto Univ. **4**, 483 (1968).
[15]E. B. Davies and B. Simon, Commun. Math. Phys. **63**, 277 (1978).
[16]J. Avron, I. Herbst, and B. Simon, Duke Math. J. **45**, 847 (1978).
[17]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, New Jersey, 1971)
[18]M. S. Birman and M. Z. Solomjak, "On estimates of singular values of integral operators," III, Vesnik Leningrad University, No. 1 (1969). pp. 35–48.
[19]T. Kato, *Perturbation Theorey for Linear Operators* (Springer, Berlin, 1966).
[20]W. Faris, *Self-Adjoint Operators, Lect. Notes in Mathematics, Vol. 433* (Springer, Berlin, 1975).
[21]I. Herbst and J. Sloan, Trans. Am. Math. Soc. **236**, 325 (1978).
[22]V. F. Kovalenko and Yu. A. Semenov, Usp. Mat. Nauk **33**, 107 (1978).
[23]M. Schechter, Lett. Math. Phys. **1**, 265 (1976).
[24]B. Simon, Proc. R. Soc. Edinburgh **29**, 267 (1977).
[25]V. A. Liskevich and Yu. A. Semenov, "An approximation theorem for the Schrödinger operators" (in preparation).
[26]B. Simon, Indian Univ. J. Math. **26**, 1067 (1977).
[27]T. Kato, Integr. Equat. and Oper. Theory, 103 (1978).
[28]B. Simon, J. Oper. Theory **1**, 37 (1979).
[29]I. C. Gohberg and M. G. Krein, "Introduction to the Theory of Linear Nonself-Adjoint Operators," AMS, 1969 (original Russian edition, 1965).
[30]B. Simon, *Lectures on Trace Ideals Methods*, London Math. Soc. Lecture Note (Cambridge U. P., Cambridge, England, 1979).
[31]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. II, Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[32]K. Ito and H. P. McKean, Jr., *Diffusion Processes and Their Sample Paths*, 2nd ed. (Springer, Berlin, 1974).
[33]W. Faris, Duke Math. J. **43**, 365 (1976).
[34]V. F. Kovalenko, M. A. Perelmuter, and Yu. A. Semenov, "Schrödinger operators with $L_w^{l/2}(\mathbb{R}^l)$-potentials (to appear).
[35]M. S. Birman, Funk. Anal. Pril. **3**, tr 1 (1969).
[36]L. Hormander, Math. Z. **146**, 69 (1976).
[37]M. Schechter, Comments Math. Helv. **49**, 84 (1974).
[38]M. Schechter, J. Math. Sco. Jpn. **28**, 71 (1976).

Self-adjointness and spectrum of Hamiltonians in nonrelativistic quantum electrodynamics

Asao Arai[a)]
Department of Pure and Applied Sciences, College of General Education, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

(Received 5 February 1980; accepted for publication 5 August 1980)

The nonrelativistic quantum electrodynamics is formulated in a mathematically rigorous way. The self-adjointness and the basic spectral property of the Hamiltonians are proved.

PACS numbers: 03.70. + k, 03.65.Db

1. INTRODUCTION

We consider a system of one nonrelativistic electron in an external potential such as the Coulomb potential interacting with a transverse radiation field. Our ultimate aim is to give a rigorous mathematical foundation, though within a nonrelativistic context, of the physical theory[1-3] of the Lamb shift and spontaneous emission of photons, which have so far been formulated only in terms of the formal perturbation theory.

As a first step, this paper gives a fundamental framework by defining the Hamiltonians of the system and proving their self-adjointness and their basic spectral property. We encounter a problem of perturbation of eigenvalues embedded in a continuous spectrum, and it is conjectured that the Lamb shift and the spontaneous emission of photons should be understood in terms of "resonances."

The outline of the paper is as follows: In Sec. 2 we give some notations and definitions. In defining the interaction of the electron with the quantized radiation field, we use the dipole approximation for simplicity and introduce an ultraviolet cutoff in order to make the Hamiltonians well defined. In Sec. 3 we prove the self-adjointness of the Hamiltonians by using the technique of approximate dressing transformation, which is usually used to control infinite momentum limit.[4-6] In Sec. 4 we analyze the spectrum of the Hamiltonians using the method of asymptotic fields[7,8] and the analytic perturbation theory. In Sec. 5 we give some remarks on the subject.

2. SOME NOTATIONS AND DEFINITIONS

The Hibert space $\mathscr{F}_{\rm EM}$ for the radiation field in the Coulomb gauge is defined as the 2-fold tensor product of the usual Fock space $\mathscr{F}$ for the neutral scalar field[4]:

$$\mathscr{F}_{\rm EM} = \mathscr{F} \otimes \mathscr{F}. \tag{2.1}$$

The tensor product of this space $\mathscr{F}_{\rm EM}$ with the space of electron wave functions $L^2(R^3)$ gives the Hilbert space $\mathscr{H}$ for our model:

$$\mathscr{H} = L^2(R^3) \otimes \mathscr{F}_{\rm EM}. \tag{2.2}$$

The Hamiltonian of the atom is given by

$$H_0^{(1)}(V) = -(1/2m)\Delta + V, \tag{2.3}$$

with m the electron bare mass and V the external potential; we assume that V is real and

$$V \in L^\infty(R^3) + L^2(R^3). \tag{2.4}$$

In terms of $a(f)$ [resp. $a^*(f)$], $f \in L^2(R^3)$, the usual annihilation (resp. creation) operator for the neutral scalar field,[4] the operatores $a^{\#(j)}(f)$, $f \in L^2(R^3)$, $j = 1, 2$, for the radiation field are defined in $\mathscr{F}_{\rm EM}$ as

$$a^{\#(1)}(f) = a^{\#}(f) \otimes I, \quad a^{\#(2)}(f) = I \otimes a^{\#}(f), \tag{2.5}$$

where $a^{\#}(f)$ denotes either $a(f)$ or $a^*(f)$. They satisfy the following commutation relations on the set of finite particle vectors in $\mathscr{F}_{\rm EM}$:

$$[a^{(j)}(f), a^{*(l)}(g)] = \delta_{jl} \int d\mathbf{k}\ f(\mathbf{k})g(\mathbf{k}),$$
$$\text{others} = 0, \quad j,l = 1,2, \quad f,g \in L^2(R^3). \tag{2.6}$$

The free Hamiltonian $H_0^{\rm EM}$ for the radiation field is defined as

$$H_0^{\rm EM} = \sum_{j=1}^{2} \int d\mathbf{k}\ |\mathbf{k}| a^{*(j)}(\mathbf{k}) a^{(j)}(\mathbf{k}), \tag{2.7}$$

$H_0^{\rm EM}$ is self-adjoint and non-negative on the maximal domain.

All the operatores in $L^2(R^3)$ (or in $\mathscr{F}_{\rm EM}$) have natural extensions to $\mathscr{H}$: If A is a densely defined closable operator in $L^2(R^3)$ (or in $\mathscr{F}_{\rm EM}$), then $A \otimes I$ (resp., $I \otimes A$) is a densely defined closed operator in $\mathscr{H}$. We write $A \otimes I$ (resp., $I \otimes A$) simply as A.

For a.e. $\mathbf{k} \in R^3$ we take $\mathbf{e}^{(j)} \in R^3$, $j = 1,2$, which represent the polarizations of the radiation, satisfying the following conditions:

$$\mathbf{e}^{(j)}(\mathbf{k}) \cdot \mathbf{e}^{(l)}(\mathbf{k}) = \delta_{jl}, \quad j,l = 1,2, \quad \mathbf{e}^{(1)}(\mathbf{k}) \times \mathbf{e}^{(2)}(\mathbf{k}) = \mathbf{k}/|\mathbf{k}|, \quad \mathbf{e}^{(1)}(-\mathbf{k}) = \mathbf{e}^{(1)}(\mathbf{k}), \quad \mathbf{e}^{(2)}(-\mathbf{k}) = -\mathbf{e}^{(2)}(\mathbf{k}).$$

Let v be a rotation invariant, real function on R^3 satisfying

$$\|v/\sqrt{|\mathbf{k}|}\|_2 < \infty, \quad \|v/|\mathbf{k}|\|_2 < \infty, \tag{2.8}$$

and then

$$\mathbf{A}_v(\mathbf{x}) = \sum_{j=1}^{2} (1/\sqrt{(2(2\pi)^3)}) \{a^{*j}(e^{-i\mathbf{k}\mathbf{x}} v \mathbf{e}^{(j)}/\sqrt{|\mathbf{k}|}) + a^{(j)}(\mathbf{e}^{i\mathbf{k}\mathbf{x}} v \mathbf{e}^{(j)}/\sqrt{|\mathbf{k}|})\} \tag{2.9}$$

is the radiation field with an ultraviolet cutoff v.

The minimal interaction between the atom and the radiation field is given in the dipole approximation by the sum of

$$H_I^{(1)}(v) = (\lambda/m)(-i\nabla) \cdot \mathbf{A}_v(\mathbf{o}) \tag{2.10}$$

and the diamagnetic term

$$H_I^{(2)}(v) = (\lambda^2/2m)\mathbf{A}_v^2(\mathbf{o}), \tag{2.11}$$

[a)]Present address: Department of Mathematics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan.

 022-2488/81/030534-04$1.00

where $\lambda \in R$ denotes the electron charge. The mass renormalization counter term is given by

$$W(v) = -\lambda^2 R(v)\Delta \tag{2.12}$$

to the second order in the electron charge with

$$R(v) = (1/24\pi^3 m^2)\|v/|\mathbf{k}|\|_2^2. \tag{2.13}$$

The renormalized Hamiltonian $H_v^{\mathrm{ren}}(V)$ for our model is given formally by

$$H_v^{\mathrm{ren}}(V) = H_0^{\mathrm{el}}(V) + H_0^{\mathrm{EM}} + H_I^{(1)}(v) + H_I^{(2)}(v) + W(v). \tag{2.14}$$

3. SELF-ADJOINTNESS OF THE HAMILTONIANS

Let

$$H_0 = -(1/2m)\Delta + H_0^{\mathrm{EM}}. \tag{3.1}$$

H_0 is self-adjoint and non-negative with $D(H_0) = D(\Delta) \cap D(H_0^{\mathrm{EM}})$. Our aim in this section is to prove the following.

Theorem 3.1: The renormalized Hamiltonian $H_v^{\mathrm{ren}}(V)$ given in (2.14) is self-adjoint with $D(H_v^{\mathrm{ren}}(V)) = D(H_0)$. Further

$$H_v^{\mathrm{ren}}(V) \geqslant \inf \Sigma(H_0^{\mathrm{el}}(V)), \tag{3.2}$$

where $\Sigma(H_0^{\mathrm{el}}(V))$ denotes the spectrum of $H_0^{\mathrm{el}}(V)$ in (2.3).

By direct computations using basic estimates[5] for the operators $a^{\#}(\cdot)$, it can be proved that the interactions $H_I^{(j)}(v)$, $j = 1,2$, are H_0-bounded:

$$\|H_I^{(j)}(v)\Psi\| \leqslant a\|H_0\Psi\| + b\|\Psi\|, \Psi \in D(H_0), \quad j = 1,2, \tag{3.3}$$

where $a > 0$ and $b > 0$ are constants. However, the relative bound a of $H_I^{(j)}(v)$ with respect to H_0 is not smaller than one unless $|\lambda|$ is sufficiently small. Therefore, for large values of $|\lambda|$, we cannot directly apply the Kato–Rellich theorem to prove the self-adjointness of $H_0 + H_I^{(1)}(v) + H_I^{(2)}(v)$. The main point of our proof is to use the dressing transformation.

We begin with the definition of the generator of the dressing transformation. Let $K > 0$ and let

$$T_K = -\frac{\lambda}{(2(2\pi)^3)^{1/2} m} \sum_{j=1}^{2} (-i\nabla)\cdot\{a^{(j)}(|\mathbf{k}|^{-3/2} v_K \mathbf{e}^{(j)}) - a^{*(j)}(|\mathbf{k}|^{-3/2} v_K \mathbf{e}^{(j)})\}, \tag{3.4}$$

where

$$v_K = \begin{cases} v(\mathbf{k}) & K \leqslant |\mathbf{k}|, \\ 0 & 0 \leqslant |\mathbf{k}| < K. \end{cases} \tag{3.5}$$

Let

$$D_F = S \otimes F \otimes F, \tag{3.6}$$

where $S = \{f \in \mathscr{S}(R^3) |$ Fourier transform of f is in $C_0^\infty(R^3)\}$, and F is the set of finite particle vectors in $\mathscr{F}$. In the same way as in Ref. 6 we can prove the following.

Lemma 3.1: D_F is a set of analytic vectors for T_K and T_K is essentially skew-adjoint on D_F.

We write the closure of $T_K \upharpoonright D_F$ simply as T_K. Let

$$H_v'^{\mathrm{ren}}(V) = H_0^{\mathrm{el}}(V) + H_0^{\mathrm{EM}} + H_I^{(1)}(v) + W(v). \tag{3.7}$$

This operator is the renormalized Hamiltonian without the diamagnetic term. We denote $H_v^{\mathrm{ren}}(0)$ [resp., $H_v'^{\mathrm{ren}}(0)$] by H_v^{ren} (resp., $H_v'^{\mathrm{ren}}$). By using Lemma 3.1, the H_0-boundedness of $H_v'^{\mathrm{ren}}$ and some limiting argumens, we can prove the following.

Lemma 3.2: (1) e^{T_K} maps $D(H_0)$ onto $D(H_0)$. (2) For all $\sigma \geqslant 0$ and $K \geqslant \sigma$,

$$e^{-T_K} H_{v_\sigma \chi_K}'^{\mathrm{ren}} e^{T_K} = H_{v_\sigma}'^{\mathrm{ren}} \quad \text{on } D(H_0), \tag{3.8}$$

where

$$\chi_K(\mathbf{k}) = \begin{cases} 1, & 0 \leqslant |\mathbf{k}| \leqslant K \\ 0, & K < |\mathbf{k}|. \end{cases} \tag{3.9}$$

This lemma can be proved in the same way as in Ref. 4.

Lemma 3.3: $H_v'^{\mathrm{ren}}(V)$ is self-adjoint with $D(H_v'^{\mathrm{ren}}(V)) = D(H_0)$ and

$$H_v'^{\mathrm{ren}}(V) \geqslant \inf \Sigma(H_0^{\mathrm{el}}(V)). \tag{3.10}$$

Proof: It is easy to see that

$$\|W(v)\Psi\| \leqslant 2m\lambda^2 R(v)\|H_0\Psi\|, \quad \Psi \in D(H_0).$$

This inequality together with (3.3) gives

$$\|(H_I^{(1)}(v\chi_K) + W(v\chi_K))\Psi\| \leqslant c_K(\|H_0\Psi\| + \|\Psi\|), \quad \Psi \in D(H_0), \tag{3.11}$$

where $c_K > 0$ is a constant such that $c_K \to 0$ as $K \to +0$. We fix the value of K such that $c_K < 1$. Then it follows from (3.11) and the Kato–Rellich theorem that $H_{v\chi_K}'^{\mathrm{ren}}$ is self-adjoint with $D(H_{v\chi_K}'^{\mathrm{ren}}) = D(H_0)$. Therefore, by Lemma 3.2, $H_v'^{\mathrm{ren}}$ is self-adjoint with $D(H_v'^{\mathrm{ren}}) = D(H_0)$. Furthermore, using (3.8), we can show that

$$H_v'^{\mathrm{ren}} \geqslant -(1/2m)\Delta. \tag{3.12}$$

We next consider $H_v'^{\mathrm{ren}}(V) = H_v'^{\mathrm{ren}} + V$. It follows from the closed graph theorem that there exists a constant $c > 0$ such that

$$\|H_0\Psi\| \leqslant c(\|H_v'^{\mathrm{ren}}\Psi\| + \|\Psi\|), \quad \Psi \in D(H_0). \tag{3.13}$$

On the other hand, we have the following well-known estimate from the assumption (2.4) on V (see, e.g., Ref. 9, p. 303):

$$\|V\Psi\| \leqslant \epsilon\|-(1/2m)\Delta\Psi\| + b(\epsilon)\|\Psi\|, \quad \Psi \in D(\Delta), \tag{3.14}$$

where $\epsilon > 0$ is arbitrary and $b(\epsilon)$ is a constant which depends on ϵ. Combining (3.13) and (3.14) we get

$$\|V\Psi\| < c\epsilon\|H_v'^{\mathrm{ren}}\Psi\| + (b(\epsilon) + c\epsilon)\|\Psi\|, \quad \Psi \in D(H_0).$$

Since $\epsilon > 0$ is arbitrary, it follows from the Kato–Rellich theorem that $H_v'^{\mathrm{ren}}(V)$ is self-adjoint and bounded below with $D(H_v'^{\mathrm{ren}}(V)) = D(H_0)$. Inequality (3.10) follows from (3.12).

Proof of Theorem 3.1: Let $L = H_0 + 1$. It follows from (3.3) that

$$\|H_v^{\mathrm{ren}}\Psi\| \leqslant c\|L\Psi\|, \quad \Psi \in D(H_0).$$

By using commutation relations and basic estimates on $a^{\#(j)}$s, we can show that for some constant $c' > 0$

$$|(L\Phi, H_v^{\mathrm{ren}}\Psi) - (H_v^{\mathrm{ren}}\Phi, L\Psi)| \leqslant c'\|L^{1/2}\Psi\| \quad \|L^{1/2}\Phi\|, \quad \Phi, \Psi \in D(H_0).$$

Therefore it follows from the Nelson commutator theorem (see, e.g., Ref. 10, Sec. X) that H_v^{ren} is essentially self-adjoint on $D(H_0)$. On the other hand, by direct computations using the positivity of $H_v'^{\mathrm{ren}}$ [cf. (3.12)] and commutation relations, we can show that for some constant $a > 0$

$$\|H_v'^{\rm ren}\Psi\|^2 + \|H_I^{(2)}(v)\Psi\|^2 \leqslant a(\|H_v^{\rm ren}\Psi\|^2 + \|\Psi\|^2), \quad \Psi\in D(H_0). \tag{3.15}$$

Since $D(H_v'^{\rm ren})\cap D(H_I^{(2)}(v)) = D(H_0)$ and $H_v'^{\rm ren}$ is closed on $D(H_0)$ by Lemma 3.3, it follows from (3.15) that $H_v^{\rm ren}$ is closed on $D(H_0)$. Thus $H_v^{\rm ren}$ is self-adjoint with $D(H_v^{\rm ren}) = D(H_0)$. The proof of the self-adjointness of $H_v^{\rm ren}(V) = H_v^{\rm ren} + V$ is similar to that of $H_v'^{\rm ren}(V)$. Since $H_I^{(2)}(v)$ is non-negative, inequality (3.2) follows from (3.10).

Remark: We can also consider the operator

$$H_v(V) = H_0^{\rm el}(V) + H_0^{\rm EM} + H_I^{(1)}(v) + H_I^{(2)}(v)$$

as well as

$$H_v'(V) = H_0^{\rm el}(V) + H_0^{\rm EM} + H_I^{(1)}(v)$$

as a Hamiltonian. Let

$$\lambda_c = (2mR(v))^{-1/2}.$$

Suppose that $|\lambda| \neq \lambda_c$. Then, by using Lemma 3.2, we can prove the following[11]:

(A) (1) $H_v'(V)$ is self-adjoint with $D(H_v'(V)) = D(H_0)$; (2) $H_v'(V)$ is bounded below if and only if $|\lambda| < \lambda_c$.

(B) $H_v(V)$ is self-adjoint and bounded below with $D(H_v(V)) = D(H_0)$.

It should be noted that the boundedness below of $H_v'(V)$ breaks down for $|\lambda| > \lambda_c$. In the case $|\lambda| = \lambda_c$ we can prove that if $V\in L^\infty(R^3)$, then $H_v(V)$ [resp., $H_v'(V)$] is essentially self-adjoint, bounded below on $D(H_0)$ and $D(\overline{H_v(V)})$ [resp., $D(\overline{H_v'(V)})$] $\supset D(H_0)$. In particular we

4. SPECTRUM OF THE HAMILTONIANS

In this section we study the spectrum of the Hamiltonians. We first prove the existence of asymptotic fields to obtain some information on the spectrum, and then analyze the spectrum by using the analytic perturbation theory, where we shall introduce an infrared cutoff in the interaction to make the problem manageable. We shall consider only the renormalized Hamiltonian $H_v^{\rm ren}(V)$. The other three Hamiltonians $H_v'^{\rm ren}(V)$, $H_v(V)$, and $H_v'(V)$ can be treated similarly.

Let

$$N = \{f \mid \|f/|\mathbf{k}|^{1/2}\|_2 < \infty\}. \tag{4.1}$$

For f in $L^2(R^3)\cap N$ and $t\in R$, we define $a_t^{\#(j)}(f)$, $j = 1,2$, by

$$a_t^{\#(j)}(f) = e^{-itH_v^{\rm ren}(V)}e^{itH_0(V)}a^{\#(j)}(f)e^{-itH_0(V)}e^{itH_v^{\rm ren}(V)}, \tag{4.2}$$

where

$$H_0(V) = H_0^{\rm el}(V) + H_0^{\rm EM}. \tag{4.3}$$

$a_t^{(j)}(f)$, $j = 1,2$ are well defined on $D(H_0)$. By the same method as in Refs. 7 and 8, we can prove the following.

Lemma 4.1: Suppose $v\in C^2(R^3)$ in addition to (2.8). Let Ψ be in $D(H_0)$ and f be in $L^2(R^3)\cap N$. Then the strong limits

s- $\lim_{t\to\pm\infty} a_t^{\#(j)}(f)\Psi \equiv a_\pm^{\#(j)}(f)\Psi$ exist and $\{H_v^{\rm ren}(V),\ a_\pm^{\#(j)}(f)\}$ satisfies the same commutation relations as that of $\{H_0^{\rm EM},\ a^{\#(j)}(f)\}$. Furthermore, if Ψ is an eigenvector of $H_v^{\rm ren}(V)$, then $a_\pm^{(j)}(f)\Psi = 0$, $j = 1,2$.

Let

$$E(v) = \inf\Sigma(H_v^{\rm ren}(V)). \tag{4.4}$$

Then, by Theorem 3.1, we have $E(v)\geqslant \inf\Sigma(H_0^{\rm el}(V))$.

Proposition 4.1: Let v be as in Lemma 4.1. If $E(v)$ is an eigenvalue of $H_v^{\rm ren}(V)$, then

$$\Sigma(H_v^{\rm ren}(V)) = \Sigma_{\rm ess}(H_v^{\rm ren}(V)) = E(v),\infty), \tag{4.5}$$

where $\Sigma_{\rm ess}(H_v^{\rm ren}(V))$ denotes the essential spectrum of $H_v^{\rm ren}(V)$.

Proof: Let Ψ be an eigenvector corresponding to the eigenvalue $E(v)$. Let $H = H_v^{\rm ren}(V) - E(v)$. For any $\mu > 0$ it follows from Lemma 4.1 that

$$(H-\mu)a_+^{*(j)}(f)\Psi = a_+^{*(j)}((|\mathbf{k}|-\mu)f)\Psi, \tag{4.6}$$

where f and $|\mathbf{k}|f$ are in $L^2(R^3)\cap N$. It is easy to find a sequence $\{f_n\}$ such that f_n, $|\mathbf{k}|f_n \in L^2(R^3)\cap N$, $\|f_n\|_2 = 1$ and $\|a_+^{*(j)}((|\mathbf{k}|-\mu)f_n)\Psi\| \to 0$ as $n\to\infty$. Put $\Psi_n = a_+^{*(j)}(f_n)\Psi$. Then, it follows from (4.6) that $\|(H-\mu)\Psi_n\| \to 0$ as $n\to\infty$. On the other hand, we have $\|\Psi_n\| = \|f_n\|_2 = 1$. Therefore μ is in $\Sigma(H)$. Since μ is arbitrary, $[0,\infty)\subseteq\Sigma(H)$. On the other hand, we have $\Sigma(H)\subseteq[0,\infty)$ by definition of H. Therefore we get $\Sigma(H) = \Sigma_{\rm ess}(H) = [0,\infty)$. This means (4.5).

By a standard theorem on the spectrum of tensor product of self-adjoint operators (e.g., Ref. 12, Sec. VIII) we have $\Sigma(H_0(V)) = (\inf\Sigma(H_0^{\rm el}(V)),\infty)$, so that $H_0(V)$ has no discrete spectrum and all the eigenvalues of $H_0(V)$ are embedded in the continuous spectrum. Thus, we have here a problem of the perturbation of eigenvalues embeded in the continuous spectrum. In the present case we can avoid in part this difficulty by introducing an infrared cutoff in the interaction, and then we can use the analytic perturbation theory. The idea of our procedure in the following is due to Ref. 5.

Let $\sigma > 0$ and let

$$K_\sigma = \{\mathbf{k}\in R^3 \mid |\mathbf{k}| \geqslant \sigma\}.$$

Then, we have

$$\mathcal{F} = \mathcal{F}(K_\sigma)\otimes\mathcal{F}(K_\sigma^c),$$

where

$$\mathcal{F}(K_\sigma) = \mathbf{C}\oplus\left[\bigoplus_{n=1}^{\infty}\left(\bigotimes_s^n L^2(K_\sigma)\right)\right],$$

$$\mathcal{F}(K_\sigma^c) = \mathbf{C}\oplus\left[\bigoplus_{n=1}^{\infty}\left(\bigotimes_s^n L^2(K_\sigma^c)\right)\right].$$

Let

$$\mathcal{H}_\sigma = L^2(R^3)\otimes[\mathcal{F}(K_\sigma)\otimes\{\Omega\}]\otimes[\mathcal{F}(K_\sigma)\otimes\{\Omega\}], \tag{4.7}$$

where

$$\Omega = \{1,0,0,\cdots\}\in\mathcal{F}(K_\sigma^c).$$

One can easily prove the following.

Lemma 4.2: (1) $\mathcal{H}_\sigma$ and $\mathcal{H}_\sigma^\perp$ reduce $H_{v_\sigma}^{\rm ren}(V)$.

(2) $E(v_\sigma) = \inf\Sigma(H_{v_\sigma}^{\rm ren}(V)\restriction\mathcal{H}_\sigma)$.

Theorem 4.1: Let $\sigma > 0$. Suppose that $H_0^{\rm el}(V)$ has the discrete spectrum $E_0 < E_1 < \cdots$, at the bottom of the spectrum with the multiplicity m_j for each E_j. Then, for each E_j such that $E_j < E_0 + \sigma$ there exists a constant $r_j(v,\sigma)$ such that if $|\lambda| < r_j(v,\sigma)$, then $H_{v_\sigma}^{\rm ren}(V)$ has m_j eigenvalues (not necessar-

ily distinct) near E_j which are analytic in λ, where $r_j(v,\sigma)$ depends on v and σ. In particular, if $|\lambda| < r_0(v,\sigma)$, then $E(v_\sigma)$ is an eigenvalue of $H_{v_\sigma}^{\rm ren}(V)$, i.e., the ground state for $H_{v_\sigma}^{\rm ren}(V)$ exists.

Proof: Since $\Sigma(H_0^{\rm EM} \upharpoonright \mathscr{H}_\sigma) = \{0\} \cup [\sigma, \infty)$, we see that the discrete spectrum of $H_\sigma(V) \upharpoonright \mathscr{H}_\sigma$ is $\{E_j | E_j < E_0 + \sigma\} (\neq \phi)$. Since $H_I^{(j)}(v)$, $j = 1,2$, and $W(v)$ are H_0-bounded and H_0 is $H_0(V)$-bounded, $H_I(v) \equiv H_I^{(1)}(v) + H_I^{(2)}(v) + W(v)$ is $H_0(v)$-bounded. Therefore $H_v^{\rm ren}(V)$ is an analytic family of type (A) for sufficiently small $|\lambda|$ (see Ref. 9, p. 377, Ref. 13, p. 16). Thus we can apply the analytic perturbation theory to $H_{v_\sigma}^{\rm ren}(V) \upharpoonright \mathscr{H}_\sigma$, taking $H_0(V) \upharpoonright \mathscr{H}_\sigma$ as the unperturbed operator and $H_I(v_\sigma) \upharpoonright \mathscr{H}_\sigma$ as the perturbation, thereby establishing the first half of the theorem. In particular, $\inf\Sigma(H_{v_\sigma}^{\rm ren}(V) \upharpoonright \mathscr{H}_\sigma)$ is an eigenvalue of $H_{v_\sigma}^{\rm ren}(V) \upharpoonright \mathscr{H}_\sigma$ if $|\lambda| < r_0(v,\sigma)$. By Lemma 4.2, (2), it follows that $E(v_\sigma)$ is an eigenvalue of $H_v^{\rm ren}(V)$ if $|\lambda| < r_0(v,\sigma)$ and thus the latter half of the theorem is proved.

By combining Theorem 4.1 with Proposition 4.1 we obtain the following.

Theorem 4.2: Let $\sigma > 0$. Suppose that $\inf\Sigma(H_0^{\rm el}(V))$ is a discrete eigenvalue of $H_0^{\rm el}(V)$. Then, there exists a constant $r(v,\sigma) > 0$ such that for all $|\lambda| < r(v,\sigma)$, $E(v_\sigma)$ is an eigenvalue of $H_v^{\rm ren}(V)$ and

$$\Sigma(H_{v_\sigma}^{\rm ren}(V)) = \Sigma_{\rm ess}(H_{v_\sigma}^{\rm ren}(V)) = E(v_\sigma), \infty). \tag{4.8}$$

Proof: We first suppose that $v \in C^2(R^3)$ and supp $v \subset K_\sigma$ in addition to (2.8) and prove (4.8) by using Theorem 4.1 and Proposition 4.1. Then, by a limiting argument, we prove (4.8) for general v.

Remark: (4.8) shows that all the eigenvalues of $H_v^{\rm ren}(V)$ with small $|\lambda|$ are embedded in the continuous spectrum.

5. CONCLUDING REMARKS

We have proved the self-adjointness of the Hamiltonians (Theorem 3.1 and Remark at the end of Sec. 3) and the basic property of the infrared cutoff Hamiltonian with small coupling constant (Theorem 4.2). But our results are not so strong and should be considered as a first step towards the construction of a more complete theory for our model including the rigorous formulation of the Lamb shift and the spontaneous emission of photons. As was mentioned in Sec. 4, we are faced with the problem of the perturbation of eigenvalues embedded in the continous spectrum. The perturbation theory of eigenvalues embedded in the continuous spectrum has been developed by some authors for some classes of operators (see, e.g., Ref. 14 and references cited therein) and it has been known that it leads us to the so-called "resonances" (see, e.g., Ref. 13, Sec. XII). From this point of view, it may be conjectured that the Lamb shift and the spontaneous emission of photons correspond to the "resonance". However, we have not yet succeeded in proving this conjecture for our model and this problem is left for future studies.

ACKNOWLEDGMENTS

The author is grateful to Professor H. Ezawa, who suggested the problem, and to Professor S. T. Kuroda for helpful discussions and suggestions.

[1] H. A. Bethe, Phys. Rev. **72**, 339–41 (1947).
[2] T. A. Welton, Phys. Rev. **74**, 1157–67 (1948).
[3] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976), 2nd ed.
[4] J. J. Cannon, J. Func. Anal. **8**, 101–52 (1971).
[5] J. Fröhlich, Fortschr. Phys. **22**, 159–98 (1974).
[6] E. Nelson, J. Math. Phys. **5**, 1190–97 (1964).
[7] R. Hoegh-Krohn, Commun. Math. Phys. **21**, 256–60 (1971).
[8] Y. Kato and N. Mugibayashi, Prog. Theor. Phys. **30**, 103–33 (1963).
[9] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.
[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[11] A. Arai, "Some Rigorous Results of Nonrelativistic Quantum Electrodynamics," preprint (University of Tokyo, 1980) (unpublished).
[12] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. I.
[13] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.
[14] L. P. Horwitz and I. M. Sigal, Helv. Phys. Act. **51**, 685–715 (1978).

Contacts of space–times

M. D. Maia
Instituto de Ciencias, Exatas, Universidade de Brasília, Departmento de Física, 70.910 Brasília, D. Federal, Brazil

(Received 31 March 1980; accepted for publication 24 October 1980)

The concept of contact between manifolds is applied to space–times of general relativity. For a given background space–time a contact approximation of second order is defined and interpreted both from the point of view of a metric pertubation and of a higher order tangent manifold. In the first case, an application to the high frequency gravitational wave hypothesis is suggested. In the second case, a constant curvature tangent bundle is constructed and suggested as a means to define a ten parameter local space–time symmetry.

PACS numbers: 04.20.Cv

1. INTRODUCTION

The Minkowski tangent spaces to a given space–time of general relativity represent the state of zero gravity. Except for the metric signature and Lorentz symmetry, these tangent spaces do not contribute substantially to the knowledge of the local geometry of the space–time itself. On the other hand, it is a known fact that the geometry of a manifold in a small neighborhood of a point can be expressed in terms of osculating or contact geometries. The construction and physical interpretation of similar contact structure for space–times of general relativity is the object of this note.

Mathematically speaking, there is one basic motivation for the construction of these contact structures. Because of the geometrical nature of the gravitational field, the two-body problem for local gravitational fields may be treated by means of the concept of contact of two space–times. In this fashion, a completely geometrical interaction can be described. The usual metric perturbation methods rely on the hypothesis of convergence of a power series expansion of the background metric. Because the experimental evidences in general relativity are scarce, this hypothesis may be critisized. By use of the contact method such difficulty can at least be minimized reducing the mentioned series to a Taylor expansion of some basic functions.

From the physical point of view, two motivations are considered. One of them looks for a better description of the high frequency gravitational wave as emitted by a collapsing background. In this case a contact approximation of second order is regarded as a linear approximation where the perturbation term depends on the curvature of the background.

The second motivation is derived from the need to define a local symmetry in space–time in substitution to the Poincaré group. The Poincaré group is a symmetry which is broken by the gravitational field. This fact makes that group unsuitable to be used as a symmetry in a quantum field theory or in a group theoretic description of particles on a curved background. The construction of tangent space–times with nonzero constant curvature may be used to define adequate groups of symmetry.

2. CONTACT APPROXIMATION

The notion of contact between manifolds was firstly introduced by Cartan.[1] It generalizes the elementary notion of contact between curves and surfaces. The basic concept can be described as follows: If $X, \bar{X}$ are two mappings of a manifold S into a homogeneous space G/H then X and $\bar{X}$ are said to have a contact of order k, if for each $q \in S$ there is a $g \in G$ such that X and $g \circ \bar{X}$ agree up to the order k at q.[2] Here this concept is addapted to the case where X and $\bar{X}$ are two local embeddings of space–times in a p-dimensional pseudo-Euclidean space $M(r,s)$ with signature $r+s$. Two problems may, in principle, be considered: In the first case X and $\bar{X}$ are the embedding of two distinct space–times S and $\bar{S}$ respectively. In the second case X and $\bar{X}$ are distinct embeddings of a single space–time S. However, since X and $\bar{X}$ are taken to be isometric embeddings, they are equivalent up to an isometry of $M(r,s)$. Therefore, the second problem becomes meaningless and only the first one will be considered. In particular, S will be taken to be a given space–time, while $\bar{S}$ will be determined from contact considerations.

The notion of k-order contact can be described in a more general sense by the use of a Grassmann manifold $G_4(M,X(q))$ in $M(r,s)$ which is the set of all four-dimensional planes through a point $X(q)$ of $M(r,s)$. It follows that $G_4(M,X(q))$ is an analytic manifold.[3] At each point $X(q)$ of $M(r,s)$ a Grassmann manifold is constructed so that a Grassmann bundle $B_4(M)$ with fibers $G_4(M,X(q))$ and base $M(r,s)$ may be defined. The embedding $X: S \to M(r,s)$ induces a map $t_X: S \to B_4(M)$ defined by $t_X(q) = X_*(T_q(s))$, where $q \in S$, X_* is the derivative map of X and $T_q(S)$ is the tangent space to S at q. If π' denotes the projection map of $B_4(M)$, it follows that $X = \pi' \circ t_X$ is another embedding of S. Therefore, a first order contact between two embeddings X and $\bar{X}$ at $q \in S$ and $\bar{q} \in \bar{S}$ can be defined by the conditions: $X(q) = \bar{X}(\bar{q})$, $t_X(q) = t_{\bar{X}}(\bar{q})$.

Higher order Grassmann bundles $M^{n+1} = B_4(M^n)$, $n = 0,1,\ldots,k$, may be defined by iteration. Let $M^0 = M(r,s)$, $M^1 = B_4(M^0),\ldots,M^{n+1} = B_4(M^n)$, so that M^{n+1} is the Grassmann manifold of 4-planes through a point of a fiber in M^n. Each of these M^{n+1} is an analytic manifold and the embedding t_X induces an embedding $t^1{}_X$ which by turn induces $t^2{}_X$ and so forth, until $t^n{}_X: S \to M^{n+1}$. If $X_*{}^{n+1}$ denotes the derivative map of $X_*{}^n$ (with $X_*{}^0 = X_*$) and π_n denotes the projection map of the tangent bundle $T^{n+1}S$, then, according to Fig. 1, the induced embeddings are given by

 022-2488/81/030538-05$1.00

$$t_X{}^n = X_*{}^n \circ \pi_n^{-1} \circ \pi_{n-1}^{-1} \circ \cdots \circ \pi_0^{-1},$$

where $\pi_0 = \pi$ is the projection map of the tangent bundle TS.

In terms of these induced embeddings, two embeddings $X, \bar{X}$ are said to have contact of at least order k at $q \in S$, $\bar{q} \in \bar{S}$, if

$$t_X{}^n(q) = t_{\bar{X}}{}^n(\bar{q}), \quad n = 0,1,2,\ldots,k.$$

Since $t_X{}^n$ is determined by the derivative map $X_*{}^n$, it follows that X and $\bar{X}$ have contact of at least order k when the k th-order Taylor polynomials of X and $\bar{X}$ agree at q and $\bar{q}$.[4] In other words, if q^i and $\bar{q}^i$ $(i = 1,\ldots,4)$ are the coordinates of q and $\bar{q}$ respectively in S and $\bar{S}$, then X and $\bar{X}$ have contact of at least order k if

$$\overset{(k)}{X} = X(q) + (x^i - q^i)X_{,i}(q) + \tfrac{1}{2}(x^i - q^i)(x^j - q^j)X_{,ij}(q) + \cdots$$
$$= \bar{X}(\bar{q}) + (\bar{x}^i - \bar{q}^i)\bar{X}_{,i}(\bar{q}) + \tfrac{1}{2}(\bar{x}^i - \bar{q}^i)(\bar{x}^j - \bar{q}^j)\bar{X}_{,ij}(q) + \cdots$$
$$= \overset{(k)}{\bar{X}},$$

where $\overset{(k)}{X}, \overset{(k)}{\bar{X}}$ denote the k th-order Taylor polynomials of X and $\bar{X}$ respectively.

The k th-order contact neighborhood of a point q in S is the set of points $x \in S$ on which the contact of order k with $\bar{S}$ is made. It has a radius λ such that

$$\lambda^k = \|x - q\|^k = (g_{ij}(x^i - q^i)(x^j - q^j))^{k/2} < \epsilon^k,$$

where g_{ij} is the metric tensor of S and ϵ a sufficiently small number so that powers of ϵ greater than k are neglected.

Taking X as a position vector in $M(r,s)$, a Cartesian frame in that space can be chosen so that the Taylor polynomial $\overset{(k)}{X}$ may be expressed in terms of its Cartesian components $\overset{(k)}{X}{}^\mu(x^i)$ (all Greek indices run from 1 to p while small case Latin indices run from 1 to 4).

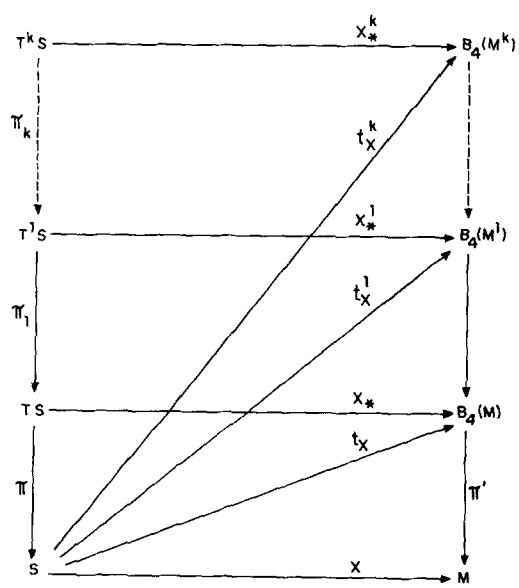


FIG. 1.

Given a space–time S with minimal embedding X in $M(r,s)$, the k th-order *contact approximation* to S at q is a space–time $\overset{(k)}{S}$ defined by the k th-order Taylor polynomial $\overset{(k)}{X}{}^\mu$. For example, the first order contact approximation $\overset{(1)}{S}$ is just the Minkowski tangent space–time to S at q generated in $M(r,s)$ by the coordinates:

$$\overset{(1)}{X}{}^\mu = X^\mu(q) + (x^i - q^i)X^\mu{}_{,i}(q). \tag{2.1}$$

The second order contact approximation $\overset{(2)}{S}$ defined by

$$\overset{(2)}{X}{}^\mu = X^\mu(q) + (x^i - q^i)X^\mu{}_{,i}(q) + \tfrac{1}{2}(x^i - q^i)(x^j - q^j)X^\mu{}_{,ij}(q) \tag{2.2}$$

(plus the condition that it satisfies Einstein's equation) represents the simplest nontrivial approximation to the space–time S at a neighborhood of q

3. GEOMETRY OF $\overset{(2)}{S}$

Since S has embedding class $p - 4$, there are $p - 4$ unit vectors $\eta_{(A)}$ in $M(r,s)$ orthogonal to S and to themselves (capital Latin indice run from 5 to p). A point of S in $M(r,s)$ is specified by a set of p Cartesian coordinates $X^\mu(x^i)$ while a point of M not in S has Cartesian coordinates

$$Z^\mu(x^i,x^A) = X^\mu(x^i) + x^A\eta^\mu{}_{(A)}(x^i), \tag{3.1}$$

where x^A are $p - 4$ parameters.

A Gaussian coordinate system in $M(r,s)$ may also be defined with coordinates $x^\alpha = (x^i,x^A)$. If $\eta_{\mu\nu}$ denote the Cartesian components of the metric tensor of $M(r,s)$, its Gaussian components are given by

$$g_{\alpha\beta}(x^i,x^A) = Z^\mu{}_{,\alpha}Z^\nu{}_{,\beta}\eta_{\mu\nu};$$

in particular, $g_{AB} = \eta^\mu{}_{(A)}\eta^\nu{}_{(B)}\eta_{\mu\nu}$. Since the vectors $\eta^\mu{}_{(A)}$ are orthonormal, then $g_{AB} = \epsilon^A\delta_{AB}$, where $\epsilon^A = \pm 1$, $A = 5,\ldots,\rho$, depending on the signature of $M(r,s)$. On the other hand, considering that $X^\mu{}_{,i}$ are vectors tangent to S

$$\eta_{\mu\nu}X^\mu{}_{,i}\eta^\nu{}_{(A)} = 0, \tag{3.2}$$

then

$$g_{iA} = (X^\mu{}_{,i} + x^B\eta^\mu{}_{(B),i})\eta^\nu{}_{(A)}\eta_{\mu\nu}$$
$$= x^B\eta^\mu{}_{(B),i}\eta^\nu{}_{(A)}\eta_{\mu\nu} = x^B P_{ABi},$$

where $P_{ABi} = \eta^\mu{}_{(B),i}\eta^\nu{}_{(A)}\eta_{\mu\nu}$. It follows that $P_{ABi} + P_{BAi} = 0$ and $g_{iA}|_S = 0$, where $|_S$ denotes the restriction to S. Finally, again using (3.2),

$$g_{ij} = X^\mu{}_{,i}X^\nu{}_{,j}\eta_{\mu\nu} + 2x^A X^\mu{}_{,(i}\eta^\nu{}_{(A),j)}\eta_{\mu\nu}$$
$$+ x^A x^B\eta^\mu{}_{(A),i}\eta^\nu{}_{(B),j}\eta_{\mu\nu}, \tag{3.3}$$

so that $g_{ij}|_S = g_{ij}(S)$. If $g^{\alpha\beta}$ denotes the inverse matrix of $g_{\alpha\beta}$, the Christoffel symbols of the first and second types in the Gaussian system are defined in the usual manner and $\Gamma_{\alpha\beta\gamma} = g_{\gamma\delta}\Gamma^\delta_{\alpha\beta}$. The restriction of some of these components to S gives some useful identities.[5] In particular, denote

$$b_{ijA} = -\Gamma_{ijA}|_S = +\tfrac{1}{2}g_{ij,A}|_S.$$

From (3.3)

$$g_{ij,A} = 2X^{\mu}{}_{,(i}\eta^{\nu}{}_{(A),j)}\eta_{\mu\nu} + 2x^{B}\eta^{\mu}{}_{(A),(i}\eta^{\nu}{}_{(B),j)}\eta_{\mu\nu},$$

where symmetrization applies to i, j only. Therefore

$$b_{ijA} = +\eta_{\mu\nu}X^{\mu}{}_{,(i}\eta^{\nu}{}_{(A),j)}.$$

Differentiating (3.2) with respect to x^i and replacing in the above expression gives

$$b_{ijA} = -\eta_{\mu\nu}\eta^{\nu}{}_{(A)}X^{\mu}{}_{,ij}|_S. \tag{3.4}$$

From this expression it follows that

$$X^{\mu}{}_{,ij} = -g^{AB}b_{ijA}\eta^{\mu}{}_{(B)} + \Gamma_{ij}{}^{k}X^{\mu}{}_{,k}.$$

At this point it may be convenient to use geodesic coordinates at q in S so that the above expression is reduced to

$$X^{\mu}{}_{,ij} = -g^{AB}b_{ijA}\eta^{\mu}{}_{(B)}. \tag{3.5}$$

The curvature radius of S at the point q, corresponding to a normal $\eta_{(A)}$ and a displacement dx^i, is a particular value of x^A such that Z^{μ} remains fixed at the center of curvature[6]:

$$\Delta Z^{\mu} = Z^{\mu}{}_{,i}dx^{i} = (X^{\mu}{}_{,i} + x^{A}\eta^{\mu}{}_{(A),i})\,dx^{i} = 0 \quad \text{(no sum)}. \tag{3.6}$$

Contracted multiplication of this equation by $\eta_{\mu\nu}X^{\nu}{}_{,j}$ and $\eta_{\mu\nu}\eta^{\nu}_{(B)}$ gives respectively

$$(g_{ij} + x^{A}b_{ijA})dx^{i} = 0, \tag{3.7}$$

and

$$P_{ABi}dx^{i} = 0. \tag{3.8}$$

Equation (3.7) has nontrivial solution $x^A = \rho^{A}_{(i)}$ when

$$\det(g_{ij} + x^{A}b_{ijA}) = 0. \tag{3.9}$$

The quantities b_{ijA} are intimately associated to the Riemannian curvature of S. This follows from the vanishing of the Riemann tensor of $M(r,s)$ writen in the Gaussian system: $R_{\alpha\beta\gamma\delta} = 0$. Then $R_{\alpha\beta\gamma\delta}|_S = 0$ and in particular $R_{ijkl}|_S = 0$, from which Gauss equation is obtained:

$$R^{i}{}_{jkl}(S) = 2g^{im}(S)g^{AB}b_{m[k|A|}b_{l]jB}. \tag{3.10}$$

From (3.5) and (2.2) it follows that

$$\overset{(2)}{X}{}^{\mu} = X^{\mu}(q) + (x^{i} - q^{i})X^{\mu}{}_{,i} - \tfrac{1}{2}(x^{i} - q^{i})(x^{j} - q^{j})\,g^{AB}b_{ijA}\eta^{\mu}{}_{(B)}. \tag{3.11}$$

Then the metric of $\overset{(2)}{S}$ is given by

$$\overset{(2)}{g}_{ij} = \overset{(2)}{X}{}^{\mu}{}_{,i}\overset{(2)}{X}{}^{\nu}{}_{,j}\eta_{\mu\nu} = g_{ij}(q) + \gamma_{ij}, \tag{3.12}$$

where

$$\gamma_{ij} = (x^{k} - q^{k})(x^{l} - q^{l})g^{AB}b_{ikA}(q)b_{jlB}(q). \tag{3.13}$$

In the case of second order contact, γ_{ij} is a small quantity of the order of λ^2. Then (3.12) may be compared to the linearized space–time solution of Einstein's equations. In fact, in a suitable coordinate system, $g_{ij}(q)$ can be written as η_{ij} so that (3.12) looks exactly like the usual linear approximation when γ_{ij} is regarded as a perturbation on the flat metric. However, there are some differences to be noted. Here the background space–time is S and the additional term γ_{ij} is a function of the curvature of S. If S were flat then γ_{ij} would be zero as a consequence of (3.10). Also note that while $g_{ij}(q)$ is calculated at a point of S, γ_{ij} is defined in the contact neighborhood.

It will be of interest to express b_{ijA} in terms of the curvature radii of the background. This can be obtained from an additional condition imposed on $\overset{(2)}{S}$. From (3.12), $g_{ij}(q) = \overset{(2)}{g}_{ij} - \gamma_{ij}$. In Eq. (3.9) it follows that

$$\det(\overset{(2)}{g}_{ij} - \gamma_{ij} + x^{A}b_{ijA}) = 0. \tag{3.14}$$

Now suppose tht for a solution $x^A = \rho^{A}_{(i)} \neq 0$ of (3.9) it is also true that $\overset{(2)}{g}_{ij} + \rho^{A}_{(i)}b_{ijA} = 0$. That is

$$b_{ijA} = -\overset{(2)}{g}_{ij}/\rho^{A}_{(i)}. \tag{3.15}$$

Since $\overset{(2)}{b}_{ijA} = b_{ijA}$, the definition of the curvature radius of $\overset{(2)}{S}$:

$$\det(\overset{(2)}{g}_{ij} + \overset{(2)}{\rho}{}^{A}_{(i)}\overset{(2)}{b}_{ijA}) = 0, \tag{3.16}$$

gives after using (3.15)

$$\det\left[\overset{(2)}{g}_{ij}(1 - \overset{(2)}{\rho}{}^{A}_{(i)}/\rho^{A}_{(i)})\right] = 0,$$

so that $\overset{(2)}{\rho}{}^{A}_{(i)} = \rho^{A}_{(i)}$. Therefore the condition (3.15) ensures that the curvature radii of S and $\overset{(2)}{S}$ coincide at q. Then (3.16) can be rewritten as

$$\det(\overset{(2)}{g}_{ij} + \rho^{A}_{(i)}b_{ijA}) = 0$$

which is trivially satisfied in view of (3.15). As a matter of fact, Eq. (3.7) for $\overset{(2)}{S}$ becomes identically satisfied. This means that with condition (3.15) all directions dx^i define a curvature line through q and consequently the point q in $\overset{(2)}{S}$ looks like an umbilic point and the suffix (i) may be dropped in $\overset{(2)}{\rho}{}^{A}_{(i)}$. Denoting $\overset{(2)}{\rho}{}^{A} = \overset{(2)}{\rho}{}^{A}_{(i)}$ for a given (i), a single curvature radius in the subspace of $M(r,s)$ normal to S may be defined by

$$\frac{1}{\overset{(2)}{\rho}{}^{2}} = g^{AB}\frac{1}{\overset{(2)}{\rho}{}^{A}}\frac{1}{\overset{(2)}{\rho}{}^{B}}. \tag{3.17}$$

4. HIGH FREQUENCY WAVES

Now consider some properties of the intrinsic geometry of $\overset{(2)}{S}$ when it is regarded as a space–time of general relativity.

The covariant metric tensor of $\overset{(2)}{S}$ is defined so that $\overset{(2)}{g}{}^{ij}\overset{(2)}{g}_{jk} = \delta^{i}{}_{k}$. Within the considered order of approximation, $\overset{(2)}{g}{}^{ij}$ may be written as

$$\overset{(2)}{g}{}^{ij} = g^{ij}(q) - g^{ik}(q)g^{jl}(q)\gamma_{kl}. \tag{4.1}$$

The condition for $\overset{(2)}{S}$ to be a solution of Einstein's equations is obtained exactly as in the case of as the linear solution, where the Minkowski metric η_{ij} is replaced by $g_{ij}(q)$.

The Christoffel symbols $\overset{(2)}{\Gamma}{}^{i}_{jk}$ of $\overset{(2)}{g}_{ij}$ are

$$\overset{(2)}{\Gamma}{}^{i}_{jk} = \tfrac{1}{2}g^{il}(q)(\gamma_{jl,k} + \gamma_{kl,j} - \gamma_{jk,l}).$$

Introducing the notations

$$\gamma = g^{kl}(q)\gamma_{kl}, \quad \psi_{ij} = \gamma_{ij} - \tfrac{1}{2}g_{ij}(q)\gamma, \quad \tau_i = g^{kl}(q)\psi_{ik,l},$$

then Einstein's tensor for $\overset{(2)}{S}$, may be written as

$$\overset{(2)}{G}_{ij} = \tfrac{1}{2}[\, g^{kl}(q)\psi_{ij,kl} - \tau_{i,j} - \tau_{j,i} + g_{ij}(q)g^{kl}(q)\tau_{k,l}\,]. \tag{4.2}$$

Choosing the usual gauge $g^{kl}(q)\psi_{ik,l} = 0$, the linear wave equation follows from Einstein's Equation $G_{ij} = KT_{ij}$. Therefore, a time-dependent ψ_{ij} may have a gravitational wave interpretation as seen by an observer in S inside the contact neighborhood with wave ripples superimposed on the curved background. Notice that the perturbation term is derived from the curvature of S. This fact can be seen from (3.10) which relates b_{ijA} to $R^{i}{}_{jkl}(S)$. Alternatively b_{ijA} can be expressed in terms of the curvature radius of S as derived from (3.9), or explicitly from (3.15), if the umbilic condition is used. Considering the last case, the perturbation term (3.13) can be written as

$$\gamma_{ij} = (x^k - q^k)(x^l - q^l)\, g^{AB} \frac{\overset{(2)}{g}_{ik}}{\overset{(2)}{\rho}{}^{A}} \frac{\overset{(2)}{g}_{jl}}{\overset{(2)}{\rho}{}^{B}} \quad . \tag{4.3}$$

Define a tensor h_{ij} by

$$\overset{(2)}{g}_{ik}\overset{(2)}{g}_{jl} = h_{ij}\, g_{kl}.$$

Then $\gamma_{ij} = (\lambda^2/\rho^2)h_{ij}$ and (3.12) becomes

$$\overset{(2)}{g}_{ij} = g_{ij}(q) + (\lambda/\rho)^2 h_{ij}. \tag{4.4}$$

This expression can be compared to the expression obtained in the high frequency limit hypothesis, according to which the metric g_{ij} is expanded in powers of the ratio of the wavelength to the curvature radius. In the case of a strong curvature the curvature radius is small and the power series is assumed to converge when the wavelength is comparatively smaller,[7,8] Although it appear to be a reasonable hypothesis, the series convergence is questionable.[9] Considering the analogy with the contact approximation, the only expansion required is the Taylor expansion of the functions X^{μ} and no additional expansion parameter is postulated. Thus, if λ is to be compared to the wavelength in the high frequency wave theory, it follows that the wavelength is nothing but a measure of the size of the contact neighborhood. Accordingly, if the contact neighborhood in S has strong curvature, then the frequency of the emitted gravitational waves has to be high to keep the ratio $(\lambda/\rho)^2$ small. Notice, however, that nothing prevents the consideration of low frequency waves as derived form (4.4). In fact, if S has weak curvature in the contact neighborhood, then ρ will be large and λ can also be a relatively large length. Therefore, it appears that (4.4) can describe a wide band local gravitational radiation emitted by the background.

5. CONSTANT CURVATURE CONTACT BUNDLE

A particularly interesting case is that when $\overset{(2)}{S}$ is specialized to have constant curvature. It was seen that when $\overset{(2)}{S}$ satisfies Eq. (3.15), it will show q as an umbilic point.

A straightforward calculation of the Riemannian curvature of $\overset{(2)}{S}$ gives

$$\overset{(2)}{R}_{ijkl} = 2[\, g^{AB} - (x^r - q^r)(x^s - q^s) \times \overset{(2)}{g}{}^{mn} g^{AC} g^{BD} b_{nrC} b_{msD}\,]\, b_{j[l|A}\, b_{i|k]B},$$

where indices inside vertical bars are not under skew symmetrization. Then, after use of (3.15) with the notation $\overset{(2)}{\rho}$, the last expression becomes

$$R_{ijkl} = g^{AB}\left(\frac{\overset{(2)}{g}_{jl}}{\overset{(2)}{\rho}{}^{A}}\frac{\overset{(2)}{g}_{ik}}{\overset{(2)}{\rho}{}^{B}} - \frac{\overset{(2)}{g}_{jk}}{\overset{(2)}{\rho}{}^{A}}\frac{\overset{(2)}{g}_{il}}{\overset{(2)}{\rho}{}^{B}}\right) + \overset{(2)}{g}{}^{mn}(x^r - q^r)(x^s - q^s) g^{AB} g^{CD} \times \frac{\overset{(2)}{g}_{nr}}{\overset{(2)}{\rho}{}^{B}}\frac{\overset{(2)}{g}_{ms}}{\overset{(2)}{\rho}{}^{D}}\left(\frac{\overset{(2)}{g}_{jk}}{\overset{(2)}{\rho}{}^{A}}\frac{\overset{(2)}{g}_{il}}{\overset{(2)}{\rho}{}^{C}} - \frac{\overset{(2)}{g}_{jl}}{\overset{(2)}{\rho}{}^{A}}\frac{\overset{(2)}{g}_{ik}}{\overset{(2)}{\rho}{}^{C}}\right)$$

or

$$\overset{(2)}{R}_{ijkl} = \frac{1}{\overset{(2)}{\rho}{}^{2}}\left(1 - \frac{\lambda^2}{\overset{(2)}{\rho}{}^{2}}\right)(\overset{(2)}{g}_{jl}\overset{(2)}{g}_{ik} - \overset{(2)}{g}_{jk}\overset{(2)}{g}_{il}). \tag{5.1}$$

Now, for a contact of second order, $\lambda^2/\overset{(2)}{\rho}{}^{4}$ is negligible inside the contact neighborhood. Then it follows that

$$\overset{(2)}{R}_{ijkl} = K^2(\overset{(2)}{g}_{jl}\overset{(2)}{g}_{ik} - \overset{(2)}{g}_{jk}\overset{(2)}{g}_{il}) \tag{5.2}$$

which characterize $\overset{(2)}{S}$ as a constant curvature manifold with $K^2 = \overset{(2)}{\rho}{}^{-2}$. Under the condition that $\overset{(2)}{S}$ is also a solution of Einstein's equations, in view of the results in Sec. 4, $\overset{(2)}{S}$ represents a linear solution with constant curvature. These constant curvature contact space–times, denoted $\overset{(2)}{S}_{q}(\rho)$ are minimally embedded in a five-dimensional subspace of $M(r,s)$ [either $M(4,1)$ or $M(3,2)$]. They can be defined in all regular point of S with curvature such that $\lambda \ll \rho^2$. This means that the diameter of the contact neighborhood has to be sufficiently small in regions of S with strong curvature, but it can

be fairly large for regions of weak curvature. Thus a second order contact bundle $\overset{(2)}{S}(\rho)$ with constant curvature with fibers $\overset{(2)}{S}$ and base S may be constructed. In particular when $\overset{(2)}{S}_q(\rho)$ is a linearized version of the de Sitter space–time with curvature radius $\overset{(2)}{\rho}$ (a function of the position in S) a de Sitter bundle is obtained. This de Sitter bundle shows a local de Sitter symmetry at each point of S. Such bundle has been proposed as an additional property of the space–times of general relativity.[10,11] Its mathematical existence and construction, as shown above, may be taken as a consequence of the local differential geometry of the space–times.

[1]E. Cartan, *La Méthode Du Repere Mobile, la Théorie des Groups Continues, et les Espaces Généralisés* (Herman, Paris, 1975).
[2]P. Griffiths, Duke Math. J. **41**, 775 (1974).
[3]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, N.J., 1951), 1st ed., p. 35.
[4]G. R. Jensen, "Higher Order Contact of Submanifolds of Homogeneous Spaces" in *Lecture Notes in Mathematics* 610 (Springer–Verlag, Berlin, 1977) 1st Edition, p. 4.
[5]M. D. Maia, Rev. Brasileira Física **8**, 428 (1978).
[6]L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P. Princeton, N.J., 1966), 6th printing, p. 143.
[7]D. R. Brill and J. B. Hartle, Phys. Rev. B **135**, 271 (1964).
[8]R. A. Isaacson, Phys. Rev. **166**, 1263 (1968).
[9]V. D. Zakharov, *Gravitational Waves in Einstein's Theory* (Halsted Wiley, New York, 1973).
[10]L. Halpern, Gen. Rel., Grav. **8**, 623 (1977).
[11]A. Böhn, "Relativistic Rotators: A Quantum Mechanical de Sitter Bundle," Preprint (1976).

Intrinsic isometry groups in general relativity

D. A. Szafron[a)]
Department of Mathematics, Portland State University, P. O. Box 751, Portland, Oregon 97207

(Received 16 July 1980; accepted for publication 19 September 1980)

We derive necessary and sufficient tensor conditions for the existence of a four-parameter isometry group G_4 which acts multiply transitively on a Riemannian V_3. We then apply these results to determine which spatially homogeneous cosmological models have induced 3-metrics which are invariant under such a four-parameter group.

PACS numbers: 04.20.Jb

1. INTRODUCTION

In this paper we define a space-time to be a four-dimensional manifold V_4 with a Lorentzian metric g which satisfies the Einstein field equations. Space-times are often found, studied and invariantly classified using n-parameter isometry groups G_n under which g is invariant. If $n \leq 6$ and G_n acts transitively on orbits which are three-dimensional spacelike hypersurfaces, then $n = 3$, 4 or 6 and the space-times are called spatially homogeneous. When $n = 6$ they are also isotropic and are called Friedmann–Robertson–Walker (FRW) models (cf. Ellis and MacCallum[1] and Eisenhart[2] for terminology.)

Historically, isometry groups have two major applications in general relativity. The imposition of an isometry group on a V_4 makes the field equations more tractable so that new exact solutions may be found. Secondly, isometry groups are used to invariantly classify known space-times. This technique fails however for space-times whose metrics are not invariant under any G_n. The Szekeres[3] space-times which admit no isometry group (cf. Bonnor, Sulaiman and Tomimura[4]) are a good example. The Szekeres space-times are not generic solutions to the field equations, since studies by Collins and Szafron[5] and Berger, Eardley and Olson[6] have shown that the three-dimensional spacelike hypersurfaces V_3^* are conformally flat. In fact, there is a subclass of the Szekeres models in which the induced 3-metric g* on each hypersurface V_3^* is invariant under a G_6^* (since each V_3^* is flat) while the 4-metric g on V_4 is not invariant under *any* isometry group G_n.

These results suggest two applications of "intrinsic symmetries". The first is to use a combination of intrinsic symmetries and extrinsic symmetries, to invariantly classify known space-times, even if they possess no isometry group G_n under which g is invariant. The second application is to impose a combination of intrinsic symmetries (spatial flatness, spatial conformal flatness, restrictions on the eigenvalues of the 3-Ricci tensors of the V_3^*'s, the existence of G_n^*'s under which the g*'s are invariant, etc.) and extrinsic symmetries (restrictions on the eigenvalues of the expansion tensor, restrictions on the acceleration vector, etc.) in order to make the field equations more tractable in the search for new exact solutions. This program has been started in Collins and Szafron,[7] Szafron and Collins,[8] Collins and Szafron,[9] Spero and Szafron[10] and Wainwright.[11]

The easiest way to impose intrinsic symmetries is by tensor conditions on the quantities of the V_3^*. "Spatial flatness", the demand that each V_3^* is flat, is equivalent to demanding that the 3-Ricci tensors of the V_3^*'s vanish. "Spatial conformal flatness", the demand that each V_3^* is conformally flat, is equivalent to demanding that the Cotton–York tensor (cf. York[12] and Wainwright[13]) vanishes.

In this paper, we begin to explore the consequences of imposing isometry groups G_n^* on the induced 3-metrics g* of spacelike hypersurfaces V_3^*. In fact, Ellis and MacCallum[14] have already begun this study from a different point of view. It is well known (cf. Eisenhart[15]) that the existence of a G_6^* under which g* of a V_3^* is invariant, is equivalent to the tensor condition: the 3-Ricci tensor is isotropic (has three equal eigenvalues) or equivalently that the trace free 3-Ricci tensor vanishes. They have imposed this tensor condition on spatially homogeneous space-times. They delineate all Bianchi–Behr types in which this condition is possible and indicate when it leads to the existence of a G_4 or G_6 under which the full space-time metric g is invariant.

We generalize these results by studying the effects of the existence of G_4^*'s under which g* on each V_3^* is invariant (the existence of a G_5^* is of course impossible, c.f. Eisenhart[16]). In order to impose this condition easily, it must first be cast in tensor form. Such a formulation is independent of the V_4 in which the V_3^*'s are imbedded and independent of the field equations. For this reason, the result should be of general interest in differential geometry. We derive such a condition in Sec. 2., and believe it to be new.

In Sec. 3., we review the results of Ellis and MacCallum on G_6^*'s and use the formulation in Sec. 2 to impose a G_4^* on the induced 3-metrics g* of the spatially homogeneous space-times with interesting results. It turns out that *every* Bianchi–Behr type II or type VI ($h = -1$) space-time has a G_4^* under which the 3-metrics g* of the hypersurfaces are invariant, but do not in general admit a G_4. We give a complete list of which Bianchi–Behr types may admit a G_4^* and under what conditions the G_4^* becomes a G_4.

2. MULTIPLY TRANSITIVE G_4 ON A V_3

In this section V_3 will be a three-dimensional Riemannian manifold with positive definite metric tensor g:

[a)]This material is based upon work supported by the National Science Foundation under Grant PHY-7911923.

$$ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta . \tag{2.1}$$

Since V_3 does not have to be imbedded in a Lorentzian V_4 we shall omit the *'s which were used in Sec. 1. Throughout this paper, Greek indices will run from 1 to 3 and Latin indices will run from 0 to 3. Semicolon denotes a covariant derivative and comma denotes a partial derivative. We denote the components of the Ricci tensor by $R_{\alpha\beta}$, the Ricci scalar by R, the trace-free Ricci tensor components by

$$S_{\alpha\beta} = R_{\alpha\beta} - \tfrac{1}{3}Rg_{\alpha\beta}, \tag{2.2}$$

and define the scalar S, by

$$S^2 = \tfrac{1}{2}S_{\alpha\beta}S^{\alpha\beta} . \tag{2.3}$$

The trace-free Cotton–York tensor $C^{\alpha\beta}$ is defined by

$$C^{\alpha\beta} = 2\eta^{\alpha\lambda\mu}(R^\beta{}_\lambda - \tfrac{1}{4}\delta^\beta{}_\lambda R)_{;\mu} . \tag{2.4}$$

We shall use the orthonormal tetrad formalism of Ellis and MacCallum[17] throughout this paper, but in this section the tetrad is really only a triad since we are in V_3. We shall prove the following:

Theorem: Let V_3 be a three-dimensional Riemannian manifold with metric tensor **g**. The following are equivalent:

(1) **g** is invariant under a four-parameter isometry group G_4, multiply transitive on V_3.

(2) There exist coordinates on V_3 in which

$$\begin{aligned} &ds^2 = (\omega^1)^2 + (\omega^2)^2 + (\omega^3)^2 , \\ &\omega^1 = dx - (zB/G)dy + (yB/G)dz , \\ &\omega^2 = (A/G)dy, \quad \omega^3 = (A/G)dz , \end{aligned} \tag{2.5}$$

where $G = 1 + \frac{1}{4}k(y^2+z^2)$, $k = 0$, 1, or -1; and A, B are constants.

(3) $R_{\alpha\beta}$ has two equal eigenvalues, the spatial gradient of S vanishes ($S_{,\mu} = 0$)

and

$$C_{\alpha\beta} = \lambda S_{\alpha\beta}, \tag{2.6}$$

where either:

(i) λ constant, $\lambda S_{\alpha\beta} \neq 0$;

(ii) $\lambda = 0$, $S_{\alpha\beta} \neq 0$, and $R_{\alpha\beta}$ has zero as its third eigenvalue;

or

(iii) $S_{\alpha\beta} = 0$.

Remarks: From Eq. (2.11), it will be seen that the spatial gradient of R also vanishes, but it is not necessary to assume this in (3).

In the coordinates (2.5) the four Killing vectors of the theorem are:

$$\begin{aligned} &\xi_1 = 4By\,\partial/\partial x + 2kyz\,\partial/\partial y + (4 + kz^2 - ky^2)\partial/\partial z , \\ &\xi_2 = -4Bz\,\partial/\partial x + (4 + ky^2 - kz^2)\partial/\partial y + 2kyz\,\partial/\partial z , \\ &\xi_3 = \partial/\partial x, \quad \xi_4 = y\,\partial/\partial z - z\,\partial/\partial y . \end{aligned} \tag{2.7}$$

The condition that $R_{\alpha\beta}$ has two equal eigenvalues is equivalent to demanding that $S_{\alpha\beta}$ has two equal eigenvalues. We can cast the eigenvalue conditions of (3) into tensor form by defining the invariant

$$T = 4S^6 - 3(S_{\alpha\beta}S^{\beta\lambda}S_\lambda{}^\alpha)^2 .$$

Then $R_{\alpha\beta}$ has two equal eigenvalues if and only if $T = 0$, and the third eigenvalue of $R_{\alpha\beta}$ is zero if and only if

$$T = 4S^6 - 3R^2(S^2 - \tfrac{1}{9}R^2)^2 .$$

Condition (3) part (iii), $S_{\alpha\beta} = 0$, occurs if and only if a G_6 is admitted.

Proof that (1) ⟶ (2): Cartan[18] has shown that if (1) is true, then we can find coordinates (2.4). On the other hand, Killing's equations

$$g_{\alpha\beta}\xi^\alpha{}_{,\lambda} + g_{\alpha\lambda}\xi^\alpha{}_{,\beta} + g_{\beta\lambda,\alpha}\xi^\alpha = 0 ,$$

in coordinates (2.5) can be solved by a long and tedious procedure and four independent solutions (2.7) can be found. [It is not too difficult to check, however, that the four vector fields (2.7) are solutions.]

Proof that (2) – (3): We solve the equations

$$\omega^\alpha{}_\lambda e_\beta{}^\lambda = \delta^\alpha{}_\beta , \tag{2.8}$$

where $\omega^\alpha{}_\lambda$ are given by (2.5) to obtain the tetrad (which is really a triad) dual to the one forms of (2.5):

$$\mathbf{e}_1 = \frac{\partial}{\partial x}, \quad \mathbf{e}_2 = \frac{G}{A}\frac{\partial}{\partial y} + \frac{zB}{A}\frac{\partial}{\partial x}, \quad \mathbf{e}_3 = \frac{G}{A}\frac{\partial}{\partial z} - \frac{yB}{A}\frac{\partial}{\partial x} . \tag{2.9}$$

Using the formulas (A1) of Appendix A, we compute the commutation coefficients of (2.9):

$$\begin{aligned} &a_1 = n_{22} = n_{33} = n_{23} = 0 , \\ &a_3 = -n_{12} = \tfrac{1}{4}zkA^{-1} , \\ &a_2 = n_{13} = \tfrac{1}{4}ykA^{-1} , \\ &n_{11} = -2BA^{-2} . \end{aligned} \tag{2.10}$$

We then use formulas (A4) and (2.10) to compute the components of the Ricci tensor in the orthonormal tetrad (2.9):

$$\begin{aligned} &R_{11} = 2B^2/A^4, \quad R_{22} = R_{33} = (kA^2 - 2B^2)/A^4 , \\ &R_{12} = R_{13} = R_{23} = 0 . \end{aligned} \tag{2.11}$$

Equations (2.2) and (2.11) then yield

$$S_{11} = (8B^2 - 2kA^2)/3A^4, \quad S_{22} = S_{33} = (kA^2 - 4B^2)/3A^4 . \tag{2.12}$$

If we compute the tetrad components of $C_{\alpha\beta}$, using (2.4) or the formulas (A5) of Appendix A (which is much easier), we find, on using (2.11) that

$$C_{\alpha\beta} = -6BS_{\alpha\beta}/A^2 . \tag{2.13}$$

The proof is thus finished.

The proof that (3) – (2) is quite complicated and we relegate the most tedious part to a preliminary lemma given in Appendix B.

Proof that (3) – (2): If the Ricci tensor is isotropic (or equivalently $S_{\alpha\beta} \equiv 0$) then it is well known that V_3 admits a G_6 (cf. Eisenhart[19]) so that (1) is satisfied. But (1) – (2) has been shown so the proof is complete. Henceforth, we assume

$$S_{\alpha\beta} \neq 0 . \tag{2.14}$$

Then by the lemma of Appendix B, we choose a tetrad which is a Ricci eigenframe in which:

$$a_1 = n_{13} - a_2 = n_{12} + a_3 = n_{23} = n_{22} = n_{33} = 0\,, \tag{2.15}$$

and

$$\lambda = 3n_{11}\,. \tag{2.16}$$

We write out the formula for $R_{22} - R_{11}$ from (A4), subject to (2.15) and (2.16):

$$R_{22} - R_{11} = 2\vartheta_2 a_2 + 2\vartheta_3 a_3 - 4a_3{}^2 - 4a_2{}^2 - \tfrac{1}{9}\lambda^2\,. \tag{2.17}$$

But $\vartheta_\alpha S = 0$, so by (2.3) and (B8),

$$\vartheta_\alpha (R_{22} - R_{11}) = 0\,,$$

so that from (2.17)

$$2\vartheta_2 a_2 + 2\vartheta_3 a_3 - 4a_3{}^2 - 4a_2{}^2 = \beta\,, \tag{2.18}$$

where β is constant.

We now introduce a coordinate system. From conditions (2.15), applied to the commutators of (A1) we have

$$[\mathbf{e}_1, \mathbf{e}_2] = [\mathbf{e}_1, \mathbf{e}_3] = 0\,,$$

so that $\mathbf{e}_1$ and $\mathbf{e}_2$ form a surface which we label $x_3 \equiv z$ = const, while $\mathbf{e}_1$ and $\mathbf{e}_3$ form a surface which we label $x^2 \equiv y$ = const. Then

$$\mathbf{e}_1 \cdot \partial/\partial z = \mathbf{e}_2 \cdot \partial/\partial z = \mathbf{e}_1 \cdot \partial/\partial y = \mathbf{e}_3 \cdot \partial/\partial y = 0\,,$$

so that we can write

$$\mathbf{e}_1 = a\,\partial/\partial x, \quad \mathbf{e}_2 = b\,\partial/\partial x + c\,\partial/\partial y, \quad \mathbf{e}_3 = r\,\partial/\partial x + s\,\partial/\partial z\,, \tag{2.19}$$

where a, b, c, r, and s are functions of x, y, z. If we apply the coordinate transformation

$$\hat{x} = \int \frac{1}{a}\partial x, \quad \hat{y} = y, \quad z = \hat{z}$$

we obtain:

$$\mathbf{e}_1 = \partial/\partial x, \quad \mathbf{e}_2 = b\,\partial/\partial x + c\,\partial/\partial y, \quad \mathbf{e}_3 = r\,\partial/\partial x + s\,\partial/\partial y\,, \tag{2.20}$$

where we have dropped the hats from x, y, z, b, c, r, and s. Applying the commutation relations (A1) to the tetrad (2.20) and using (2.15) and (2.16) leads to:

$$\begin{aligned} &b_x = c_x = r_x = s_x = 0\,,\\ &a_3 = -n_{12} = (s/2c)c_z, \quad a_2 = n_{13} = (c/2s)s_y\,,\\ &n_{11} = \tfrac{1}{3}\lambda = cr_y - sb_z + (bs/c)c_z - (cr/s)s_y\,, \end{aligned} \tag{2.21}$$

where subscripts denote partial differentiation. Equation (2.18) can now be rewritten, using (2.21), as

$$s(sc_z/c)_z + (cs_y/s)_y - (cs_y/s)^2 - (sc_y/c)^2 = \beta\,. \tag{2.22}$$

A computation of the 1-forms dual [use Eq. (2.8)] to the tetrad (2.20) yields

$$\omega^1 = dx - (b/c)dy\ y - (r/s)dz\,,$$

$$\omega^2 = dy/c, \quad \omega^3 = dz/s\,. \tag{2.23}$$

Consider the line element

$$ds^2 = (\omega^2)^2 + (\omega^3)^2\,. \tag{2.24}$$

A calculation of the Gaussian curvature of a 2-space with line element (2.24), subject to (2.23) yields

$$R_{2323}/(g_{22}g_{33} - g_{23}{}^2) = \beta\,.$$

Thus, (2.24) is a metric on a 2-space of constant curvature. Eisenhart[20] has shown that we can perform a coordinate transformation

$$\hat{y} = \hat{y}(y,z), \quad \hat{z} = \hat{z}(y,z)$$

so that

$$\omega^2 = d\hat{y}/F, \quad \omega^3 = d\hat{z}/F\,,$$

where $F = 1 + \frac{1}{4}\beta(\hat{y}^2 + \hat{z}^2)$. We modify the transformation slightly and drop $\hat{}$'s so that

$$\omega^2 = (A/G)dy, \quad \omega^3 = (A/G)dz\,,$$

where A is constant and $G = 1 + \frac{1}{4}k(y^2 + z^2)$; $k = 0$, 1 or -1. Our 1-forms (2.23) now have the form

$$\begin{aligned} &\omega^1 = dx - (b/G)dy - (r/G)dz\,,\\ &\omega^2 = (A/G)dy, \quad \omega^3 = (A/G)dz\,, \end{aligned} \tag{2.25}$$

where b and r are different arbitrary functions of y and z than appear in (2.23).

We now perform a coordinate transformation

$$\hat{x} = x + f(y,z), \quad \hat{y} = y, \hat{z} = z$$

so that

$$\begin{aligned} &\omega^1 = d\hat{x} - (f_y + bG^{-1})dy - (f_z + rG^{-1})dz\,,\\ &\omega^2 = (A/G)dy, \quad \omega^3 = (A/G)dz\,. \end{aligned} \tag{2.26}$$

We wish to choose $f(y,z)$ to satisfy the two conditions

$$f_y + b/G = zB/G, \quad f_z + r/G = -yB/G\,, \tag{2.27}$$

where B is some constant. The integrability condition for (2.27) is $f_{yz} = f_{zy}$, that is

$$Gr_y - Gb_z + bG_z - rG_y = -2B\,. \tag{2.28}$$

But on comparing (2.23) and (2.25) we see that

$$c = s = (G/A)\,,$$

so that (2.21) becomes:

$$Gr_y - Gb_z + bG_z - rG_y = \tfrac{1}{3}\lambda A\,. \tag{2.29}$$

So the integrability condition (2.28) can be satisfied by choosing $B = -\frac{1}{6}\lambda A$. Dropping $\hat{}$'s, our 1-forms (2.26) become (2.5) and the theorem is proved.

3. SPATIALLY HOMOGENEOUS SPACE-TIMES WITH G_4^*'s

In this section we assume that V_4 is a four-dimensional manifold with a Lorentzian metric $\mathbf{g}$ invariant under a G_3 simply transitive on spacelike hypersurfaces V_3^*. That is, we are looking at all spatially homogeneous space-times except the Kantowski–Sachs models.[21] These space-times have been invariantly classified by the Bianchi–Behr type of groups which they admit. We give this classification in Table I. We seek the answers to two questions:

(1) What are the Bianchi–Behr types of the V_4 which can admit a G_4^* under which the induced metrics $\mathbf{g}^*$ of the spacelike hypersurfaces are invariant?

(2) In which of these space-times is the metric $\mathbf{g}$ invariant under G_4^* so that we really have a G_4?

Included in the results will be the case studied by Ellis and MacCallum[22] for G_6^*'s.

We will perform all calculations in the canonical te-

TABLE I. The Bianchi–Behr classification of spatially homogeneous spacetimes. This information is taken from Ellis and MacCallum.[24] The cannonical tetrad has $a_2=a_3=n_{12}=n_{13}=0$, $a_1=a$, $n_{11}=n_1$, $n_{22}=n_2$, $n_{33}=n_3$ with:

Group Class	Group Type	a	n_1	n_2	n_3	Bianchi type
A	I	0	0	0	0	I
	II	0	+	0	0	II
	VII_0	0	+	+	0	VII
	VI_0	0	+	−	0	VI
	IX	0	+	+	+	IX
	VIII	0	+	+	−	VIII
B	V	+	0	0	0	V
	IV	+	0	0	+	IV
	VII_h	+	0	+	+	VII
	VI_h	+	0	+	−	VI (III if $h=-1$)

trad of Table I. We begin by considering space-times which admit a group from class A. From Eqs. (A4) in the canonical tetrad of Table I:

$$\begin{aligned} R^*_{11}&=\tfrac{1}{2}(n_1{}^2-n_2{}^2-n_3{}^2)+n_2n_3\,,\\ R^*_{22}&=\tfrac{1}{2}(n_2{}^2-n_1{}^2-n_3{}^2)+n_1n_3\,,\\ R^*_{33}&=\tfrac{1}{2}(n_3{}^2-n_1{}^2-n_2{}^2)+n_1n_2\,, \end{aligned} \tag{3.1}$$

and

$$R^*_{12}=R^*_{23}=R^*_{13}=0\,. \tag{3.2}$$

From the theorem and Eq. (3.2), one of the following conditions is necessary:

$$\begin{aligned} &\text{(i) } R^*_{11}=R^*_{22}\neq R^*_{33},\quad C^*_{11}=C^*_{22}\,,\\ &\text{(ii) } R^*_{11}=R^*_{33}\neq R^*_{22},\quad C^*_{11}=C^*_{33}\,,\\ &\text{(iii) } R^*_{22}=R^*_{33}\neq R^*_{11},\quad C^*_{22}=C^*_{33}\,,\\ &\text{(iv) } R^*_{11}=R^*_{22}=R^*_{33}\,. \end{aligned} \tag{3.3}$$

Substitution of conditions (3.3) into (A5) with (3.1) yields the corresponding four possibilities:

$$\begin{aligned} &\text{(i) } n_1=n_2\neq n_3,\quad n_3\neq 0\,,\\ &\text{(ii) } n_1=n_3\neq n_2,\quad n_2\neq 0\,,\\ &\text{(iii) } n_2=n_3\neq n_1,\quad n_1\neq 0\,,\\ &\text{(iv) } n_1=n_2=n_3,\ \text{or } n_1=n_2,\ n_3=0 \text{ or } n_1=n_3,\ n_2=0\\ &\text{or } n_2=n_3,\ n_1=0\,. \end{aligned} \tag{3.4}$$

By the theorem, any of the conditions (3.4) is also sufficient. Condition (iv) is necessary and sufficient for the existence of a G^*_6. We now look at each type from class A subject to the conditions (3.4).

Type I: All space-times satisfy (iv) so there is a G^*_6.

Type II: All space-times satisfy (iii) so there is a G^*_4.

Type VII_0: A subset ($n_1=n_2$) of the space-times satisfy (iv), so they possess a G^*_6. Since this subset has $n_1=n_2\neq 0$, the Jacobi identities (A3) imply that $\theta_2=\theta_3$. This subset is therefore locally rotationally symmetric (cf. Ellis and MacCallum[23]) so there is a $G^*_4\subseteq G^*_6$ which is really a G_4.

Type VI_0: No space-times satisfy any of the conditions (3.4) so no G^*_4 can exist.

Type IX: A subset (which by renumbering takes the form $n_1=n_2\neq n_3, n_3\neq 0$) satisfies one of the first three conditions so there exists a G^*_4. Again by (A3) the G^*_4 is really a G_4. A second subset ($n_1=n_2=n_3$) satisfies (iv) so there is a G^*_6. However $n_1=n_2=n_3\neq 0$ and the Jacobi identities (A3) imply $\theta_1=\theta_2=\theta_3$ so that the shear tensor vanishes thus the G^*_6 is a G_6 (these are some FRW models).

Type VIII: A subset ($n_1=n_2$) satsifies (i) so there is a G^*_4 which by (A3) is really G_4.

We now turn to those space-times which admit a class B group. From Eqs. (A4) in the canonical tetrad of Table I:

$$\begin{aligned} R^*_{11}&=-\tfrac{1}{2}(n_2-n_3)^2-2a_1{}^2\,,\\ R^*_{22}&=\tfrac{1}{2}(n_2{}^2-n_3{}^2)-2a_1{}^2\,,\\ R^*_{33}&=\tfrac{1}{2}(n_3{}^2-n_2{}^2)-2a_1{}^2\,,\\ R^*_{23}&=a_1(n_2-n_3)\,,\\ R^*_{12}&=R^*_{13}=0\,. \end{aligned} \tag{3.5}$$

In order to impose the necessary condition that $R^*_{\alpha\beta}$ has two equal eigenvalues we look at the characteristic equation

$$\det[R^*_{\alpha\beta}-\phi\,\delta_{\alpha\beta}]=0\,. \tag{3.6}$$

Since $R^*_{12}=R^*_{13}=0$, (3.6) becomes

$$(R^*_{11}-\phi)[(R^*_{22}-\phi)(R^*_{33}-\phi)-R^{*2}_{23}]=0\,. \tag{3.7}$$

There are two cases in which (3.7) may have two equal solutions for ϕ. Case one is when R^*_{11} is a repeated root, in which case

$$(R^*_{22}-R^*_{11})(R^*_{33}-R^*_{11})-R^{*2}_{23}=0\,,$$

which upon substituting from (3.5) implies

$$(n_2-n_3)(n_2n_3+a_1{}^2)=0\,. \tag{3.8}$$

The second case is when

$$(R^*_{22}-\phi)(R^*_{33}-\phi)-(R^*_{23})^2=0$$

has two equal roots for ϕ. Using the quadratic formula leads to the equivalent condition

$$R^*_{23}=R^*_{22}-R^*_{33}=0\,. \tag{3.9}$$

By (3.5), this becomes

$$n_2=n_3\,. \tag{3.10}$$

But condition (3.10) implies $S^*_{\alpha\beta}=0$ which is necessary and sufficient for the existence of a G^*_6. If $n_2\neq n_3$ then by (3.8)

$$n_2n_3+a_1{}^2=0\,. \tag{3.11}$$

A long calculation using the general definition of $C^*_{\alpha\beta}$ given in Appendix A [we cannot use (A5) since we are not in a Ricci eigenframe] leads to

$$C^*_{\alpha\beta}=3(n_2+n_3)S^*_{\alpha\beta}\,. \tag{3.12}$$

So by the theorem, condition (3.11) is also sufficient for the existence of a G^*_4. We now look at each type from class B subject to the condition (3.10) or (3.11).

Type V: All space-times satisfy (3.10) so there is a G^*_6.

Type IV: No space-times satisfy (3.10) or (3.11) so no G_4^* can exist.

Type VII$_h$: A subset ($n_2=n_3$) satisfy condition (3.10) s there exists a G_6^*. Since $n_2=n_3\neq 0$, the Jacobi identities (A3) imply that $\theta_2=\theta_3$, so there exists a $G_4^*\subseteq G_6^*$ which is really a G_4.

Type VI$_h$: All space-times with $h=-1$ satisfy (3.11) so there exists a G_4^*. No space-times with other values of h possess a G_4^*.

ACKNOWLEDGMENTS

I am indebted to M.A.H. MacCallum for many helpful suggestions and especially to J. Wainwright whose preprint of Ref. 11 provided the groundwork for part of the theorem in this paper.

APPENDIX A

This Appendix contains formulas related to the orthonormal tetrad formalism as presented by Ellis and MacCallum[25] and augmented by Wainwright.[26] We have assumed that our tetrad $\mathbf{e}_i$, $i=0,1,2,3$ has $\mathbf{e}_0=\mathbf{u}$ hypersurface orthogonal so that the vorticity tensor ω vanishes.

The commutators of $\mathbf{e}_i$ are:

$$\begin{aligned}
[\mathbf{e}_0,\mathbf{e}_1]&=\dot{u}^1\mathbf{e}_0-\theta_1\mathbf{e}_1-(\sigma_{12}-\Omega_3)\mathbf{e}_2-(\sigma_{13}+\Omega_2)\mathbf{e}_3,\\
[\mathbf{e}_0,\mathbf{e}_2]&=\dot{u}^2\mathbf{e}_0-(\sigma_{12}+\Omega_3)\mathbf{e}_1-\theta_2\mathbf{e}_2-(\sigma_{23}-\Omega_1)\mathbf{e}_3,\\
[\mathbf{e}_0,\mathbf{e}_3]&=\dot{u}^3\mathbf{e}_0-(\sigma_{13}-\Omega_2)\mathbf{e}_1-(\sigma_{23}+\Omega_1)\mathbf{e}_2-\theta_3\mathbf{e}_3,\\
[\mathbf{e}_1,\mathbf{e}_2]&=(n_{13}-a_2)\mathbf{e}_1+(n_{23}+a_1)\mathbf{e}_2+n_{33}\mathbf{e}_3,\\
[\mathbf{e}_2,\mathbf{e}_3]&=n_{11}\mathbf{e}_1+(n_{12}-a_3)\mathbf{e}_2+(n_{13}+a_2)\mathbf{e}_3,\\
[\mathbf{e}_3,\mathbf{e}_1]&=(n_{12}+a_3)\mathbf{e}_1+n_{22}\mathbf{e}_2+(n_{23}-a_1)\mathbf{e}_3.
\end{aligned}\tag{A1}$$

The Jacobi identities:

$$[\mathbf{X},[\mathbf{Y},\mathbf{Z}]]+[\mathbf{Y},[\mathbf{Z},\mathbf{X}]]+[\mathbf{Z},[\mathbf{X},\mathbf{Y}]]=0,$$

applied to the vector fields $\mathbf{e}_\alpha$, $\mathbf{e}_\beta$, and $\mathbf{e}_\lambda$ are

$$\begin{aligned}
&\partial_1 n_{11}+\partial_2 n_{12}+\partial_3 n_{31}+\partial_2 a_3-\partial_3 a_2-2(n_{11}a_1+n_{12}a_2+n_{31}a_3)=0,\\
&\partial_2 n_{22}+\partial_3 n_{23}+\partial_1 n_{12}+\partial_3 a_1-\partial_1 a_3-2(n_{22}a_2+n_{23}a_3+n_{12}a_1)=0,\\
&\partial_3 n_{33}+\partial_1 n_{31}+\partial_2 n_{23}+\partial_1 a_2-\partial_2 a_1-2(n_{33}a_3+n_{31}a_1+n_{23}a_2)=0.
\end{aligned}\tag{A2}$$

In addition, for a spatially homogeneous space-time, with a *canonical tetrad* some of the Jacobi identities applied to the vector fields $\mathbf{e}_0,\mathbf{e}_\alpha,\mathbf{e}_\beta$ yield:

$$\begin{aligned}
&\partial_0 n_{11}+(\theta_2+\theta_3-\theta_1)n_{11}=0,\\
&\partial_0 n_{22}+(\theta_3+\theta_1-\theta_2)n_{22}=0,\\
&\partial_0 n_{33}+(\theta_1+\theta_2-\theta_3)n_{33}=0.
\end{aligned}\tag{A3}$$

The 3-Ricci tensor $R^*_{\alpha\beta}$ on the hypersurfaces orthogonal to $\mathbf{e}_0$ are given by:

$$\begin{aligned}
R^*_{11}&=2\partial_1 a_1+\partial_2(a_2-n_{13})+\partial_3(a_3+n_{12})+2(a_2n_{13}-a_3n_{12})\\
&\quad+\tfrac{1}{2}(n_{11}^2-n_{22}^2-n_{33}^2)+n_{22}n_{33}-2n_{23}^2-2(a_1^2+a_2^2+a_3^2),\\
R^*_{22}&=2\partial_2 a_2+\partial_1(n_{23}+a_1)+\partial_3(a_3-n_{12})+2(a_3n_{12}-a_1n_{23})\\
&\quad+\tfrac{1}{2}(n_{22}^2-n_{11}^2-n_{33}^2)+n_{11}n_{33}-2n_{13}^2-2(a_1^2+a_2^2+a_3^2),\\
R^*_{33}&=2\partial_3 a_3+\partial_1(a_1-n_{23})+\partial_2(a_2+n_{13})+2(a_1n_{23}-a_2n_{13})\\
&\quad+\tfrac{1}{2}(n_{33}^2-n_{11}^2-n_{22}^2)+n_{11}n_{22}-2n_{12}^2-2(a_1^2+a_2^2+a_3^2),\\
R^*_{12}&=\tfrac{1}{2}\partial_1(a_2+n_{13})+\tfrac{1}{2}\partial_2(a_1-n_{23})-\tfrac{1}{2}\partial_3(n_{11}-n_{22})+a_3(n_{11}-n_{22})\\
&\quad+a_2n_{23}-a_1n_{13}+n_{12}(n_{11}+n_{22}-n_{33})+2n_{13}n_{23},\\
R^*_{13}&=\tfrac{1}{2}\partial_1(a_3-n_{12})+\tfrac{1}{2}\partial_3(a_1+n_{23})+\tfrac{1}{2}\partial_2(n_{11}-n_{33})+a_2(n_{33}-n_{11})\\
&\quad+a_1n_{12}-a_3n_{23}+n_{13}(n_{11}-n_{22}+n_{33})+2n_{12}n_{23}.\\
R^*_{23}&=\tfrac{1}{2}\partial_2(a_3+n_{12})+\tfrac{1}{2}\partial_3(a_2-n_{13})+\tfrac{1}{2}\partial_1(n_{33}-n_{22})+a_1(n_{22}-n_{33})\\
&\quad+a_3n_{13}-a_2n_{12}+n_{23}(-n_{11}+n_{22}+n_{33})+2n_{12}n_{13}.
\end{aligned}\tag{A4}$$

The Cotton–York tensor (1.2) is in general:

$$C^{*\alpha\beta}=2\eta^{\mu\nu(\alpha}[\partial_\nu-a_\nu]R^{*\beta)}{}_\mu+4\eta^{\lambda(\alpha}R^{*\beta)}{}_\lambda+\eta^{\alpha\mu\nu}\eta^{\beta\kappa\sigma}R^*_{\mu\kappa}n_{\nu\sigma}-R^*n_{\alpha\beta}-n_{\mu\nu}R^{*\mu\nu}\delta_{\alpha\beta}.$$

However, in a 3-Ricci eigenframe ($R^*_{12}=R^*_{13}=R^*_{23}=0$) it simplifies to:

$$\begin{aligned}
C^*_{11}&=2n_{11}R^*_{11}+(-n_{11}-n_{22}+n_{33})R^*_{22}+(-n_{11}+n_{22}-n_{33})R^*_{33},\\
C^*_{22}&=(-n_{11}-n_{22}+n_{33})R^*_{11}+2n_{22}R^*_{22}+(n_{11}-n_{22}-n_{33})R^*_{33},\\
C^*_{33}&=(-n_{11}+n_{22}-n_{33})R^*_{11}+(n_{11}-n_{22}-n_{33})R^*_{22}+2n_{33}R^*_{33},\\
C^*_{12}&=-\partial_3(R^*_{11}-R^*_{22})+n_{12}(R^*_{11}+R^*_{22}-2R^*_{33})+a_3(R^*_{11}-R^*_{22}),\\
C^*_{13}&=-\partial_2(R^*_{33}-R^*_{11})+n_{13}(R^*_{11}-2R^*_{22}+R^*_{33})+a_2(R^*_{33}-R^*_{11}),\\
C^*_{23}&=-\partial_1(R^*_{22}-R^*_{33})+n_{23}(-2R^*_{11}+R^*_{22}+R^*_{33})+a_1(R^*_{22}-R^*_{33}).
\end{aligned}\tag{A5}$$

APPENDIX B

In this Appendix, we prove the following lemma:

Lemma: Let V_3 be a three-dimensional Riemannian manifold. Assume that $R_{\alpha\beta}$ (or equivalently $S_{\alpha\beta}$) has two equal eigenvalues and that $S_{\alpha\beta}$ is not identically zero. Assume that the spatial gradient of $S^2=\frac{1}{2}S_{\alpha\beta}S^{\alpha\beta}$ vanishes ($S_{,\mu}=0$) and

$$C_{\alpha\beta}=\lambda S_{\alpha\beta},\tag{B1}$$

with either:

(i) λ constant, $\lambda\neq 0$,

or (ii) $\lambda=0$ and $R_{\alpha\beta}$ has zero as its third eigenvalue.

Then there exists a Ricci eigenframe ($R_{12}=R_{13}=R_{23}=0$) in which

$$a_1=n_{13}-a_2=n_{12}+a_3=n_{23}=n_{22}=n_{33}=0,\tag{B2}$$

and in which

$$\lambda=3n_{11}.\tag{B3}$$

Proof: We begin by choosing a Ricci eigenframe in which

$$R_{12}=R_{13}=R_{23}=0,\quad R_{11}\neq R_{22}=R_{33}.\tag{B4}$$

The remaining tetrad freedom is a rotation:

$$\begin{aligned}
\hat{\mathbf{e}}_1&=\mathbf{e}_1,\\
\hat{\mathbf{e}}_2&=\mathbf{e}_2\cos\phi+\mathbf{e}_3\sin\phi,\\
\hat{\mathbf{e}}_3&=-\mathbf{e}_2\sin\phi+\mathbf{e}_3\cos\phi,
\end{aligned}\tag{B5}$$

and the following quantities are invariant under the rotation:

$$I_1 = a_1\,,$$
$$I_2 = (n_{13} - a_2)^2 + (n_{12} + a_3)^2\,, \tag{B6}$$
$$I_3 = n_{11}\,,$$
$$I_4 = (n_{22} - n_{33})^2 + 4n_{23}{}^2\,,$$

(cf. Collins and Szafron[27] or Ellis[28]). Conditions (B1) and (B4) imply

$$C_{23} = C_{22} - C_{33} = 0\,,$$

which, using (A5), becomes

$$n_{23} = n_{22} - n_{33} = 0\,. \tag{B7}$$

Definition (2.2) together with (B4) implies

$$S_{11} = \tfrac{2}{3}(R_{11} - R_{22}),\quad S_{22} = S_{33} = -\tfrac{1}{3}(R_{11} - R_{22})\,, \tag{B8}$$

so that the first three equations from (A5) now become

$$C_{11} = \lambda S_{11}, C_{22} = C_{33} = \lambda S_{22} = \lambda S_{33}\,, \tag{B9}$$

where

$$\lambda = 3n_{11}\,. \tag{B10}$$

Equations (B1) and (B4) imply

$$C_{12} = C_{13} = 0\,, \tag{B11}$$

while the conditions $\vartheta_2 S = \vartheta_3 S = 0$ together with (B8) yields

$$\vartheta_2(R_{11} - R_{22}) = \vartheta_3(R_{11} - R_{22}) = 0\,. \tag{B12}$$

Applying (B11) and (B12) to the fourth and fifth equations of (A5), and recalling (B4), we find

$$n_{13} - a_2 = n_{12} + a_3 = 0\,. \tag{B13}$$

We now use a tetrad rotation (B5) to set

$$n_{22} = 0\,, \tag{B14}$$

and Eq. (B7) provides

$$n_{33} = 0\,. \tag{B15}$$

From the partial list of invariants (B6), we see that (B7), (B10), and (B13) are preserved. It remains only to show that $a_1 = 0$.

The first Jacobi identity (A2) is now

$$n_{11}a_1 = 0\,. \tag{B16}$$

In case (i), $\lambda \neq 0$ so (B10) and (B16) imply

$$a_1 = 0\,, \tag{B17}$$

and the lemma is proved. In case (ii), $\lambda = 0$ so that (B10) implies

$$n_{11} = 0\,, \tag{B18}$$

and $R_{11} = 0$ converts the first equation of (A4) to

$$\vartheta_1 a_1 = a_1{}^2\,. \tag{B19}$$

Adding and subtracting the second and third equations from (A2) with twice the fourth and fifth equations from (A4), now yields

$$\vartheta_3 a_1 = \vartheta_2 a_1 = 0;\quad \vartheta_1 a_3 = a_1 a_3,\quad \vartheta_1 a_2 = a_1 a_2\,. \tag{B20}$$

Applying ϑ_1 to the second equation from (A4) and using (B18), (B19), (B20), (B13), (B14), (B15), and the commutation relations (A2) yields

$$\vartheta_1 R_{22} = 2a_1 R_{22}\,. \tag{B21}$$

The conditions $\vartheta_1 S = R_{11} = 0$ together with (B8) imply

$$\vartheta_1 R_{22} = 0\,. \tag{B22}$$

Equations (B21), (B22), (B4) and the fact that $R_{11} = 0$ yield

$$a_1 = 0\,, \tag{B23}$$

so the lemma is proved.

[1]G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[2]L. P. Eisenhart, Riemannian Geometry (Princeton U. P., Princeton, N. J., 1926).
[3]P. Szekeres, Commun. Math. Phys. **41** 55 (1975).
[4]W. B. Bonnor, A. H. Sulaiman and N. Tomimura, Gen. Rel. Grav. **8**, 549 (1977).
[5]D. A. Szafron and C. B. Collins, J. Math. Phys. **20**, 2354 (1979).
[6]B. K. Berger, D. M. Eardley, and D. W. Olson, Phys. Rev. D **16**, 3086 (1977).
[7]C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347 (1979).
[8]See Ref. 5.
[9]C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2362 (1979).
[10]A. Spero and D. A. Szafron, J. Math. Phys. **19**, 1536 (1978).
[11]J. Wainwright, J. Phys. A: Math. Gen. **12**, 2015 (1979).
[12]J. W. York, Phys. Rev. Lett. **26**, 1656 (1971).
[13]See Ref. 11.
[14]See Ref. 1, p. 130.
[15]See Ref. 2, p. 88.
[16]See Ref. 2, p. 246.
[17]See Ref. 1.
[18]E. Cartan, *Lecons sur la Geometrie des Espaces de Riemann* (Gauthier–Villars, Paris, 1928), see p. 304.
[19]See Ref. 2, p. 88.
[20]See Ref. 2, p. 86.
[21]These space-times admit a multiply transitive G_4 with three-dimensional spacelike orbits, but no simply transitive subgroup G_3. Since they possess a G_4 they automatically have a $G_4^\star$.
[22]See Ref. 1, p. 130.
[23]They have a tetrad of the form of Eqs. (7.1) of Ref. 1.
[24]See Ref. 1, p. 114.
[25]See Ref. 1.
[26]See Ref. 11.
[27]See Ref. 7, p. 2352.
[28]G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967), see the Appendix.

Fiber bundles, superspace, and the gravitational field

D. K. Ross
Physics Department, Iowa State University, Ames, Iowa 50011

(Received 22 May 1980; accepted for publication 15 September 1980)

We develop a new theory of the pure gravitational field by treating superspace as a fiber bundle with space–time as the base space, and Fermi space, with anticommuting Majorana spinor coordinate θ^i, as the typical fiber. In the fiber bundle geometry, a spin-3/2 field arises automatically. It comes in through the commutation relations of the basis vectors in the horizontal lift basis, or through the metric in the local direct-product basis. The Lagrangian is taken to be the scalar curvature of the fiber bundle, in analogy with general relativity with no source terms. The theory is self sourced and the spin-3/2 field and usual spin-2 field appear as gauge fields. The theory is also completely basis invariant. The field equations correctly describe a spin-3/2 field coupled to general relativity, with the correct "energy–momentum tensor" of the spin-3/2 field appearing automatically. We thus end up with a very simple, geometrical theory which contains far fewer fields that the geometrical work of Arnowitt and Nath while keeping their elegance of formulation. The resulting field equations are similar to those of simple supergravity with only a spin-3/2 field appearing in addition to the spin-2 Einstein field, however, supersymmetry invariance seems to play little or no direct role in the present theory.

PACS numbers: 04.50. + h, 11.30.Pb

I. INTRODUCTION

Supersymmetry[1,2] and its possible role in gravitation theory has received considerable recent attention. Salam and Strathdee[3] introduced the idea of using linear transformations in an eight-dimensional superspace $z^A = \{x^\mu, \theta^i\}$ to describe the usual supersymmetry transformation. Here x^μ are the usual Bose space–time coordinates and θ^i are Majorana spinor anticommuting Fermi coordinates. Arnowitt and Nath[4] then extended this to arbitrary coordinate transformations in superspace which leave the line element

$$ds^2 = dz^A\, g_{AB}\, dz^B \tag{1}$$

invariant. In this elegant geometrical theory, the Lagrangian is taken to be the scalar curvature of superspace and the single gauge superfield g_{AB} contains all the physical fields in the theory. To accommodate electricity and magnetism, or other internal symmetry groups, the Fermi sector is enlarged by adding an internal symmetry index to θ^i to give θ^{iq}. Recently the arbitrary n-point Green's functions of spontaneously broken gauge supersymmetry have been shown[5] to be ultraviolet finite to arbitrary loop order for $N \geqslant 2$, where $4N$ is the number of Fermi coordinates. This augurs well for a renormalizable theory of gravitation. The chief disadvantage of this theory is that the superfield g_{AB} contains a large number of higher order ordinary fields even in the case of pure gravity. If electricity and magnetism are put in, not one, but a large number of additional fields appear. One hopes that nature has not been so extravagant.

An alaternative approach to a renormalizable theory of gravitation has been the supergravity work of Freedman, van Nieuwenhuizen, and Ferrara[6] and of Deser and Zumino.[7] This has been extended to larger gauge groups and elaborated by numerous authors.[8] The simple version of this theory (hereinafter denoted FvFDZ) constructs a Lagrangian which is explicitly invariant under a supersymmetry transformation of the fields. The action describes massless spin-2 and spin-3/2 Majorana fields and is the sum of the Einstein and Rarita–Schwinger[9] actions. This theory seems to be 1- and 2-loop finite[10] but 3-loop divergent.[11] Thus the spin-3/2 field seems to improve greatly the renormalizability of the gravitational field but problems remain. This approach lacks the geometrical elegance of the Arnowitt and Nath theory but has the great advantage of containing far fewer fields, so that higher order fields do not have to be neglected whenever an explicit calculation is made.

In this paper we propose a new fiber bundle theory of the gravitational field which is similar to the Nath and Arnowitt[4,5,12] theory in its geometrical elegance but which contains far fewer physical fields. In addition to the usual spin-2 Einstein field, a spin-3/2 field appears quite naturally, so that we end up with something similar to simple supergravity. We will be concerned in this paper only with a pure gravitational field. Coupling to matter fields will be taken up in a subsequent paper.

In Sec. II below, we discuss the fiber bundle geometry. The field equations are worked out in Sec. III, and we conclude with a discussion of results in Sec. IV.

II. BASIC GEOMETRY

Rather than treat eight-dimensional superspace as a general Riemannian manifold as Arnowitt and Nath do, we will take it to have the structure of a fiber bundle **P**, with real space–time M as the base manifold and the four-dimensional Fermi space F, with anticommuting Majorana spinor coordinates θ^i, as the typical fiber. This approach is much closer to how a gauge group is conventionally associated with space–time and has the enormous benefit of introducing far fewer physical fields as we shall see. We are essentially taking superspace to be a kind of generalized topological product of F and M.

Cho[13] has written a very interesting paper unifying gravitation with a nonabelian gauge group. We will use this paper, appropriately modified, in the development which follows.

 022-2488/81/030549-08$1.00

We choose a coordinate basis[14] $\xi_\mu = \partial_\mu$ for the base manifold M with commutation relations

$$[\xi_\mu, \xi_\nu] = [\partial_\mu, \partial_\nu] = 0 . \tag{2}$$

We also assume that M is a metric manifold with metric $g_{\mu\nu}$. This $g_{\mu\nu}$ will depend on x^μ but not on the Fermi coordinate θ^i. Similarly, we choose a coordinate basis for the Fermi space F with the basis vectors satisfying anticommutators,

$$\{\xi_i, \xi_j\} = \{\partial_i, \partial_j\} = 0 . \tag{3}$$

F can also be taken to be a metric manifold with metric g_{ij} which depends on θ^i, but contains no space–time dependent physical fields. g_{ij} can be expanded in general in a finite power series in θ^i with contributions up through four powers of the θ's. For the structure of the fiber bundle $\mathbf{P}$,[15] we consider the fibration $\lambda = (\mathbf{P}, M, \Pi)$ where the base space M and the bundle space $\mathbf{P}$ are smooth manifolds and Π is a C^∞ mapping of $\mathbf{P}$ into M. Π is surjective and satisfies the condition of local triviality: for any $x \in M$ there exists an open neighborhood U of x, a manifold F, and an isomorphism (diffeomorphism) ϕ of $\Pi^{-1}(U)$ onto $U \times F$ such that $\Pi(\phi^{-1}(x, \theta)) = x$ for all $x \in U$ and $\theta \in F$. Π is called the projection of the fibration and for each $x \in M$, the inverse image $P_x = \Pi^{-1}(x)$ is a closed submanifold of $\mathbf{P}$ called the fiber over x. It represents the local gauge degree of freedom at x. Local triviality implies that all $P_{x'}$ are diffeomorphic to P_x for all $x' \in U$, a neighborhood of x. The fibration can be viewed as a bundle of fibers or fiber bundle, especially since all the fibers are diffeomorphic to the same manifold F, the typical fiber.

To go further, we now need a connection Γ in $\mathbf{P}$. We first note that the basis ξ_i of F can be mapped naturally into the fiber bundle.[13] The resulting fields in the bundle will be denoted by ξ_i^*. These are tangent to the fiber space and form a subspace, called the vertical subspace V_p of the tangent space $T_p(\mathbf{P})$ to $\mathbf{P}$. A connection[13,14] is then taken to be a choice of a subspace of $T_p(\mathbf{P})$ called the horizontal subspace H_p at each point of $\mathbf{P}$ such that

(a) The tangent space $T_p(\mathbf{P})$ is the direct sum of V_p and H_p.

(b) For all $\theta \in F$ and $p \in \mathbf{P}$, $H_{p\theta} = R_\theta H_p$, where $R_\theta : (p, \theta) \in \mathbf{P} \times F \to p\theta \in \mathbf{P}$ with $p(\theta\phi) = (p\theta)\phi$ for all $\theta, \phi \in F$, $p \in \mathbf{P}$.

(c) H_p is smooth on $\mathbf{P}$.

We can now define the horizontal lift of any vector A at $x \in M$. The horizontal lift of A will be a unique vector $\hat{A} \in H_p$ at each point of the fiber $\Pi^{-1}(x)$ such that $\Pi(\hat{A}) = A$.[14] We can horizontally lift our basis vectors ξ_μ in M and denote these by $\hat{\xi}_\mu$. These $\hat{\xi}_\mu$ along with the ξ_i^* defined above, form a natural basis for the fiber bundle called the horizontal lift basis with commutation relations in the bundle

$$\begin{aligned} &\{\xi_i^*, \xi_j^*\} = 0, \\ &[\xi_i^*, \hat{\xi}_\mu] = 0 , \\ &[\hat{\xi}_\mu, \hat{\xi}_\nu] = -F^i_{\mu\nu}\, \xi_i^* . \end{aligned} \tag{4}$$

[We will use Greek letters μ, ν, ρ, etc. to denote Bose indices, lower case Latin indices i, j, k, etc. to denote Fermi indices, and upper case Latin letters A, B, C, etc. to denote a fiber bundle index which ranges over both Bose and Fermi indices. The latter takes on eight values. We can keep track of of anticommuting Fermi variables by associating a Grassman parity a, b, etc. with each fiber bundle index A, B, etc., as in the work of Arnowitt and Nath.[4] $a = 1$ if A is Fermi and $a = 0$ if A is Bose. If we change the order of two quantities H_{AB} and $Q^C{}_{DE}$, for example, we have $H_{AB} Q^C{}_{DE} = (-1)^{(a+b)(c+d+e)} Q^C{}_{DE} H_{AB}$.] We will use this horizontal lift basis extensively below because of the simplications it makes in actual calculations. Other bases are also possible, of course, and the theory is completely basis-invariant as we shall see. One other basis in particular is useful, especially in understanding gauge covariant derivatives. This is the local direct-product basis, which we discuss below.

The first two commutation relations in (4) follow immediately from the definition of the respective basis vectors. The third relation is the statement that $[\hat{\xi}_\mu, \hat{\xi}_\nu]$ is vertical. This follows since the projection with Π of any horizontal component into M must give $[\xi_\mu, \xi_\nu]$ which vanishes from (2). This third commutation relation is quite important for us, since it brings in $F^i_{\mu\nu}$, which will be related to the spin-3/2 field. We see that the $\hat{\xi}_\mu$ form a noncoordinate basis. In the local direct-product basis, as we shall see, the corresponding commutator vanishes. In this latter basis, $F^i_{\mu\nu}$ enters the theory completely through the metric γ_{AB} of the fiber bundle [see (6'), (23), and the definition of B^i_μ following (20)]. In the horizontal lift basis, the metric is simple and $F^i_{\mu\nu}$ enters the theory through the commutation relations of the basis vectors as in (4).

Let us now turn to the metric γ_{AB} in the fiber bundle. We shall require γ_{AB} to be compatible with the metrics $g_{\mu\nu}$ and g_{ik} on M and F by requiring[13,16]

$$\begin{aligned} &\gamma_{AB}\, \hat{\xi}_\mu^{\,A}\, \hat{\xi}_\nu^{\,B} = g_{\mu\nu} , \\ &\gamma_{AB}\, \xi_i^{*A}\, \xi_k^{*B} = g_{ik} , \\ &\gamma_{AB}\, \hat{\xi}_\mu^{\,A}\, \xi_k^{*B} = 0 , \end{aligned} \tag{5}$$

where this last condition is that the horizontal and vertical subspaces be orthogonal. This is a basis-invariant definition. In the horizontal lift basis, this metric is particularly simple:

$$\gamma_{AB} = \begin{pmatrix} g_{\mu\nu} & 0 \\ 0 & g_{ik} \end{pmatrix}, \qquad \gamma^{AB} = \begin{pmatrix} g^{\mu\nu} & 0 \\ 0 & g^{ik} \end{pmatrix}. \tag{6}$$

γ^{AB} is not in this simple diagonal form in any other basis. In the local direct-product basis, we have for comparison that

$$\gamma_{AB} = \begin{pmatrix} g_{\mu\nu} + g_{ik}\, B^i_\mu\, B^k_\nu & B^i_\mu\, g_{ik} \\ g_{ik}\, B^k_\nu & g_{ik} \end{pmatrix}, \tag{6'}$$

where we used (5) and (21). This is much more difficult to use in calculations than (6).

The horizontal lift basis gives rise to one further simplification. Note that $g_{\mu\nu,i} = 0$ and $g_{ik,\mu} = 0$, where a comma denotes a partial derivative. These follow from our definitions of the M and F manifolds and of their respective metrics. We can take these over into the fiber bundle by writing $\gamma_{\mu\nu,i} = 0$ and $\gamma_{ik,\hat{\mu}} = 0$ in the horizontal lift basis. These are now directional derivatives along ξ_i^* and $\hat{\xi}_\mu$ respectively. They follow from the isomorphism of ξ_i and ξ^*_i and from

the definition of the horizontal lift, and are basis-invariant statements. If we work in some other barred basis, $\gamma_{il,\hat{\mu}} = 0$ will still hold, but in terms of quantities in that basis $\gamma_{il,\bar{\mu}} \neq 0$. The horizontal lift basis thus simplifies the fiber bundle metric in its form and in its dependence on x_μ and θ_i.

At this point we have two quantities which may serve as physical fields, $g_{\mu\nu}(x)$ and $F^i_{\mu\nu}(x,\theta)$. Note that $g_{\mu\nu}(x)$ is not a superfield but a simple function of x_μ. $F^i_{\mu\nu}(x,\theta)$, however, does look like a superfield and may contain large numbers of physical fields when expanded in a power series in θ_i. To determine the fiber space dependence of $F^i_{\mu\nu}$, let us look at the Jacobi identities. One of these is

$$[\xi^*_i,[\hat{\xi}_\mu,\hat{\xi}_\nu]]+[\hat{\xi}_\mu,[\hat{\xi}_\nu,\xi^*_i]]+[\hat{\xi}_\nu,[\xi^*_i,\hat{\xi}_\mu]]=0. \tag{7}$$

Using (4) gives

$$[\xi^*_i,-F^j_{\mu\nu}\xi^*_j]=0, \tag{8}$$

which becomes

$$-(\partial^*_i F^j_{\mu\nu})\xi^*_j+F^j_{\mu\nu}\{\xi^*_i,\xi^*_j\}=0, \tag{9}$$

and finally

$$-\partial^*_i F^j_{\mu\nu}=0 \quad \text{or} \quad F^j_{\mu\nu,i}=0. \tag{10}$$

Thus $F^j_{\mu\nu}$ is a simple function of x_μ and we end up with just $g_{\mu\nu}(x)$ and $F^j_{\mu\nu}(x)$ for physical fields. This is in sharp contrast with the work of Arnowitt and Nath where the superfield $g_{AB}(x,\theta)$ contains a large number of physical fields.

A second Jacobi identity is

$$[\hat{\xi}_\rho,[\hat{\xi}_\mu,\hat{\xi}_\nu]]+[\hat{\xi}_\mu,[\hat{\xi}_\nu,\hat{\xi}_\rho]]+[\hat{\xi}_\nu,[\hat{\xi}_\rho,\hat{\xi}_\mu]]=0. \tag{11}$$

This becomes

$$\hat{\xi}_\rho F^i_{\mu\nu}+\hat{\xi}_\mu F^i_{\nu\rho}+\hat{\xi}_\nu F^i_{\rho\mu}=0, \tag{12}$$

or

$$F^i_{\mu\nu,\rho}+F^i_{\nu\rho,\mu}+F^i_{\rho\mu,\nu}=0, \tag{13}$$

where we must remember that the partial derivatives refer to the horizontal lift basis. To work out the remaining two Jacobi identities, we need $[\xi^*_i,\xi^*_j]$, which is not given in (4). This can be written

$$[\xi^*_i,\xi^*_k]=\nabla_{\xi^*_i}\xi^*_k-\nabla_{\xi^*_k}\xi^*_i, \tag{14}$$

where $\nabla_{\xi^*_n}$ is the covariant derivative in the ξ_i direction, and we have assumed zero torsion. This involves the Christoffel symbol Γ^l_{ik} in the Fermi sector. Thus from (36) below we have

$$[\xi^*_i,\xi^*_k]=2\Gamma^l_{ik}\xi^*_l. \tag{15}$$

The Jacobi identity involving ξ^*_i, ξ^*_j, and ξ^*_k then becomes

$$\Gamma^l_{jk,i}+\Gamma^l_{ki,j}+\Gamma^l_{ij,k}=0. \tag{16}$$

Finally, the Jacobi identity involving ξ^*_i, ξ^*_j, and $\hat{\xi}_\mu$ becomes

$$\hat{\xi}_\mu\Gamma^k_{ij}=0 \quad \text{or} \quad \Gamma^l_{ij,\mu}=0. \tag{17}$$

This last equation is satisfied identically since $g_{ij,\mu}=0$ in the horizontal lift basis and Γ^i_{jk} is a function of g_{ik} only, from (37) below.

We would now like to introduce potentials. To do this we need to define connection forms and cross sections of the fiber bundle. Given a connection Γ, we can always define a connection form ω such that[13]

$$\omega^i(\xi^*_k)=\omega^i_a\xi^{*a}_k=\delta^i_k,$$
$$\omega^i(\hat{\xi}_\mu)=\omega^i_a\hat{\xi}^a_\mu=0, \tag{18}$$

where $\omega\equiv\omega^i\xi_i$ and ω^i are real valued 1-forms. A cross section $\sigma(x,\theta)$ is then a submanifold of $\mathbf{P}$ which is diffeomorphic to an open subset U of M with $\Pi(\sigma(x,\theta))=x$. Let $\sigma\cdot\xi_\mu$ be a vector in the tangent space $T_{\sigma(x,\theta)}(\mathbf{P})$ to $\mathbf{P}$ at $\sigma(x,\theta)$ induced by $\sigma: x\in U\to\sigma(x,\theta)\in\Pi^{-1}(U)$. We can then define a connection form $A^{(\sigma)}$ on U [or $\sigma(x,\theta)$] by[13]

$$A^{(\sigma)}(\xi_\mu)\equiv A^{(\sigma)i}_\mu\xi_i=\omega(\sigma\cdot\xi_\mu)=\omega^i(\sigma\cdot\xi_\mu)\xi_i. \tag{19}$$

The $A^{(\sigma)i}_\mu$ are then identified as our vector potentials. They depend on the choice of the cross section $\sigma(x,\theta)$. We can define the trivial cross section $\sigma_\theta(x)$ as the set of points $p=(x,\theta)$ with fixed θ. For each θ there corresponds such a cross section. We can also introduce the local direct product basis $\bar{\xi}_\mu$ and $\bar{\xi}_i=\xi^*_i$. For all θ the $\bar{\xi}_\mu$ form a basis of the tangent space of the trivial cross sections $\sigma_\theta(x)$.[13] The commutation relations for this basis are

$$\{\bar{\xi}_i,\bar{\xi}_j\}=0,$$
$$[\bar{\xi}_i,\bar{\xi}_\mu]=0, \tag{20}$$
$$[\bar{\xi}_\mu,\bar{\xi}_\nu]=0.$$

We can define potentials in this basis as $B^i_\mu(x,\theta)\equiv A^{(\sigma_\theta)i}_\mu=\omega^i(\bar{\xi}_\mu(x,\theta))$, and using (18) write

$$\hat{\xi}_\mu=\bar{\xi}_\mu-\bar{\xi}_i\,\omega^i(\bar{\xi}_\mu)=\bar{\xi}_\mu-\bar{\xi}_i B^i_\mu(x,\theta)=D_\mu. \tag{21}$$

Thus $\hat{\xi}_\mu$ is the usual gauge covariant derivative if one writes it out in the more usual local direct-product basis rather than working in the horizontal lift basis as we have been doing. We will continue to work in the horizontal lift basis, where this derivative will often be denoted by a comma, but this shows the role of $\hat{\xi}_\mu$ as a gauge covariant derivative. Thus a major advantage of the horizontal lift basis is that the derivatives are automatically gauge covariant.

Now $F^i_{\mu\nu}$ can be written in terms of $A^{(\sigma)i}_\mu$ by writing (4) as

$$D_\mu\hat{\xi}_\nu-D_\nu\hat{\xi}_\mu=-F^i_{\mu\nu}\xi^*_i. \tag{22}$$

We can map this into the tangent space $T_{\sigma(x)}(\mathbf{P})$ using the mapping σ as above and then operate with ω^i using (18) to get

$$F^i_{\mu\nu}=D_\nu A^{(\sigma)i}_\mu-D_\mu A^{(\sigma)i}_\nu. \tag{23}$$

Either $F^i_{\mu\nu}$ or $A^{(\sigma)i}_\mu$ can be used to describe the spin-3/2 field. We will work primarily with the gauge invariant quantity $F^i_{\mu\nu}$ in the following, although the $A^{(\sigma)i}_\mu$ is more closely related to the ψ_μ spinor fields of supergravity.[6,7] Note that $F^i_{\mu\nu}$ is essentially the curvature 2-form associated with the connection 1-form ω.[14]

III. FIELD EQUATIONS

The fiber bundle space $\mathbf{P}$ has been organized into an eight-dimensional Riemann space above. We can calculate the curvature of the bundle and use the eight-dimensional extension of the Hilbert action of general relativity for our action principle. Before we do this, we need to define covar-

iant derivatives in terms of Christoffel symbols Γ^{A}_{BC} and write out the curvature tensor $R_{ABC}{}^{D}$ in terms of Γ^{A}_{BC} and $F^{i}_{\mu\nu}$. We have the double complication of having anticommuting variables and also a noncoordinate basis (the price we pay for using the horizontal lift basis).

The covariant derivative of a contravariant vector V^{A} can be defined as[4]

$$V^{A}{}_{;B} \equiv V^{A}{}_{,B} + V^{C}\Gamma^{A}_{CB} . \tag{24}$$

(In the horizontal lift basis, this derivative is both gauge covariant and generally covariant.) Unfortunately we then have two types of covariant vectors, $U^{+}_{B} = V^{A} g_{AB}$ and $U_{B} = g_{BA} V^{A}$ with slightly different covariant derivatives

$$U_{A;B} = U_{A,B} - (-1)^{d(d+a)} U_{D} \Gamma^{D}_{AB} \tag{25}$$

and

$$U^{+}_{A;B} = U^{+}_{A,B} - (-1)^{a(a+d)} U^{+}_{D} \Gamma^{D}_{AB} . \tag{26}$$

The behavior of higher-rank tensors can be seen by considering products of different kinds of vectors. In particular, the covariant derivative of the metric γ_{AB} in the fiber bundle is

$$\begin{aligned}\gamma_{AB;C} = {} & \gamma_{AB,C} - (-1)^{b(b+d)} \gamma_{AD} \Gamma^{D}_{BC} \\ & - (-1)^{(d+b)(a+d)} \gamma_{DB} \Gamma^{D}_{AC} = 0 ,\end{aligned} \tag{27}$$

since it has one lower index of each type.[4] We can now solve (27) for Γ^{D}_{BC} remembering that we are interested in general in a noncoordinate basis (4). In Arnowitt and Nath,[4] we have the following symmetry properties,

$$\gamma_{BC} = (-1)^{b+c+bc} \gamma_{CB} \tag{28}$$

and

$$\Gamma^{C}_{BD} = (-1)^{b+d+bd+(b+d)c} \Gamma^{C}_{DB} . \tag{29}$$

(28) will hold for us but (29) must be modified because of our noncoordinate basis to give

$$(-1)^{b+d+bd+(b+d)c} \Gamma^{C}_{DB} - \Gamma^{C}_{BD} \equiv H^{C}_{DB} , \tag{30}$$

where H^{C}_{DB} must be determined in terms of $F^{i}_{\mu\nu}$ from the commutation relations (4). Solving (27) for Γ^{D}_{BC}, watching anticommuting variables, and using (30), gives after a laborious calculation that

$$\begin{aligned}\Gamma^{E}_{AB} = {} & (-1)^{be} \tfrac{1}{2}[(-1)^{bd} \gamma_{AD,B} \\ & + (-1)^{a+b+ab+ad} \gamma_{BD,A} - \gamma_{AB,D}] \gamma^{DE} \\ & - \tfrac{1}{2} H^{E}_{BA} - (-1)^{be+(b+c)(a+c)} \tfrac{1}{2} \gamma_{CB} H^{C}_{DA} \gamma^{DE} \\ & - (-1)^{b(b+c+e)} \tfrac{1}{2} \gamma_{AC} H^{C}_{DB} \gamma^{DE} .\end{aligned} \tag{31}$$

This reduces to the Arnowitt and Nath[4] form if $H^{E}_{BA} = 0$ and to the usual form[7] for $\Gamma^{\alpha}_{\beta\gamma}$ in a noncoordinate basis if all the indices are Bose indices.

We can now use

$$[\xi_{A}, \xi_{B}] = \nabla_{\xi_{A}} \xi_{B} - \nabla_{\xi_{A}} \xi_{A} , \tag{32}$$

where $\xi_{A} = (\hat{\xi}_{\mu}, \xi^{*}_{i})$ are basis vectors obeying (4), to get H^{A}_{BC} in terms of $F^{i}_{\mu\nu}$. We have assumed zero torsion. We have

$$\nabla_{\xi_{A}} \xi_{B} = \Gamma^{C}_{BA} \xi_{C} (-1)^{a+a(b+c)} \tag{33}$$

for the covariant derivative of the basis vector ξ_{B}. Thus, using (30),

$$\begin{aligned}[\xi_{A}, \xi_{B}] = {} & \Gamma^{C}_{BA} \xi_{C} (-1)^{a+a(b+c)} \\ & - \Gamma^{C}_{AB} \xi_{C} (-1)^{b+b(a+c)} \\ = {} & [(-1)^{b+a+ba+(b+a)c} \Gamma^{C}_{AB} - H^{C}_{AB}] \\ & \times \xi_{C} (-1)^{a+a(b+c)} \\ & - \Gamma^{C}_{AB} \xi_{C} (-1)^{b+b(a+c)} .\end{aligned} \tag{34}$$

(34) and hence (30) are consistent with the commutation relations (4) if

$$H^{C}_{AB} = \begin{cases} F^{i}_{\alpha\beta}, & \text{if } C = i,\ A = \alpha,\ B = \beta \\ 0, & \text{otherwise} \end{cases} . \tag{35}$$

We also note that

$$\begin{aligned}[\xi^{*}_{i}, \xi^{*}_{j}] = {} & -2\Gamma^{C}_{ij} \xi_{C} (-1)^{c} \\ = {} & 2\Gamma^{l}_{ij} \xi^{*}_{l} ,\end{aligned} \tag{36}$$

since $\Gamma^{\alpha}_{ij} = 0$ from (31) and our work below. This was used in (15) above.

Using (6) to write γ_{AB} in terms of $g_{\alpha\beta}$ and g_{ij} and using $\gamma_{\mu\nu,i} = 0$ and $\gamma_{ik,\mu} = 0$ in the horizontal lift basis, we can write out the various components of Γ^{E}_{AB} in (31) as

$$\begin{aligned}& \Gamma^{\epsilon}_{\alpha\beta} = \tfrac{1}{2}[g_{\alpha\delta,\beta} + g_{\beta\delta,\alpha} - g_{\alpha\beta,\delta}] g^{\delta\epsilon} , \\ & \Gamma^{i}_{\alpha\beta} = \tfrac{1}{2} F^{i}_{\alpha\beta} , \\ & \Gamma^{\beta}_{\alpha j} = -\Gamma^{\beta}_{j\alpha} = -\tfrac{1}{2} g_{ij} F^{i}_{\delta\alpha} g^{\delta\beta} , \\ & \Gamma^{i}_{\alpha j} = \Gamma^{i}_{j\alpha} = 0 , \\ & \Gamma^{k}_{ij} = -\tfrac{1}{2}[-g_{il,j} + g_{jl,i} - g_{ij,l}] g^{lk} , \\ & \Gamma^{\epsilon}_{ij} = 0 .\end{aligned} \tag{37}$$

Before we can write down field equations, we still must define the curvature tensor $R_{ABC}{}^{D}$ in terms of Γ^{A}_{BC} and $F^{i}_{\mu\nu}$. We can get this by parallel transport of a vector around a closed loop consisting of five sides $dx \equiv dx^{A}\xi_{A}$, $dx' \equiv dx'^{A}\xi_{A}$, $[dx, dx']$, $-dx$, and $-dx'$, where dx^{A} are components of a small vector in our basis ξ_{A}.[17] We must again be careful of anticommuting variables and take into account that we have a noncoordinate basis so that $[dx, dx'] \neq 0$. The result is

$$\begin{aligned}R^{A}{}_{FGB} = {} & \Gamma^{A}_{FG,B} (-1)^{bg} - \Gamma^{A}_{FB,G} \\ & + \Gamma^{C}_{FG} \Gamma^{A}_{CB} (-1)^{g(a+b+c)} \\ & - \Gamma^{C}_{FB} \Gamma^{A}_{CG} (-1)^{b(a+c)} \\ & - \Gamma^{A}_{FE} H^{E}_{GB} (-1)^{g+b+bg+a(b+g)} ,\end{aligned} \tag{38}$$

where H^{E}_{GB} is given in (35). This agrees with the result of Misner, Thorne, and Wheeler[17] when we have all Bose coordinates and with the result of Arnowitt and Nath[4] when $H^{E}_{GB} = 0$ (coordinate basis). (I am using the sign conventions of Ref. 4.)

We can now define

$$R_{AB} \equiv (-1)^{c} R^{C}{}_{ABC} \tag{39}$$

and the curvature scalar

$$R \equiv (-1)^{a} \gamma^{AB} R_{BA} . \tag{40}$$

Our variational principle for the field equations will then be taken to be

$$\delta \int (-\gamma)^{1/2} R d^{8}z = 0 , \tag{41}$$

where we vary with respect to the metric γ^{AB} and

$\gamma \equiv \det \gamma_{AB}$. Explicit calculation shows that the field equations which follow from (41) are

$$R_{AB} = 0 . \tag{42}$$

It should be emphasized that the variational principle (41) is completely basis-invariant. We will work in the horizontal lift basis but any other basis would lead to the same final results. Also note that (41) is very geometrical and is a simple and elegant generalization of general relativity to the fiber bundle. No "source" terms are present so that the theory is completely self-sourced. The spin-3/2 field $F^{i}_{\mu\nu}$ which is buried in (42) arises completely naturally in the geometry of the fiber bundle and does not have to be put into (41) by hand as in the supergravity theories.[6,7] Both $g_{\mu\nu}$ and $F^{i}_{\mu\nu}$ arise as gauge fields, with the fundamental gauge group being the group of arbitrary coordinate transformation in the fiber bundle.

We are now in a position to write out our field equations (42) in terms of $g_{\alpha\beta}(x)$, g_{ij}, and $F^{i}_{\mu\nu}(x)$. Putting (38) and (39) into (42) and using (37) gives

$$R_{\mu\nu} = R^{(E)}_{\mu\nu} + \tfrac{1}{2} g_{ji} F^{j}_{\delta\mu} F^{i}_{\nu\beta} g^{\delta\beta} + \tfrac{1}{2}(F^{i}_{\nu\mu,i} + F^{j}_{\nu\mu} \Gamma^{i}_{ji}) = 0 , \tag{43}$$

$$R_{\mu i} = -\tfrac{1}{2} g_{ji}(F^{j}_{\delta\mu} g^{\delta\alpha})_{,\alpha} - \tfrac{1}{2} g_{ji} F^{j}_{\delta\mu} g^{\delta\alpha} \Gamma^{\beta}_{\alpha\beta} + \tfrac{1}{2} g_{ji} F^{j}_{\delta\alpha} g^{\delta\beta} \Gamma^{\alpha}_{\mu\beta} = 0 , \tag{44}$$

$$R_{ij} = R^{(E')}_{ij} - (\tfrac{1}{2} g_{ik} F^{k}_{\pi\alpha} g^{\pi\mu})(\tfrac{1}{2} g_{lj} F^{l}_{\delta\mu} g^{\delta\alpha}) = 0 , \tag{45}$$

where

$$R^{(E)}_{\mu\nu} \equiv \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\delta}_{\mu\nu}\Gamma^{\alpha}_{\delta\alpha} - \Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\alpha\nu} \tag{46}$$

and

$$R^{(E')}_{ij} \equiv \Gamma^{k}_{ij,k} + \Gamma^{k}_{ik,j} + \Gamma^{k}_{ij}\Gamma^{l}_{kl} + \Gamma^{l}_{ik}\Gamma^{k}_{lj} . \tag{47}$$

We note that the curvature tensor (39) has the symmetries

$$R_{\alpha i} = -R_{i\alpha} \quad \text{and} \quad R_{ij} = -R_{ji} , \tag{48}$$

but that $R_{\mu\nu} \neq R_{\nu\mu}$ because of the $F^{i}_{\mu\nu}$ terms. In fact (43) has the form of the sum of a symmetric part and an antisymmetric part. These must vanish separately so that (43) becomes

$$R^{(E)}_{\mu\nu} + \tfrac{1}{2} g_{ji} F^{j}_{\delta\mu} F^{i}_{\nu\beta} g^{\delta\beta} = 0 \tag{43a}$$

and

$$F^{i}_{\nu\mu,i} + F^{j}_{\nu\mu} \Gamma^{i}_{ji} = 0 . \tag{43b}$$

Now from our definitions of covariant derivatives in (24)–(26) we can write

$$F^{i}_{\alpha\beta;C} = F^{i}_{\alpha\beta,C} + F^{j}_{\alpha\beta} \Gamma^{i}_{jC} - F^{i}_{\alpha\delta} \Gamma^{\delta}_{\beta C} - F^{i}_{\delta\beta} \Gamma^{\delta}_{\alpha C} , \tag{49}$$

where C is a general index. Using (49) and (37) shows that (43b) can be written as

$$F^{i}_{\mu\nu;i} = 0 \tag{50}$$

and (44) can be written as

$$g_{ij}(F^{j\nu}{}_{\mu})_{;\nu} = 0. \tag{51}$$

Thus our field equations (43a), (45), (50), and (51) can be written in completely covariant form.

If we look at (43a) we see that $R^{(E)}_{\mu\nu}$, $F^{j}_{\delta\mu}$, and $g^{\delta\beta}$ are all functions of x_μ only while g_{ji} is a function of θ_i only (a power series with terms up to θ^4). (43a) thus implies that g_{ji} must be a constant matrix and not a function θ_i and can be written as

$$g_{ji} = Q\eta_{ji} , \tag{52}$$

where Q is a constant and $\eta_{ji} = (-C^{-1})_{ji}$. C is the charge conjugation matrix ($C^* = C^{-1}$ and $C^T = -C$) and is the only constant Fermi sector matrix which is available to play this role.[4] Note that the conclusion (52) *holds only in the horizontal lift basis.* It would not hold in some other basis. Using (52) and (10), our field equations can be simplified to the following form:

$$R^{(E)}_{\alpha\beta} - \tfrac{1}{2} g_{\alpha\beta} R^{(E)} \equiv -(8\pi\kappa/c^2) T_{\alpha\beta} , \tag{53}$$

$$(F^{j\nu}{}_{\mu})_{;\nu} = 0 , \tag{54}$$

$$F^{k\mu}{}_{\alpha} F^{l\alpha}{}_{\mu} = 0 , \tag{55}$$

with

$$-c^2 T_{\mu\nu} = \eta_{ji} F^{j}_{\alpha\mu} F^{i\alpha}{}_{\nu} - \tfrac{1}{4} g_{\mu\nu} \eta_{ji} F^{j}_{\alpha\beta} F^{i\alpha\beta} . \tag{56}$$

We used (55) to write (56) in traceless form and identified

$$Q = 16\pi\kappa/c^4 . \tag{57}$$

$R^{(E)}_{\alpha\beta}$ and $R^{(E)}$ refer to the Einstein quantities. (Notice that our fiber bundle equations are sourceless since $R_{AB} = 0$.) We also must satisfy the cyclic condition on $F^{i}_{\mu\nu}$ from the Jacobi identity (13). The Jacobi identity (16) and the field equation (43b) are identically satisfied using (52) and (10). We thus end up with a very simple set of equations. These equations were derived in the horizontal lift basis. Any other basis will give a *set* of equations which are equivalent to these, although individual components of R_{AB} look different in different bases.[13] The variational principle (41) is basis invariant since R is a scalar, and hence the theory is completely basis invariant. We can write out R as

$$R = R^{(E)}_{\alpha\beta} \gamma^{\alpha\beta} + R^{(E')}_{ij} \gamma^{ij} - \tfrac{1}{4} \gamma_{ji} F^{j}_{\delta\mu} F^{i}_{\beta\nu} \gamma^{\delta\beta} \gamma^{\mu\nu} , \tag{58}$$

$R^{(E)}_{\mu\nu}$ and $R^{(E')}_{ij}$ are given in (46) and (47). This R would take the same form in any basis.

It is important to note that our final set of equations (13), (53)–(55) is invariant under general x_μ coordinate transformations and under "gauge transformations" of the $A^{(\sigma)i}_{\mu}$ in (23). Since the $A^{(\sigma)i}_{\mu}$ are the coefficients of cross section dependent connection forms, different gauges correspond to different choices of the cross section.[18] More explicitly, $F^{i}_{\mu\nu}$ given in (23) is invariant under a change in the potentials of the form

$$A^{(\sigma)i}_{\mu} \rightarrow A^{(\sigma)i}_{\mu} + D_\mu \epsilon^i(x) , \tag{59}$$

where $\epsilon^i(x)$ is a spinor which depends only on x_μ. It is easy to show that (57) corresponds to a change in cross section $\sigma(x, \theta)$ in (19). Under this transformation

$$\begin{aligned} \delta F^{i}_{\mu\nu} &= D_\nu(D_\mu \epsilon^i(x)) - D_\mu(D_\nu \epsilon^i(x)) \\ &= [D_\nu , D_\mu] \epsilon^i(x) = [\hat{\xi}_\nu , \hat{\xi}_\mu] \epsilon^i(x) \\ &= -F^{j}_{\nu\mu} \xi^*_j \epsilon^i(x) = 0 , \end{aligned} \tag{60}$$

since $\xi^*_j \epsilon^i(x) = 0$. All of our derivatives are also gauge covariant. Our field equations are thus invariant under "gauge transformations" of the form (59).

As in general relativity, we have differential Bianchi identities which reduce the number of independent equations which the $g_{\mu\nu}(x)$ and $F^{i}_{\mu\nu}(x)$ must satisfy. We can write

the variation of (41) with respect to γ_{BA} as

$$\int (-\gamma)^{1/2}(-1)^a[R^{AB} - \tfrac{1}{2}\gamma^{AB}R]\,\delta\gamma_{BA}\,d^8z = 0\,. \quad (61)$$

Under an infinitesimal coordinate transformation,

$$z^A = z^{A'} + \xi^A(z)\,, \quad (62)$$

we have

$$\delta\gamma_{AB} \equiv \gamma'_{AB}(z) - \gamma_{AB}(z) = \gamma_{AC}\,\xi^C_{,B} + (-1)^{a+b+ab}\gamma_{BC}\,\xi^C_{,A} + \gamma_{AB,C}\,\xi^C\,. \quad (63)$$

This can be put in the form

$$\delta\gamma_{AB} = \xi_{A;B} + (-1)^{a+b+ab}\xi_{B;A} + \gamma_{Ai}\,\xi^\delta H^i_{\delta B} + (-1)^{a+b+ab}\gamma_{Bi}\,\xi^\delta H^i_{\delta A}\,, \quad (64)$$

where we used (30) and $H^i_{\delta B}$ is given by (35). Putting (64) into (61) and noting that the result must hold for arbitrary ξ_A leads finally to the Bianchi identity

$$(-1)^a(R^{AB} - \tfrac{1}{2}\gamma^{AB}R)_{;A} \equiv 0\,. \quad (65)$$

This represents eight identities.

At this point it is interesting to see if the number of independent equations versus the number of unknown quantities makes sense. If we use (23) to write $F^i_{\mu\nu}$ in terms of A^i_μ, we have 16 unknown components of A^i_μ, and ten unknown components of $g_{\mu\nu}$ to be determined. The cyclic Jacobi identity (13) is satisfied identically for $F^i_{\mu\nu}$ of the form (23). We are left with the ten equations (53) and the 16 equations (54). These 26 equations are reduced to 18 independent equations if the Bianchi indentities (65) are taken into account. Since we have 26 unknowns, we have eight degrees of freedom that remain to be specified. These are just four x_μ coordinate conditions plus four gauge fixing conditions for A^i_μ, so that the numbers work out as one might expect.

IV. DISCUSSION

Our final set of equations (13), (53)–(56) represents a massless spin-3/2 field coupled to general relativity. Unfortunately, the spin-3/2 field is in a very unfamiliar representation. To verify that we do indeed have a spin-3/2 field, we can start with the massless, spin-3/2, two-component spinor equations of Corson,[9]

$$\partial_{ru}\,\Phi^{ust} = 0 \quad (66)$$

and

$$\partial^{ru}\chi_r{}^{st} = 0 \quad (67)$$

along with the complex conjugate equations

$$\partial_{\dot u\dot r}\,\Phi^{\dot u\dot s\dot t} = 0 \quad (68)$$

and

$$\partial^{\dot u\dot r}\chi_{\dot r}{}^{\dot s\dot t} = 0\,. \quad (69)$$

(We use Latin letters near the end of the alphabet to denote a two-component spinor index.) These spinors must be symmetric under interchange of indices to prevent lower spins from appearing also. These spinorial derivatives are defined, for example, as

$$\partial_{ru} \equiv \partial_\gamma\, g^\gamma{}_{ru} \quad (70)$$

with the $g^\gamma{}_{ru}$ given in terms of the Pauli spin matrices and the unit matrix as in Corson.[19] Now, extending the developing of Corson a bit, we can write a general antisymmetric tensor with one 2 component spinor index as

$$G^t_{\mu\nu} = C_0[\,g_\mu{}^s{}_r\,g_{\nu\dot s p}\,\Phi^{rpt} + g_{\mu\dot r}{}^s\,g_{\nu\dot p s}\,\Phi^{\dot r\dot p\dot t}\,] \quad (71)$$

along with the dual

$$\check G^t_{\mu\nu} = -C_0[\,g_\mu{}^s{}_r\,g_{\nu\dot s p}\,\Phi^{rpt} - g_{\mu\dot r}{}^s\,g_{\nu\dot p s}\,\Phi^{\dot r\dot p\dot t}\,]\,, \quad (72)$$

where C_0 is some constant.

A second general antisymmetric tensor with one dotted spinor index can be written as

$$H_{\mu\nu i} = -C_0[\,g_\mu{}^s{}_r\,g_{\nu\dot s p}\,\chi^{rp}{}_i + g_{\mu\dot r}{}^s\,g_{\nu\dot p s}\,\chi^{\dot r\dot p}{}_i\,] \quad (73)$$

along with the dual $\check H_{\mu\nu i}$. It is now easy to see by adding and subtracting the divergences of (71) and (72) and the divergence of (73) and its dual equation that the spin-3/2 field equations given in (66)–(69) are completely equivalent to the equations

$$\partial^\mu G_{\mu\nu}{}^t = \partial^\mu \check G_{\mu\nu}{}^t = \partial^\mu H_{\mu\nu i} = \partial^\mu \check H_{\mu\nu i} = 0\,. \quad (74)$$

From the way in which $G_{\mu\nu}{}^t$ and $H_{\mu\nu i}$ are defined, we can now combine these to give a four-component spinor

$$F^i_{\mu\nu} = \begin{pmatrix} G_{\mu\nu}{}^t \\ H_{\mu\nu i} \end{pmatrix}, \quad (75)$$

where $i = 1,2,3,4$. Outr spin-3/2 field equations in flat space then become

$$\partial^\mu F^i_{\mu\nu} = \partial^\mu \check F^i_{\mu\nu} = 0\,, \quad (76a)$$

with the antisymmetric field $F^i_{\mu\upsilon}$ satisfying

$$F^i_{\mu\nu}F_i^{\mu\nu} = \check F^i_{\mu\nu}\check F_i^{\mu\nu} = 0\,. \quad (76b)$$

(76b) follows from the symmetry properties of the basic spinors in (66)–(69). These symmetry properties are necessary to prevent the appearance of lower spin fields. The derivation of (76a) and (76b), can be reversed leading back to (66)–(69), which shows that (76a) and (76b) do indeed represent a field with spin 3/2 and only spin 3/2 present. One advantage of this representation is that it generalizes readily to curved space giving finally

$$(F^{i\nu}{}_\mu)_{;\nu} = 0 \quad (77)$$

and the dual

$$(\check F^{i\nu}{}_\mu)_{;\nu} = 0\,, \quad (78)$$

where a semicolon denotes a covariant derivative. It is important to note that $F^i_{\mu\nu}$ itself represents the physical spin-3/2 field in this representation, and that this physical field satisfies first-order, ghostless equations as we would expect. The reason is that the equivalent spinor equations (66)–(69) are uncoupled linear equations for this massless case (and only for this case). [In the massive case our Eqs. (74) acquire a vector potential term on the right and become second order equations in terms of the spinors Φ or χ. In this case C_0 becomes proportional to the mass in (71). Using the massive version of (66)–(69) shows that $F^t_{\mu\nu}$, for example, is then proportional to a derivative of χ and (74) become second order equations in terms of χ. Thus our description (74) works for the present massless gauge fields but is inferior to a Rarita–Schwinger formulation in the massive case.] It is interesting

that the spin-3/2 field of FvFDZ, $\Psi_{\mu}^{\ i}$, can also be written in terms of the Φ and χ spinors in equations similar to (71)–(73), and we have a crude correspondence $F_{\mu\nu}^{\ i} \sim \gamma_{\mu}\ \Psi_{\nu}^{\ i}$ between their field and ours. Unfortunately, no neat mathematical relation between $F_{\mu\nu}^{\ i}$ and $\Psi_{\nu}^{\ i}$ seems to exist.

Equation (77) is now identical with our field equation for $F^{j\nu}{}_{\mu}$ given in (54). Also (78) is equivalent to our Jacobi relation (13). Similarly, we can show in a rather tedious calculation using two-component spinors that our unfamiliar energy–momentum tensor given in (56) is equivalent to the usual energy–momentum tensor for a spin-3/2 field. We notice in this context, that the effective Lagrangian for the spin-3/2 field in (58) and the field equations (54) both have a Maxwell-like form so that it is not suprising that the energy–momentum tensor (56) also has a Maxwell-like form in this representation. We should comment that our remaining equation (55) holds for a sourceless massless spin-3/2 field in the same way that $F^{\mu\nu}F_{\mu\nu} = \mathbf{E}^2 - \mathbf{B}^2 = 0$ holds for solutions of the sourceless Maxwell equations. It is thus not really a field equation. Thus we conclude that our equations do indeed represent a massless spin-3/2 field coupled to general relativity.

The role of supersymmetry transformations in this theory also merits discussion. In Arnowitt and Nath,[4] the usual supersymmetry transformation

$$x^{\mu'} = x^{\mu} + \tfrac{1}{2}i\bar{\epsilon}\gamma^{\mu}\theta\,, \qquad \theta^{i'} = \theta^{i} + \epsilon^{i}\,, \tag{79}$$

where ϵ^{i} is an infinitesimal Majorana spinor parameter, plays the same role as the Poincaré transformations play in the transition from special to general relativity. The fundamental gauge group is thus the group of arbitrary coordinate transformations in supersymmetry space for Arnowitt and Nath. The supersymmetry transformation (70) thus becomes submerged only to resurface later in their work as a result of spontaneous symmetry breaking. Similarly, in our work our field equations do not possess supersymmetry invariance. Whether or not supersymmetry plays any role after spontaneous symmetry breaking in the present theory remains to be seen since we have not yet worked out spontaneous symmetry breaking effects. It seems rather unlikely that supersymmetry will emerge, however, since spontaneous symmetry breaking itself may be a problem in this self-sourced theory. No scalar fields are now present for spontaneous symmetry breaking, and it is not at all clear how they can arise in a natural way.

Another potential problem in the absence of supersymmetry invariance is that our spin-2-3/2 system may suffer from acausal propagation. It is well known that massless high spin gauge theories need to be treated with great care.[20] The related questions of supersymmetry invariance and behavior of our spin-2-3/2 system clearly deserves more attention in subsequent work.

The similarity between this work and simple $N = 1$ supergravity is striking, even though our field equations do not possess supersymmetry invariance. We have a theory which is coordinate system and basis invariant under transformations in superspace and which ends up quite naturally with a spin 3/2, massless field coupled to general relativity. Beyond that, however, the similarities cease. First, it would certainly be nice if supersymmetry invariance could be accomodated in some fashion in the present theory. Second, our spin-3/2 field seems to bear no simple relationship to that of FvFDZ,[6,7] primarily because of our peculiar representation. Third, torsion plays a different role in the two theories. We have explicitly taken our eight-dimensional torsion to be zero, as in the work of Arnowitt and Nath. Four-dimensional torsion, however, plays a key role in supergravity especially in the first-order formulation. Our effective Langrangian is "nonminimally" coupled in terms of Rarita–Schwinger fields with this "nonminimal" coupling following from the fiber bundle structure of the theory. Thus we end up with something more akin to the torsionless original version[6] of supergravity with nonminimal four Fermion couplings present, than to the first-order formulation with torsion.[6,7]

A number of things remain to be considered in future papers. Since massless spin-3/2 fields very likely do not exist in nature, spontaneous symmetry breaking probably plays a role and gives a large mass to the spin-3/2 field. As we mentioned above, spontaneous symmetry breaking may be a problem in this theory, however. We have also discussed only a pure gravitational field above. How matter fields are to be put in is yet to be seen. Hopefully one can do this without destroying the geometrical elegance of the theory, perhaps by way of an associated fiber bundle. Ultimately, renormalizability must be considered. Other theories suggest that the spin-3/2 field may make the theory more renormalizable than general relativity, especially if supersymmetry can somehow be accomodated.

To summarize, we have a very simple and elegant geometrical formulation of superspace using a fiber bundle. A spin-3/2 field rather magically appears in the fiber bundle geometry in addition to the usual spin-2 gravitational field. These fields are both gauge fields and the basic theory is sourceless. A certain economy is achieved since these fields are the only ones to appear rather than the very large number of fields which appear in the work of Arnowitt and Nath.[4] The resulting equations are similar to those of the simple supergravity of FvFDZ,[6,7] but our field equations do not possess supersymmetry invariance. This paper shows how a self-sourced theory of gravitation can be formulated in superspace in a fiber bundle approach. Whether or not the present theory leads to a renormalizable theory of the gravitational field remains to be seen.

[1] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974); B. Zumino, in *Proceedings of the XVII International Conference on High Energy Physics*, London, 1974, edited by J. R. Smith (Rutherford Laboratory, Chilton, Didcot, Berkshire, England, 1974), pp.1–254.

[2] D. V. Volkov and V. A. Soroka, Zh. Eksp. Teor. Fiz. Pis'ma Red. **18**, 529 (1973) [JETP Lett. **18**, 312 (1973)].

[3] A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); B **80**, 499 (1974); Phys. Rev. D **11**, 1521 (1975).

[4] P. Nath and R. Arnowitt, Phys. Lett. B **56**, 177 (1975); R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B **56**, 81 (1975); P. Nath and R.

Arnowitt, J. Phys. (Paris) **37**, C2-**75** (1976); R. Arnowitt and P. Nath, Gen. Rel. Grav. **7**, 89 (1976); P. Nath, in *Gauge Theories and Modern Field Theory*. Proceedings of the Boston Conference, 1975, edited by R. Arnowitt and P. Nath (MIT Press, Cambridge, Mass., 1976), p. 281; R. Arnowitt and P. Nath, Phys. Rev. Lett. **36**, 1526 (1976); Phys. Rev. D **15**, 1033 (1977); P. Nath and R. Arnowitt, Phys. Rev. D **18**, 2759 (1978).

[5]P. Nath and R. Arnowitt, Phys. Rev. Lett. **42**, 138 (1979).

[6]D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976); D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D **14**, 912 (1976).

[7]S. Deser and B. Zumino, Physics Lett. B **62**, 335 (1976).

[8]A recent review talk by B. Zumino at Liblice, Czechoslovakia contains numerous references. Columbia University Preprint CU-TP-132, June 1978.

[9]W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).

[10]See the review by M. Grisaru and P. van Nieuwenhuizen in *Deeper Pathways in High-energy Physics*, Proc. Orbis Scientiae, Coral Gables, 1977 (Plenum, New York, 1977).

[11]S. Deser, J. H. Kay, and K. S. Stelle, Phys. Rev. Lett. **38**, 527 (1977).

[12]P. Nath and R. Arnowittt, Phys. Lett. B **65**, 73 (1976).

[13]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).

[14]The reader is referred to the following for general mathematical background. S. Sternberg, *Lectures on Differential Geometry* (Prentice–Hall, Englewoood Cliffs, N. J., 1964); S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963); R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds* (Academic, New York, 1965); W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, (Springer, New York, 1977).

[15]M. E. Mayer, Ref. 14.

[16]T. Kaluza, Sitzber. Preuss. Akad. Wiss., 966 (1921); O. Klein, Z. Physik **37**, 895 (1926); B. DeWitt, *Dynamical Theories of Groups and Fields* (Gordon and Breach, New York, 1965), p. 139; R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968); A. Trautman, Rep. Math. Phys. **1**, 29 (1970).

[17]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[18]A. Trautman, Ref. 16.

[19]E. M. Corson, *Tensors, Spinors, and Relativistic Wave-Equations*, (Blackie, Glasgow, 1953).

[20]T. Cukierda and J. Lukierski, J. Math Phys. **11**, 46 (1970); C. Fronsdal and J. Fang, Phys. Rev. D **18**, 3630 (1978); F. A Berends, J. W. van Holden, P. van Nieuwenhuizen, Phys. Lett. B **83**, 188 (1979), Nucl. Phys. B **154**, 261 (1979); D. H. Tchrakian, Lett. Nuovo Cimento **24**, 18 (1979); T. Curtright, Phys. Lett B **85**, 219 (1979); C. Aragone and S. Deser, Phys. Lett. B **86**, 161 (1979).

Gauge groups in a fiber bundle approach to superspace and the gravitational field

D. K. Ross
Physics Department and Ames Laboratory, USDOE, Iowa State University, Ames, Iowa 50011

(Received 22 May 1980; accepted for publication 15 September 1980)

A semisimple Lie group is incorporated into our fiber bundle model of superspace and the gravitational field. The Lagrangian density is taken to be the scalar curvature of the $(8 + n)$-dimensional fiber bundle where n is the dimensionality of the Lie group. The resulting field equations describe a nonabelian gauge field correctly coupled to the spin-3/2 gravitino and the spin-2 graviton.

PACS numbers: 04.50. + h, 11.30.Pb

I. INTRODUCTION

We have recently formulated a fiber bundle model[1] of superspace in which four-dimensional Fermi space F, with anticommuting Majorana spinor coordinates θ^i, plays the role of the typical fiber and space–time plays the role of the base space. In this model, superspace becomes a fiber bundle rather than a general eight-dimensional Riemannian manifold as in the gauge supersymmetry work of Nath and Arnowitt.[2] The Lagrangian density is taken to be the scalar curvature of the fiber bundle. A spin 3/2 field appears rather magically, and the resulting "self-sourced" field equations describe this gravitino field correctly coupled to the spin 2 field of general relativity. No other fields are present. Although the theory bears a strong resemblance to simple $N = 1$ supergravity,[3] supersymmetry invariance is not present.

The purpose of the present paper is to extend our fiber bundle model by including a gauge group such as SO(8),[4] SO(3),[5] SU(4),[6] etc., all of which have been discussed in the context of supergravity. We need to do this if we are to have any hope of obtaining a physically realistic theory. So far we have only a pure gravitational field. Nath and Arnowitt[2,7] put in a gauge group by enlarging the number of Fermi coordinates from θ^i to θ^{iq}, where q refers to the gauge group. More conventional supergravity[4–6] puts in a gauge group by constructing Lagrangians whose couplings are invariant under larger irreducible representations of the global supersymmetry algebra.

For us, there are several possible ways of putting a gauge group G (a general semisimple Lie group) into our fiber bundle. These include: (1) a hierarchical model in which our previous fiber bundle plays the role of the base space of a new, larger fiber bundle with principal fiber G,(2) a model with space-time as the base space and a Fermi space with an enlarged number of coordinates θ^{iq},(3) a model with an eight-dimensional general Riemannian manifold with coordinates (x_μ, θ_i) as the base space and G as the principal fiber, and (4) a model with space–time as the base space and the direct product $F \times G$ as the typical fiber, where F is four-dimensional Fermi space and G is the Lie group. Of these models only (4) leads to a reasonable theory with no extraneous fields beyond what one would expect. We will explore model (4) in the following. Note that (4) is essentially the method of Nath and Arnowitt[2] correctly transcribed into fiber bundle language whereas, contrary to expectations, (2) is not.

II. STRUCTURE OF THE FIBER BUNDLE

Cho[8] has written a very nice paper in which he puts a Lie group into general relativity by way of a fiber bundle. We will follow his approach below. Choose a set of n linearly independent left invariant vector fields ξ_a for a basis of the Lie group G.[9] Thus we have $\forall\, a \in G$, $L_a \xi_b = \xi_b$, where L_a: $b \in G \to ab \in G$. These ξ_a are also a basis of the Lie algebra $\mathcal{G}$ of G which forms an n-dimensional vector space. They have commutation relations

$$[\xi_a, \xi_b] = f^c_{ab} \xi_c, \tag{1}$$

where the f^c_{ab} are the structure constants of the group G. This is a noncoordinate basis. Following our work in I and II, the four-dimensional Fermi space F with anticommuting Majorana coordinates θ^i, can be considered to be a group. We can choose a coordinate basis for F with the basis vectors satisfying anticommutation relations,

$$\{\xi_i, \xi_j\} = 0. \tag{2}$$

(In the following a,b,c,d,e will denote n-dimensional group indices; i,j,k,l will denote four-dimensional Fermi indices; lower case Greek letters will denote four-dimensional space–time indices; upper case A,B,C,D,E,F,G, will denote $(8 + n)$-dimensional fiber bundle indices; and r,s will denote the combination of group plus Fermi indices taking on $4 + n$ values.) If we now consider the direct product $F \times G$, we will have

$$[\xi_a, \xi_i] = 0 \tag{3}$$

also. (1), (2), and (3) can then be considered to be the commutation relations of the basis vectors of $F \times G$. Since both commutators and anticommutators appear, (1), (2), and (3) are the basis of a Grassman algebra.

Now we want to consider a fiber bundle $\underline{P}$ whose base space M is space–time and whose typical tiber is $F \times G$. We consider the fibration $\lambda = (\underline{P}, M, \Pi)$, where the base space M and the bundle space $\underline{P}$ are smooth manifolds and the projection Π is a C^∞ mapping of $\underline{P}$ into M. Π is surjective and locally trivial: for any $x \in M$ there exists an open neighborhood U of x, a manifold $F \times G$ and an isomorphism (diffeo-

 022-2488/81/030557-05$1.00

morphism)Φ of $\Pi^{-1}(U)$ onto $U \times (F \times G)$ such that $\Pi(\Phi^{-1}(x,a) = x$ for all $x \in U$ and $a \in F \times G$ (we denote elements of F by θ, elements of G by a, and elements of $F \times G$ by α, $\mathscr{B}$, etc. $\Pi^{-1}(x)$ is a closed submanifold of $\underline{P}$ called the fiber over x. The fibers are diffeomorphic to $F \times G$, the typical fiber.

Following Cho, the basis vectors ξ_a and ξ_i of $F \times G$ can now be mapped naturally into the fiber bundle $\underline{P}$ where they will be designated $\xi_a{}^*$ and $\xi_i{}^*$. These are tangent to the fiber space and form a subspace, called the vertical subspace V_p of the tangent space $T_p(\underline{P})$ to $\underline{P}$. In the fiber bundle, they have the commutation relations

$$[\xi_a{}^*, \xi_b{}^*] = f^c_{ab} \xi_c{}^*, \tag{4}$$

$$\{\xi_i{}^*, \xi_j{}^*\} = 0, \tag{5}$$

and

$$[\xi_i{}^*, \xi_a{}^*] = 0. \tag{6}$$

To go further, we need to define a connection Γ. This is a choice of a horizontal subspace H_p at each point of $\underline{P}$ such that

a) The tangent space $T_p(\underline{P})$ is the direct sum of V_p and H_p.

b) For all $\alpha \in F \times G$ and $p \in \underline{P}$, $H_{p\alpha} = R_\alpha H_p$ where R_α: $(p,a) \in \underline{P} \times (F \times G) \to p\alpha \in \underline{P}$ with $p(\alpha\mathscr{B}) = p(\alpha)\mathscr{B}$ for all $\alpha, \mathscr{B} \in F \times G$, $p \in \underline{P}$.

c) H_p is smooth on $\underline{P}$.

Using this connection, we can define the horizontal lift of any vector A at $x \in M$ as the unique vector $\hat{A} \in H_p$ at each point of the fiber $\Pi^{-1}(x)$ such that $\Pi(\hat{A}) = A$. We can also define a coordinate basis $\xi_\mu = \partial_\mu$ in the base space with

$$[\xi_\mu, \xi_\nu] = 0. \tag{7}$$

These ξ_μ can be horizontally lifted into the bundle and denoted there by $\hat{\xi}_\mu$. $(\hat{\xi}_\mu, \xi_i{}^*, \xi_a{}^*)$ form a very convenient basis for the fiber bundle with commutation relations[1]

$$[\hat{\xi}_\mu, \hat{\xi}_\nu] = -F^i_{\mu\nu} \xi_i{}^* - F^a_{\mu\nu} \xi_a{}^*, \tag{8}$$

$$[\hat{\xi}_\mu, \xi_j{}^*] = 0, \tag{9}$$

$$[\hat{\xi}_\mu, \xi_a{}^*] = 0, \tag{10}$$

along with (4), (5), and (6). This horizontal lift basis is very convenient for calculations though not the only possible basis, of course. (8) is simply the statement that $[\hat{\xi}_\mu, \hat{\xi}_\nu]$ must be vertical. The coefficients $F^i_{\mu\nu}$ and $F^a_{\mu\nu}$ will play the role of physical fields in the following.

Given a connection Γ, we can always define a connection form ω such that

$$\omega^r(\xi_s{}^*) = \delta^r_s, \tag{11}$$

$$\omega^r(\hat{\xi}_\mu) = 0, \tag{12}$$

where r and s are summed over both Fermi (i) and group (a) indices. As in Papers I and II, we can define a cross section $\sigma(x)$ as a submanifold of $\underline{P}$ which is diffeomorphic to an open subset U of M with $\Pi(\sigma(x)) = x$. If $\sigma \cdot \xi_\mu$ is a vector in the tangent space to $\underline{P}$ at $\sigma(x)$ induced by $\sigma: x \in U \to \sigma(x) \in \Pi^{-1}(U)$, we can define a connection form $A^{(\sigma)}$ on U [or on $\sigma(x)$] by

$$A^{(\sigma)}(\xi_\mu) \equiv A^{(\sigma)}{}_\mu{}^r \xi_r = \omega(\sigma \cdot \xi_\mu) = \omega^r(\sigma \cdot \xi_\mu)\xi_r. \tag{13}$$

$A^{(\sigma)}{}_\mu{}^r$ are our cross-section dependent potentials. Now as in I and II, $\hat{\xi}_\mu$ is the gauge covariant derivative D_μ when we write it out in the local direct product basis. We can define the trivial cross section $\sigma_{(\alpha)}$ as the set of points $p = (x,\alpha)$ with α fixed, where α is an element of $F \times G$. If $\bar{\xi}_\mu$ and $\bar{\xi}_r$ denote basis vectors in the local direct product basis, we have

$$\hat{\xi}_\mu = \bar{\xi}_\mu - \bar{\xi}_r \omega^r(\bar{\xi}_\mu) = \bar{\xi}_\mu - \bar{\xi}_r B^r_\mu \equiv D_\mu, \tag{14}$$

where $B^r_\mu \equiv A^{\sigma_{(\alpha)}r}_\mu$ (see I and II for details). We see that

$$D_\mu = \bar{\xi}_\mu - \bar{\xi}_i B^i_\mu - \bar{\xi}_a B^a_\mu \tag{15}$$

and hence involves both the potential $B_\mu{}^i$ associated with the spin 3/2 gravitino field and the potential $B_\mu{}^a$ associated with the gauge group G.

Our fields $F^r_{\mu\nu}$ can now be written in terms of the potentials B^r_μ in the local direct product basis. Using (8) and (14), we can write

$$F^i_{\mu\nu} = \partial_{\bar{\mu}} B^i_\nu - \partial_{\bar{\nu}} B^i_\mu \tag{16}$$

and

$$F^a_{\mu\nu} = \partial_{\bar{\mu}} B^a_\nu - \partial_{\bar{\nu}} B^a_\mu + f^a_{cb} B^c_\mu B^b_\nu \tag{17}$$

so that $F^i_{\mu\nu}$ and $F^a_{\mu\nu}$ take the forms we might expect. $\partial_{\bar{\mu}}$ denotes a partial derivative in the local direct product basis. In the following, we will deal primarily with $F^i_{\mu\nu}$ and $F^a_{\mu\nu}$ rather than the potentials.

We can denote a metric for the base space M as $g_{\mu\nu}(x)$ and for the four-dimensional Fermi space as $g_{ij}(\theta)$. It is important that these can depend at most only on x and θ respectively. The metric for the semisimple Lie group G can be written as

$$g_{ab} = f^d_{ac} f^c_{db} W, \tag{18}$$

where W is a constant to be determined later and f^d_{ac} are structure constants for the group. [Note that the definition of g_{ab} will not work for the abelian group U(1). In that case we can let g_{11} (U(1) is one-dimensional) float until we get to the field equations. Quantities such as $g_{11} F^1_{\mu\delta} F^1_{\nu\beta}$ will appear in these equations. If we let $g_{11} = 1$ and $F^1_{\mu\delta} = F_{\mu\delta}$ (the electromagnetic field tensor), our resulting field equations correctly describe electricity and magnetism coupled to the spin 2 graviton and the spin 3/2 gravitino fields. In fact, in this case, ${}^{(E^n)}R_{ab}$ vanishes identically, so that the cosmological constant term in (63) never even appears.] We can define a metric γ_{AB} on the fiber bundle, which is compatible with the above metrics and the fact that $F \times G$ is the typical fiber, as

$$\gamma_{AB} \hat{\xi}^A_\mu \hat{\xi}^B_\nu = g_{\mu\nu}, \tag{19}$$

$$\gamma_{AB} \xi^{*A}_i \xi^{*B}_k = g_{ik}, \tag{20}$$

$$\gamma_{AB} \xi^{*A}_a \xi^B_b = g_{ab}, \tag{21}$$

$$\gamma_{AB} \hat{\xi}^A_\mu \xi^{*B}_i = 0, \tag{22}$$

$$\gamma_{AB} \hat{\xi}^A_\mu \xi^{*B}_a = 0, \tag{23}$$

$$\gamma_{AB} \xi^{*A}_i \xi^{*B}_a = 0. \tag{24}$$

This is a basis invariant definition. In the horizontal lift basis, γ_{AB} is particularly simple and we have

$$\gamma_{AB} = \begin{pmatrix} g_{\mu\nu} & 0 & 0 \\ 0 & g_{ik} & 0 \\ 0 & 0 & g_{ab} \end{pmatrix}. \tag{25}$$

In another basis, off diagonal terms would appear.

At this point, we have $g_{\mu\nu}(x)$ and the superfields $F^i_{\mu\nu}(x,\theta,a)$ and $F^a_{\mu\nu}(x,\theta,a)$ which can play the role of physical fields. The fiber space dependence of $F^i_{\mu\nu}$ and $F^a_{\mu\nu}$ is given by the Jacobi identites, which will greatly simplify these superfields. There are ten Jacobi identites, one of which is

$$[\hat{\xi}_\mu,[\hat{\xi}_\nu,\xi_i{}^*]]+[\hat{\xi}_\nu,[\xi_i{}^*,\hat{\xi}_\mu]]+[\xi_i{}^*,[\hat{\xi}_\mu,\hat{\xi}_\nu]]=0. \tag{26}$$

Using (8) and (9), this becomes

$$[\xi_i{}^*,-F^r_{\mu\nu}\xi_r{}^*]=0 \tag{27}$$

which can be written as

$$-(\partial_i{}^*F^j_{\mu\nu})\xi_j{}^*-(\partial_i{}^*F^a_{\mu\nu})\xi_a{}^*+F^j_{\mu\nu}\{\xi_i{}^*,\xi_j{}^*\}=0. \tag{28}$$

Using (5) gives finally

$$\partial_i{}^*F^j_{\mu\nu}=0 \tag{29}$$

and

$$\partial_i{}^*F^a_{\mu\nu}=0. \tag{30}$$

Thus these two fields are not functions of the Fermi coordinates θ^i. The other nine Jacobi identities can be worked out similarly using the commutation relations and the expression

$$[\xi_i{}^*,\xi_j{}^*]=2\Gamma^k_{ij}\xi_k{}^*, \tag{31}$$

where Γ^k_{ij} is the Christoffel symbol. We summarize as follows: $F^i_{\mu\nu}$ can only be a function of x^μ; $F^a_{\mu\nu}$ can only be a function of x^μ and group coordinates a; Γ^k_{ij} can oly be a function of θ^i. In addition we must satisfy the following equations:

$$f^d_{bc}f^e_{ad}+f^d_{ca}f^e_{bd}+f^d_{ab}f^e_{cd}=0, \tag{32}$$

$$\partial_a{}^*F^b_{\mu\nu}+F^c_{\mu\nu}f^b_{ac}=0, \tag{33}$$

$$\partial_{\hat{\mu}}F^i_{\nu\delta}+\partial_{\hat{\nu}}F^i_{\delta\mu}+\partial_{\hat{\delta}}F^i_{\mu\nu}=0, \tag{34}$$

$$\partial_{\hat{\mu}}F^a_{\mu\delta}+\partial_{\hat{\nu}}F^a_{\delta\mu}+\partial_{\hat{\delta}}F^a_{\mu\nu}=0, \tag{35}$$

$$\partial_i{}^*\Gamma^l_{jk}+\partial_j{}^*\Gamma^l_{ki}+\partial_k{}^*\Gamma^l_{ij}=0. \tag{36}$$

(32) is merely the statement that $f_{abc}\equiv g_{ae}f^e_{bc}$ must be fully antisymmetric. (33) is the well-known gauge covariance of nonabelian gauge field and (36) will be found later to be satisfied identically in the horizontal lift basis. (34) and (35) are the statements that $F^i_{\nu\delta}$ and $F^a_{\nu\delta}$ can be written in terms of potentials. If we wish to deal with the fields themselves rather than potentials, (34) and (35) behave like additional field equations.

It is important to note that the Jacobi identities tell us that $F^i_{\mu\nu}$ and $F^a_{\mu\nu}$ are *not* superfields but ordinary space–time fields. This vastly reduces the number of possible physical fields in the theory to just $g_{\mu\nu}(x)$,$F^i_{\mu\nu}(x)$, and $F^a_{\mu\nu}(x)$.

We still need Christoffel symbols before we can write out our field equations. As in I and II, we have the double complication of having anticommuting variables and also a noncoordinate basis, since we chose to work in the horizontal lift basis where the metric is particularly simple. From I and II, we have

$$\begin{aligned}\Gamma^E_{AB}=&(-1)^{B'E'}\tfrac{1}{2}[(-1)^{B'D'}\gamma_{AD,B}+(-1)^{A'+B'+A'B'+A'D'}\gamma_{BD,A}\\&-\gamma_{AB,D}]\gamma^{DE}-\tfrac{1}{2}H^E_{BA}-(-1)^{B'E'+(B'+C')(A'+C')}\tfrac{1}{2}\gamma_{CB}H^C_{DA}\gamma^{DE}\\&-(-1)^{B'(B'+C'+E')}\tfrac{1}{2}\gamma_{AC}H^C_{DB}\gamma^{DE}\end{aligned} \tag{37}$$

for the Christoffel symbol in the fiber bundle. A comma denotes the directional derivative in the horizontal lift basis. A, B, etc. are fiber bundle indices and in the present paper include $(\mu,i,a,)$. A' is the Grassman parity associated with index A and is $+1$ if A is Fermi (i) and zero if A is Bose (μ or a). H^C_{DB} is given by

$$H^C_{DB}\equiv(-1)^{B'+D'+B'D'+(B'+D')C'}\Gamma^C_{DB}-\Gamma^C_{BD}, \tag{38}$$

and vanishes in a coordinate basis. In the horizontal lift basis, we have

$$H^D_{AB}=\begin{cases}F^i_{\mu\nu} & \text{if } A=\mu, \quad B=\nu, \quad D=i\\ F^a_{\mu\nu} & \text{if } A=\mu, \quad B=\nu, \quad D=a\\ -f^d_{ab} & \text{if } A=a, \quad B=b, \quad D=d\\ 0 & \text{otherwise.}\end{cases} \tag{39}$$

Writing out the Christoffel symbols gives

$$\Gamma^\nu_{\mu\gamma}=\tfrac{1}{2}[g_{\mu\delta,\gamma}+g_{\gamma\delta,\mu}-g_{\mu\gamma,\delta}]g^{\delta\nu}, \tag{40}$$

$$\Gamma^k_{ij}=-\tfrac{1}{2}[-g_{il,j}+g_{jl,i}-g_{ij,1}]g^{lk}, \tag{41}$$

$$\begin{aligned}\Gamma^e_{ab}=&\tfrac{1}{2}[g_{ad,b}+g_{bd,a}-g_{ab,d}]g^{de}+\tfrac{1}{2}f^e_{ba}\\&+\tfrac{1}{2}g_{cb}f^c_{da}g^{de}+\tfrac{1}{2}g_{ac}f^c_{db}g^{de},\end{aligned} \tag{42}$$

$$\Gamma^i_{\mu\nu}=\tfrac{1}{2}F^i_{\mu\nu}, \tag{43}$$

$$\Gamma^a_{\mu\nu}=\tfrac{1}{2}F^a_{\mu\nu}, \tag{44}$$

$$\Gamma^\nu_{\mu i}=-\Gamma^\nu_{i\mu}=-\tfrac{1}{2}g_{ji}F^j_{\delta\mu}g^{\delta\nu}, \tag{45}$$

$$\Gamma^\nu_{a\mu}=\Gamma^\nu_{\mu a}=-\tfrac{1}{2}g_{ab}F^b_{\delta\mu}g^{\delta\nu}. \tag{46}$$

The remaining 18 Christoffel symbols all vanish.

III. FIELD EQUATIONS

The variational principle for the field equations will be taken to be

$$\delta\int\sqrt{-\gamma}\,R\,d^{8+n}z=0, \tag{47}$$

where R is the scalar curvature of the $8+n$-dimensional fiber bundle. We vary with respect to the metric γ^{AB}, and $\gamma\equiv\det\gamma_{AB}$. No cosmological constant term is included. The field equations which follow from this are

$$R_{FG}=(-1)^{A'}R^{A}{}_{FGA}=0. \tag{48}$$

We can calculate the curvature tensor $R^{A}{}_{FGB}$ by transporting a vector around a closed loop being careful of our noncoordinate basis as in I and II. This leads to

$$\begin{aligned}R_{FG}=&(-1)^{A'}\Gamma^{A}_{FG,A}(-1)^{A'G'}-(-1)^{A'}\Gamma^{A}_{FA,G}\\&+(-1)^{A'}\Gamma^{C}_{FG}\Gamma^{A}_{CA}(-1)^{G'C'}-(-1)^{A'}\Gamma^{C}_{FA}\Gamma^{A}_{CG}(-1)^{A'(A'+C')}\\&-(-1)^{A'}\Gamma^{A}_{FE}H^{E}_{GA}(-1)^{G'+A'+A'G'+E'(A'+G')}=0.\end{aligned} \tag{49}$$

Using (39) through (46) allows us to write this out. Exactly as in I and II, we find that the field equations imply that g_{ij} must be of the form

$$g_{ij}=Q\eta_{ij}, \tag{50}$$

where Q is a constant and $\eta_{ji}=(-C^{-1})_{ji}$. C is the charge conjugation matrix and is the only constant Fermi sector matrix which is available to play this role. Conclusion (50) holds *only* in the horizontal lift basis and is another way that this basis simplifies calculations. After considerable simplification, the field equations then can be written as

$$^{(E)}R_{\mu\nu}+\tfrac{1}{2}Q\eta_{ji}F^{j}_{\delta\mu}F^{i}_{\nu\beta}g^{\delta\beta}+\tfrac{1}{2}g_{ab}F^{a}_{\delta\mu}F^{b}_{\nu\beta}g^{\delta\beta}=0, \tag{51}$$

$$F^{a}_{\nu\mu;a}=0, \tag{52}$$

$$F^{j\alpha}{}_{\mu;\alpha}=0, \tag{53}$$

$$F^{i\mu\alpha}F^{j}_{\mu\alpha}=0, \tag{54}$$

$$F^{a\alpha}{}_{\nu;\alpha}=0, \tag{55}$$

$$F^{a\pi\alpha}F^{j}_{\pi\alpha}=0, \tag{56}$$

$$^{(E'')}R_{ab}-\tfrac{1}{4}g_{ae}g_{bd}F^{e\pi\alpha}F^{d}{}_{\alpha\pi}=0, \tag{57}$$

where

$$^{(E)}R_{\mu\nu}\equiv\Gamma^{\alpha}_{\mu\nu,\alpha}-\Gamma^{\alpha}_{\mu\alpha,\nu}+\Gamma^{\delta}_{\mu\nu}\Gamma^{\alpha}_{\delta\alpha}-\Gamma^{\alpha}_{\mu\beta}\Gamma^{\beta}_{\alpha\nu} \tag{58}$$

is the Einstein curvature tensor and

$$\begin{aligned}&^{(E'')}R_{ab}\\&\equiv\Gamma^{d}_{ab,d}-\Gamma^{e}_{ae,b}+\Gamma^{e}_{ab}\Gamma^{d}_{ed}-\Gamma^{e}_{ad}\Gamma^{d}_{eb}+\Gamma^{d}_{ae}f^{e}_{bd}\\&=\tfrac{1}{4}f^{e}_{ad}f^{d}_{eb}.\end{aligned} \tag{59}$$

The semicolon denotes a covariant derivative where, for example,

$$\begin{aligned}F^{a}_{\nu\mu;a}\equiv&F^{a}_{\nu\mu,a}+F^{b}_{\nu\mu}\Gamma^{a}_{ba}-F^{a}_{\pi\mu}\Gamma^{\pi}_{\nu a}-F^{a}_{\nu\pi}\Gamma^{\pi}_{\mu a}\\=&F^{a}_{\nu\mu,a}+F^{b}_{\nu\mu}\Gamma^{a}_{ba},\end{aligned} \tag{60}$$

using (46) and the antisymmetry of $F^{b}_{\nu\mu}$. Equations (51) and (52) arise from the symmetric and antisymmetric parts of the original $R_{\mu\nu}$ in (49), respectively. (53) arises from $R_{\mu i}=0$ and so forth. These field equations are completely covariant.

In addition to these equations, we must also satisfy the Jacobi identities. Using (60) and the fact that $\Gamma^{e}_{ae}=f^{e}_{ea}$, we see that field equation (52) is equivalent to Jacobi identity (33). (50) implies that (36) is satisfied identically. Thus, we have only (34) and (35) remaining of the Jacobi identities.

If we choose to work in more conventional units, we can rescale

$$F^{j}_{\delta\mu}\to(16\pi\kappa/c^{4})^{1/2}F^{j}_{\delta\mu} \tag{61}$$

and

$$F^{a}_{\delta\mu}\to(16\pi\kappa/c^{4})^{1/2}F^{a}_{\delta\mu}. \tag{62}$$

We can manipulate (51) into the form

$$^{(E)}R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}\,^{(E)}R-\tfrac{1}{2}g_{\mu\nu}\,^{(E'')}R^{a}{}_{a}=(-8\pi\kappa/c^{2})T_{\mu\nu}, \tag{63}$$

where the energy–momentum tensor is given by

$$\begin{aligned}-c^{2}T_{\mu\nu}=&\eta_{ji}F^{j}_{\mu\delta}F^{i\delta}_{\nu}-\tfrac{1}{4}g_{\mu\nu}\eta_{ji}F^{j}_{\alpha\beta}F^{i\alpha\beta}\\&+g_{ab}F^{a}_{\mu\delta}F^{b}{}_{\nu}{}^{\delta}-\tfrac{1}{4}g_{\mu\nu}g_{ab}F^{a}_{\beta\delta}F^{b\beta\delta},\end{aligned} \tag{64}$$

and we let $Q=1$ in (50). The other field equations remain unchanged under this rescaling except for (57) which becomes

$$^{(E'')}R_{ab}-(4\pi\kappa/c^{4})g_{ae}g_{bd}F^{e}_{\delta\beta}F^{d\beta\delta}=0. \tag{65}$$

We notice that $^{(E'')}R^{a}{}_{a}$ plays the role of a cosmological constant in (63). From (59) and (18), we have

$$^{(E'')}R^{a}{}_{a}=\tfrac{1}{4}g^{a}{}_{a}/W=n/(4W), \tag{66}$$

where n is the dimensionality of the group space. We see that we can make $^{(E'')}R^{a}{}_{a}$ as small as we like by picking the scale factor W in the group metric sufficiently large. If W is taken to be large, we see that (65) requires

$$g_{ae}g_{bd}F^{e}{}_{\delta\beta}F^{d\beta\delta}\to 0. \tag{67}$$

We end up with the set of equations (52)–(56), (63), (64), (67), and the Jacobi identities (34) and (35). If $F^{a}_{\mu\nu}$ is set equal to zero, we return to the equations of I and II, which correctly describe a spin-3/2 field $F^{i}_{\mu\nu}$ coupled to general relativity in a rather unfamiliar representation. This spin-3/2 field arises quite naturally from the Fermi part of the typical fiber. If $F^{a}_{\mu\nu}$ is not set equal to zero, the complete set of equations describes a nonabelian gauge field correctly coupled to the spin-2 graviton and spin-3/2 gravitino fields. The part of the energy–momentum tensor (64) concerning the $F^{a}_{\mu\nu}$ field is correct,[10] (52) [or (33)] correctly describes the gauge covariance of the $F^{a}_{\mu\nu}$, and (55) gives the correct field equation for $F^{a}_{\mu\nu}$ in the present sourceless case. Equation (67) is all right

for these massless, sourceless nonabelian gauge fields and is analogous to the statement $F^{\alpha\beta}F_{\alpha\beta} \equiv \vec{E}^2 - \vec{B}^2 = 0$ for solutions of the massless, sourceless Maxwell equations.

The only unexpected equation to turn up is (56) which couples the gauge field to the gravitino field. In the U(1) case, this would couple photons to spin-3/2 gravitinos in an unacceptable way. We must also remember, however, that massless spin-3/2 fields have not been observed in nature. Through spontaneous symmetry breaking, this field presumably will grow a superheavy mass. Because of the small range of such a massive gravitino, the effects of (56) will be felt only at very high energies. Nonetheless, (56) may prove important in future work, in which we must address spontaneous symmetry breaking and the renormalizability of the theory. Spontaneous symmetry breaking may be a problem in this self sourced theory since no scalar fields are now present, and it is not at all clear where they can come from in a natural way. If supersymmetry can somehow be accomodated in this theory, other work on supergravity and gauge supersymmetry suggests that this theory should be more renormalizable than general relativity. Otherwise, renormalizabilty may be a problem also.

As in I and II, we can work out the Bianchi identities and we find

$$(-1)^{A'}(R^{AB} - \tfrac{1}{2}g^{AB}R)_{;A} \equiv 0, \tag{68}$$

where now this represents $8 + n$ equations (4 Bose + 4 Fermi + n gauge indices). We can easily check that the number of unknown field components minus the number of independent field equations gives the number of degrees of freedom expected for the present case.

To conclude, we find that a semisimple Lie group can be incorporated into our fiber bundle model of the gravitational field with relatively little trouble. Thus the foundations for an extended theory with a physically realistic particle content have been laid.

[1] D. K. Ross, Proc. Stony Brook Supergravity Conference (North-Holland, Amsterdam) to be published 1980; D. K. Ross, J. Math. Phys., preceeding paper. We denote these papers as I and II in the text.
[2] P. Nath and R. Arnowitt, Phys. Lett. B **56**, 177 (1975).
[3] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976); D. Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D **14**, 912 (1976).
[4] D. Z. Freedman and B. De Wit, Nucl. Phys. B **130**, 105 (1977); E. Cremmer, J. Scherk, and J. Schwarz, Phys. Lett. B **84**, 83 (1979).
[5] D. Z. Freedman, Phys. Rev. Lett. **38**, 105 (1977).
[6] J. Sherk, E. Cremmer, and S. Ferrara, Phys. Lett. B **74**, 61 (1978).
[7] R. Arnowitt and P. Nath, Phys. Rev. D **15**, 1033 (1977).
[8] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).
[9] The reader is referred to the following for general mathematical background. S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964); S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963); R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds* (Academic, New York and London, 1965); W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories* (Springer, New York, 1977).
[10] B. S. De Witt, *Dynamical Theory of Groups* (Gordon and Breach, New York, 1965), p. 137.

A note on lattice random walks with an excluded point

George H. Weiss
National Institutes of Health, Bethesda, Maryland 20205

(Received 21 April 1980; accepted 31 October 1980)

We derive some asymptotic results for the moments of displacement of a random walk on a lattice with a single absorbing point. In one dimension we are able to find an asymptotic expression for the probability distribution of random walks that have not been trapped.

PACS numbers: 0520.Gg, 02.50.Ey

Rubin[1] appears to have been the first to investigate random walks on a lattice with the origin excluded, i. e., which never return to their starting point. Although he derived an expression for the generating function of such random walks he presented results in detail only for the mean-square displacement when the starting point of the random walk is symmetrically located around the origin. Subsequently, Montroll[2,3] developed equation for a generalization of the model to study statistics of the first passage time till trapping. In this note we present more detailed results for the moments and distribution of displacement in the Rubin–Montroll model on an infinite lattice.

The random walk will be assumed to take place on a translationally invariant homogeneous lattice. The single step transition probabilities will be denoted by $p(\mathbf{j})$ with the corresponding structure function

$$\lambda(\boldsymbol{\theta}) = \sum_{\mathbf{j}} p(\mathbf{j}) \exp(\mathbf{j}\cdot\boldsymbol{\theta}). \tag{1}$$

Let the excluded point be $\mathbf{s}$, let $U_n(\mathbf{r})$ be the probability that the random walker is at $\mathbf{r}$ at step n on the lattice with the excluded point, and let $P_n(\mathbf{r})$ be the state probability on a regular lattice, i.e., one without an excluded point. The $U_n(\mathbf{r})$ satisfy the recursion relation

$$U_{n+1}(\mathbf{r}) = \sum_{\mathbf{l}} U_n(\mathbf{l})p(\mathbf{r}-\mathbf{l}) - U_n(\mathbf{s})p(\mathbf{s}-\mathbf{r}), \quad \mathbf{r}\neq\mathbf{s} \tag{2}$$

$$U_{n+1}(\mathbf{s}) = \sum_{\mathbf{l}} U_n(\mathbf{l})p(\mathbf{s}-\mathbf{l}) + U_n(\mathbf{s})[1-p(\mathbf{0})].$$

Hence if we define the generating function

$$U(\boldsymbol{\theta};z) = \sum_{n=0}^{\infty}\sum_{\mathbf{r}} U_n(\mathbf{r})z^n \exp(i\mathbf{r}\cdot\boldsymbol{\theta}), \tag{3}$$

we find, following Rubin[1] and Montroll,[2,3] that

$$U(\boldsymbol{\theta};z) = \frac{1}{1-z\lambda(\boldsymbol{\theta})} + \frac{P(\mathbf{s};z)}{(1-z)P(\mathbf{0};z)}\left(1 - \frac{(1-z)}{1-z\lambda(\boldsymbol{\theta})}\right)e^{i\mathbf{s}\cdot\boldsymbol{\theta}}, \tag{4}$$

where

$$P(\mathbf{s};z) = \frac{1}{(2\pi)^D}\int_{-\pi}^{\pi}\cdots\int \frac{\exp(-i\mathbf{s}\cdot\boldsymbol{\theta})}{1-z\lambda(\boldsymbol{\theta})}\, d^D\boldsymbol{\theta} \tag{5}$$

in D dimensions. Equation (2) implies the result

$$\sum_{n=0}^{\infty} U_n(\mathbf{r})z^n \equiv V(\mathbf{r};z) = P(\mathbf{r};z) - \frac{P(\mathbf{s};z)P(\mathbf{r}-\mathbf{s};z)}{P(\mathbf{0};z)}, \quad \mathbf{r}\neq\mathbf{s},$$

$$V(\mathbf{s};z) = \frac{P(\mathbf{s};z)}{(1-z)P(\mathbf{0};z)}. \tag{6}$$

It is readily verified from these relations that since

$$\sum_{\mathbf{r}} P(\mathbf{r};z) = 1/(1-z), \tag{7}$$

it follows that the $V(\mathbf{r};z)$ satisfy the same relation. Equation (6) now allows us to determine the asymptotic distribution of the U_n for n large. Let us first note that if $F_n(\mathbf{s})$ is the probability that the random walker reaches $\mathbf{s}$ for the first time at step n, that

$$\sum_{n=0}^{\infty} F_n(\mathbf{s})z^n = F(\mathbf{s};z) = P(\mathbf{s};z)/P(\mathbf{0};z), \quad \mathbf{s}\neq\mathbf{0}. \tag{8}$$

When $z=1$ we know that $F(\mathbf{s};1)=1$ for $D=1,2$ and $F(\mathbf{s};1)<1$ for $D\geqslant 3$. Maradudin, *et al*[4] and Joyce [5–8] have determined the analytic form for $F(s;z)$ when $|z-1|$ is small and

$$\sum_{\mathbf{r}} r_i^2 p(\mathbf{r}) < \infty, \quad i = 1,2,\cdots,D, \tag{9}$$

and Lindenberg, *et al.*[9] provided similar results for $D>3$. In the present analysis we make the assumption of finite variance transitions as in Eq. (9), and also that in the absence of traps, the average displacement in a single step is zero. Under these hypotheses the contribution from the first term on the right hand side of Eq. (4) is clearly the Gaussian distribution that appears in the theory of unrestricted random walks. In order to determine the nature of the remaining contribution we must study the properties of $F(\mathbf{s};z)$ as $z\to 1$.

In the case $D=1$ it is straightforward to calculate moments of the distribution by using Eq. (4) as a generating function. The probability that a random walk remains untrapped at step n, call it G_n, has the generating function

$$\sum_{n=0}^{\infty} G_n z^n = 1 - F(\mathbf{s};z)/(1-z). \tag{10}$$

The properties of this generating function have been established by Lindenberg, *et al.*[9] who showed that the asymptotic form of G_n *is*

$$G_n \sim |s|\sqrt{2}/(\sigma\sqrt{\pi n}), \tag{11}$$

where σ is the standard deviation of the single step transition probabilities. One can easily verify from Eq. (4) that

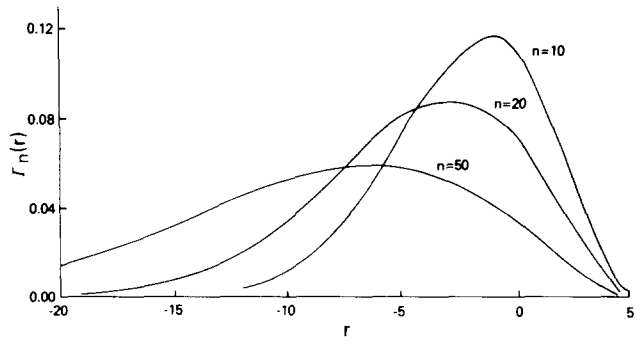


FIG. 1. Curves of the probability distribution of the end-to end distance of a random walk on a 1 − D lattice with a trapping point at $s = 4$ and $n = 10,20$, and 50. The random walker is initially at the origin.

$$\sum_{r=-\infty}^{\infty} rV(r;z) = 0. \tag{12}$$

This implies that the mean position of a random walker conditional on its not having been trapped is

$$\langle r(n)\rangle_u = -s(1-G_n)/G_n \sim -\sigma(\pi n/2)^{1/2}\operatorname{sgn} s. \tag{13}$$

That is to say, the mean position of untrapped walks reflects the fact that they must move steadily away from the trapping point in order to survive. A similar calculation shows that the leading term of the conditional variance is

$$\sigma_n^2 \sim 2\sigma^2 n, \tag{14}$$

which is independent of the trapping point. Lower order terms do depend on s. The distribution of position of the untrapped random walker at step n is easily calculated by noticing that the joint generating function excluding the point s is

$$U(\theta;z) - U_s(\theta)e^{is\theta} = [1 - F(s;z)e^{is\theta}]/[1 - z\lambda(\theta)]. \tag{15}$$

Hence the conditional probability is

$$\Gamma_n(r) = \left[P_n(r) - \sum_{l=0}^{n} F_{n-l}(s)P_l(r-s)\right] \Big/ \sum_{l=n+1}^{\infty} F_l(s), \tag{16}$$

as is otherwise obvious. An approximation to $\Gamma_n(r)$ valid for large n can be obtained by using the continuum limits for $P_n(r)$ and $F_n(r)$, i.e.,

$$P_n(r) \sim \frac{1}{\sigma(4\pi n)^{1/2}} e^{-r^2/(4n\sigma^2)}, \quad F_n(s) \sim \frac{|s|}{\sigma^3(4\pi n)^{3/2}} e^{-s^2/(4n\sigma^2)}. \tag{17}$$

In Fig. 1 we show curves of $\Gamma_n(r)$ for $\sigma = 1$, $s = 4$, and $n = 10$, 20, and 50. As n increases the maximum of the curve shifts to the left (for fixed s) as one would expect from Eq. (13).

In dimensions greater than one we may use Eq. (10) to show that

$$G_n \sim \frac{2\pi\sigma_1\sigma_2[P(\mathbf{0};1) - P(\mathbf{s};1)]}{\ln n} \tag{18}$$

so that in two dimensions, using the asymptotic form for $G_n(\mathbf{s})$ derived by Lindenberg, *et al*,[9]

$$\langle r_j(n)\rangle = -\frac{-s_j \ln n}{2\pi\sigma_1\sigma_2[P(\mathbf{0};1) - P(\mathbf{s};1)]}. \tag{19}$$

Just as in one dimension the surviving random walks are those that move away from the trap, but at a slower rate. In $D \geqslant 3$ dimensions, since $\lim G_n(\mathbf{s}) = G(\mathbf{s}) < 1$ the untrapped random walkers are influenced by the presence of the trap, but their average position does not move off to ∞ as $n \to \infty$. The variances and covariances of position can also be calculated from Eq. (4). Since we have assumed symmetry in the underlying transition probabilities, the generating function of second order moments is

$$\sum_{n=0}^{\infty} \langle r_l(n)r_j(n)\rangle z^n = \left.\frac{-\partial^2 U}{\partial\theta_l\partial\theta_j}\right|_{\theta=0} = \frac{\delta_{jl}\sigma_j^2 z[1 - F(\mathbf{s};z)]}{(1-z)^2}. \tag{20}$$

In $D \geqslant 3$ dimensions $\lim_{z\to 1} F(\mathbf{s},z) = F(\mathbf{s}) < 1$, which is the probability of being trapped at $\mathbf{s}$. Hence the elements of the variance–covariance matrix are asymptotic to

$$\langle r_l(n)r_j(n)\rangle \sim \delta_{jl}\sigma_j^2[1 - F(\mathbf{s})]n. \tag{21}$$

This expression is valid for all random walks including those trapped at $\mathbf{s}$. If we consider only the untrapped random walks, we find

$$\langle r_l(n)r_j(n)\rangle_u - \langle r_l(n)\rangle_u\langle r_j(n)\rangle_u \sim \delta_{jl}\sigma_j^2 n, \tag{22}$$

so that to leading order the variance of position is unaffected by the trapping point,and covariances are $O(n)$. This result agrees with Rubin's more detailed calculation. The extension of these results to study the effects of a finite number of excluded points is formally simple,[2] but the development rapidly leads to tedious algebra.

[1] R. J. Rubin, J. Math. Phys. **8**, 576 (1967).
[2] E. W. Montroll, J. Phys. Soc. Jpn. Suppl., **26**, 6 (1969).
[3] E. W. Montroll, J. Math. Phys. **10**, 753 (1969).
[4] A. A. Maradudin, E. W. Montroll, G. H. Weiss, R. Herman and H. W. Milnes, *Green's Functions for Monatonic Simple Cubic Lattices*(Acad. Roy. de Belgique, Brussels,1960).
[5] G. S. Joyce, J. Phys. **C 4**, L53 (1971).
[6] G. S. Joyce, J. Phys. **A 5**, L65 (1972).
[7] G. S. Joyce, J. Math. Phys. **12**, 390 (1971).
[8] G. S. Joyce, Phil. Trans. R. Soc. London, Ser. A **273**, 585 (1973).
[9] K. Lindenberg, V. Seshadri, K. E. Shuler, and G. H. Weiss, J. Stat. Phys. **23**, 11 (1980).

Lattice Green's functions for a linear chain with next nearest neighbor interactions

A. A. Bahurmuz and P. D. Loly
Department of Physics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada

(Received 29 May 1980; accepted for publication 17 October 1980)

The lattice Green's functions for a linear chain with next nearest neighbor (NNN) interactions,

$$G_l(t,A) = \frac{1}{\pi}\int_0^{\pi} \frac{\cos lx\, dx}{t + i\epsilon - (\cos x + A\cos 2x)}$$

are studied for various values of A, the ratio of NNN to nearest neighbor (NN) interactions. It is shown that $G_l(t,-A) = (-1)^{l+1}[G_l(-t,A)]^*$, thereby allowing attention to be restricted to the positive range of A which further divides into the two cases, (a) $0 \leqslant A \leqslant 1/4$, with band edge singularities, and, (b) $1/4 < A < \infty$, with an additional in-band singularity. Exact results are then obtained for the imaginary parts of G_l and for the real parts of G_0 and G_1. Finally a recurrence relation enables the real parts of G_l for arbitrary l to be found from those for $l = 0$ and 1.

PACS numbers: 05.50. + q, 71.20. + c, 63.10. + a

I. INTRODUCTION

It is well known that lattice Green's functions (LGF) play an important role in a broad range of phenomena in solid state physics, and accordingly, their evaluation has been the subject of extensive studies.[1] The last decade has seen substantial progress in the analytic treatment of those for two- and three-dimensional lattice problems with nearest neighbor (NN) interactions.[2] In the general case, characterized by longer ranged interactions, one has to resort to the tedious and approximate numerical methods of multiple integration for which there are nowadays a number of quite practical schemes.[3] The inclusion of next nearest neighbor (NNN) interactions to the NN problem provides a simple and physical way of overcoming[4] special features of the nearest neighbor cases, e.g., the logarithmic singularities of the NN BCC and FCC tight binding densities of states, and also provides an enriched singularity structure to the LGF's. It is with these thoughts that the present authors studied the inclusion of NNN interactions in the linear chain and found the unique opportunity for the exact solution of the LGF's for an arbitrary ratio of the NNN and NN interactions.

Section 2 presents some general characteristics of these LGF's including their critical points, associated van Hove singularities,[5] and a recursion relation which shows that one need only evaluate the first two LGF's in order to obtain all others. Section 3 gives the derivation of the exact expressions for the LGF's and the conclusions are given in Sec. 4.

II. GENERAL PROPERTIES OF LGF'S

The general form of the LGF for a one-dimensional system with NNN interactions is

$$G_l(t,A) = \frac{1}{\pi}\int_0^{\pi} \frac{\cos lx\, dx}{t + i\epsilon - (\cos x + A\cos 2x)}, \tag{1}$$

where ϵ is a positive infinitesimal and A is the ratio of NNN to NN interactions. Although l may take any integral value, we can derive the following recursion relation from Eq. (1) by multiplying the integrand by $t - (\cos x + A\cos 2x)$ and rearranging with trigonometric identities:

$$\delta_{l0} = tG_l - \tfrac{1}{2}(G_{l+1} + G_{l-1}) - \tfrac{1}{2}A\,(G_{l+2} + G_{l-2}). \tag{2}$$

It then follows that we need only evaluate G_l for $l = 0$ and 1 since the remaining LGF's can be obtained from these two by successive application of Eq. (2) and the fact that $G_{-l} = G_l$. Furthermore, we need only consider positive values of A since the LGF's for negative A are related to those for positive A by:

$$G_l(t,-A) = (-1)^{l+1}[G_l(-t,A)]^*, \tag{3}$$

where * stands for complex conjugate. Equation (3) may be easily obtained from Eq. (1) by making the substitution $x = \pi - y$.

Before we derive expressions for G_l we discuss briefly the critical points of the dispersion function and the associated van Hove singularities.[5] The present dispersion function,

$$F(x) = \cos x + A\cos 2x, \tag{4}$$

has a qualitatively different behavior for $A > 1/4$ than for $A < 1/4$ as shown in Fig. 1. Critical points are defined as those points in the Brillouin Zone (BZ) where the group velocity dF/dx vanishes. The resulting singularities in the spectral functions are known as van Hove singularities. For $A < 1/4$, the dispersion function is monotonic with turning points, maximum and minimum, at $x = 0$ and π respectively. For $A > 1/4$, $F(x)$ has maxima at $x = 0$ and π and a minimum in between at $x_0 = \cos^{-1}(-1/4A)$. The values of $F(x)$ at the turning points are as follows:

$$\begin{aligned} F(0) &= 1 + A = t_2, \\ F(\pi) &= -1 + A = t_1, \\ F(x_0) &= -(1 + 8A^2)/8A = t_0. \end{aligned} \tag{5}$$

For $A = 1/4$ it follows that $x_0 = \pi$, $t_0 = t_1$, and at this point $F(x)$ has a point of inflexion. The band represented by $F(x)$ extends from t_1 to t_2 for $A \leqslant 1/4$ and from t_0 to t_2 for $A > 1/4$. The van Hove singularities for $A < 1/4$ are at the band edges whereas for $A > 1/4$ there is an additional singularity within the band at t_1.

 022-2488/81/030564-05$1.00

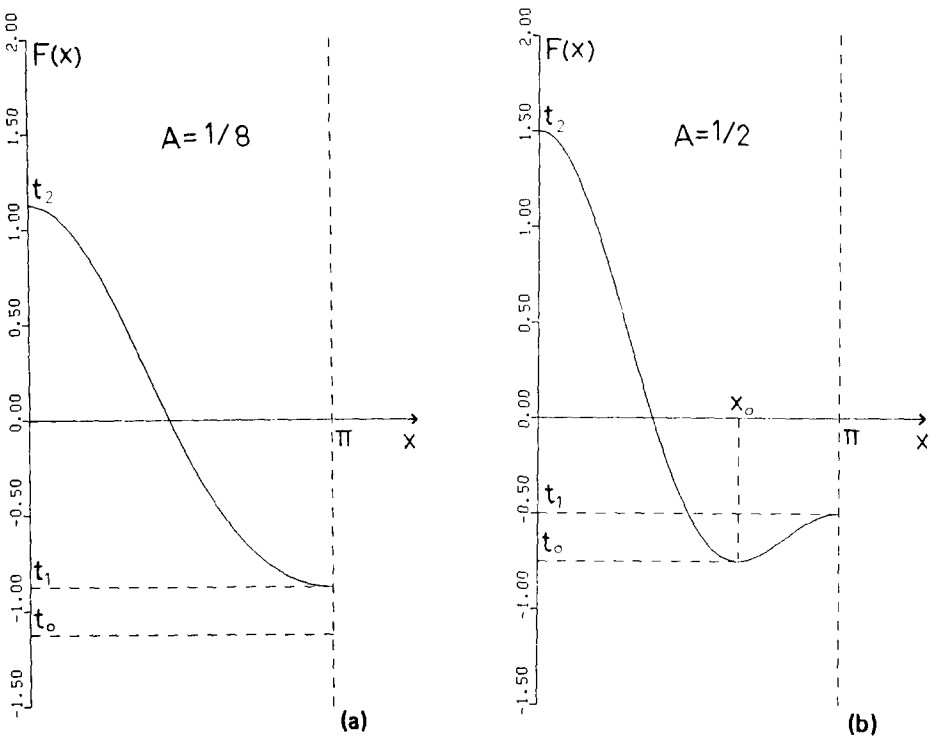


FIG. 1. Dispersion function $F(x) = \cos x + A\cos 2x$ for the linear chain with NNN interactions.

III. EXACT EXPRESSIONS FOR LGF'S

To derive expressions for the real and imaginary parts of G_l, we first rewrite the denominator in the integrand of Eq. (1) as follows:

$$g(x) = t - (\cos x + A\cos 2x) = -2A\,(\cos x - a_+)(\cos x - a_-), \tag{6}$$

where

$$a_\pm = \{-1 \pm [D(t)]^{1/2}\}/4A \tag{7}$$

and

$$D(t) = 1 + 8A\,(A + t) = 8A\,(t - t_0). \tag{8}$$

We note that since $D(t)$ is negative for $t < t_0$, $a_\pm$ are real for $t \geqslant t_0$ and a pair of complex conjugates for $t < t_0$. As we will see, the variation of $a_\pm$ with t is crucial in understanding the behavior of G_l. Figure 2 shows this variation and we distinguish three types of behavior: (a) $0 \leqslant A < 1/4$, (b) $A = 1/4$ and (c) $A > 1/4$. For (a) and (b), $|a_+| \leqslant 1$ within the band $(t_1 \leqslant t \leqslant t_2)$ while for (c), $|a_+| \leqslant 1$ throughout the band $(t_0 \leqslant t \leqslant t_2)$ and $|a_-| \leqslant 1$ within part of the band $(t_0 \leqslant t \leqslant t_1)$.

(i) Imaginary part of G_l

From Eq. (1), the imaginary part of G_l is given by

$$\mathrm{Im}G_l(t,A) = -\int_0^\pi dx\,\cos lx\,\delta\{g(x)\} = -\sum_n \frac{\cos lx_n}{|g'(x_n)|}, \tag{9}$$

where

$$g'(x) = dg/dx = 2A\,(1 - \cos^2 x)^{1/2} \times [(\cos x - a_+) + (\cos x - a_-)] \tag{10}$$

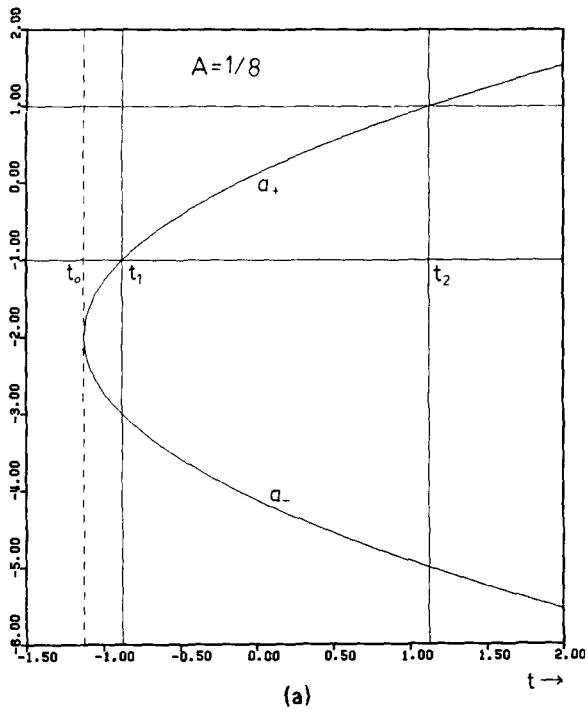
and where x_n are the roots of $g(x) = 0$. From Eq. (6) we see that the two possible roots are given by $\cos x_\pm = a_\pm$ and, since $|\cos x| \leqslant 1$, roots exist only for $|a_\pm| \leqslant 1$. Hence from Fig. 2, we see that for $A \leqslant 1/4$, a_+ is the only root and occurs within the band $(t_1 \leqslant t \leqslant t_2)$ while for $A > 1/4$, the a_+ root exists throughout the band $(t_0 \leqslant t \leqslant t_2)$ and the a_- root exists within part of the band $(t_0 \leqslant t \leqslant t_1)$. Hence, using Eqs. (9) and (10) we obtain the following expressions for the imaginary part of G_l

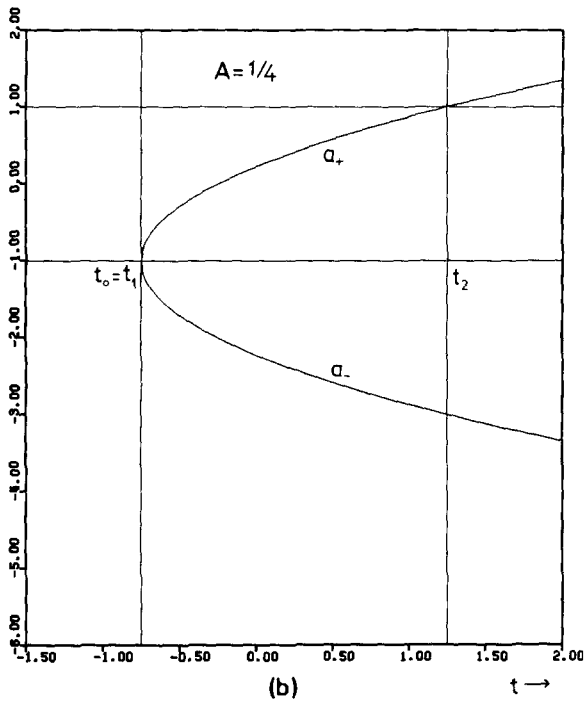
$$A \leqslant 1/4: \mathrm{Im}G_l(t,A) = -\frac{\cos lx_+}{D^{1/2}(1 - a_+^2)^{1/2}},\quad t_1 \leqslant t \leqslant t_2 \tag{11}$$

$$A > 1/4: \mathrm{Im}G_l(t,A) = -\frac{\cos lx_+}{D^{1/2}(1 - a_+^2)^{1/2}},\quad t_1 \leqslant t \leqslant t_2 \tag{12a}$$

$$= -\frac{\cos lx_+}{D^{1/2}(1 - a_+^2)^{1/2}} - \frac{\cos lx_-}{D^{1/2}(1 - a_-^2)^{1/2}},\quad t_0 \leqslant t \leqslant t_1 \tag{12b}$$

where $\cos x_\pm = a_\pm$. Outside the band $\mathrm{Im}G_l = 0$. The behavior of $\mathrm{Im}G_l$ within the band is shown in Fig. 3. For $A < 1/4$, the only poles occur when $|a_+| \to 1[(1 - a_+^2)^{1/2} \to 0]$ and we see from Fig. 2a that this occurs at t_1 and t_2. Near the pole, for $t = t_1 + \delta$ or $t = t_2 - \delta$ (where δ is small and positive) the behavior is inverse square root, i.e., $\mathrm{Im}G_l \sim \delta^{-1/2}$. For $A > 1/4$, the poles occur when $(1 - a_\pm^2)^{1/2}$ or $D^{1/2}$ goes to zero and the divergencies are also of the form $\delta^{-1/2}$ for $t = t_0 + \delta$ $(D^{1/2} \sim \delta^{1/2})$, $t = t_1 - \delta$ $[(1 - a_-^2)^{1/2} \sim \delta^{1/2}]$, and $t = t_2 - \delta$ $[(1 - a_+^2)^{1/2} \sim \delta^{1/2}]$. We note that for $A = 1/4$, we have $t_0 = t_1$ and the divergence around t_0 is stronger than inverse square root as in this case





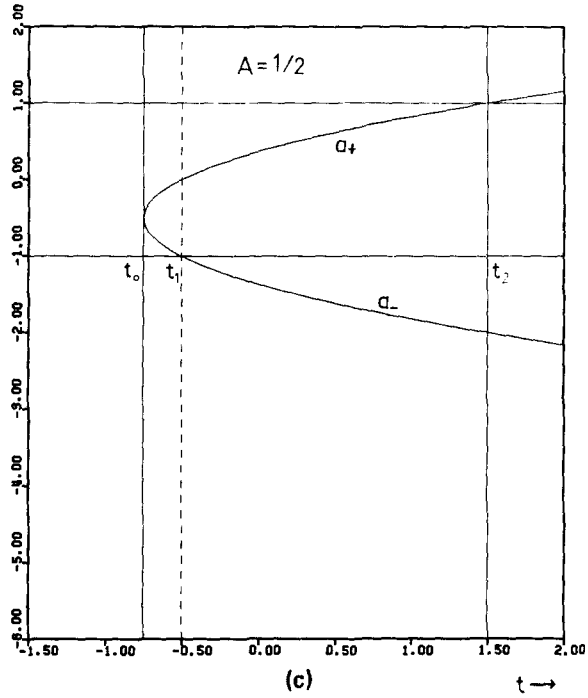


FIG. 2. Variation of $a_\pm$ with t [Eq. (7)].

$D^{1/2}\sim\delta^{1/2}$ and $(1-a_+^2)^{1/2}\sim\delta^{1/4}$ giving a $\delta^{-3/4}$ variation for Im G_l.

(ii) Real part of G_l

From Eqs. (1) and (6), the real part of G_l is given by

$$\mathrm{Re}G_l(t,A) = -D^{-1/2} \times\left[\frac{P}{\pi}\int_0^{\pi} dx\left(\frac{\cos lx}{\cos x - a_+} - \frac{\cos lx}{\cos x - a_-}\right)\right], \tag{13a}$$

where P stands for principal part. As mentioned in the previous section, a recursion relation [Eq. (2)] allows one to obtain all the LGF's from those for $l=0$ and 1. Hence, we only evaluate the real part of G_l for these two cases, i.e.

$$\mathrm{Re}G_0(t,A) = -D^{-1/2} \times\left[\frac{P}{\pi}\int_0^{\pi} dx\left(\frac{1}{\cos x - a_+} - \frac{1}{\cos x - a_-}\right)\right] \tag{13b}$$

and, with the help of $\cos x = (\cos x - a_\pm) + a_\pm$.

$$\mathrm{Re}G_1(t,A) = -D^{-1/2} \times\left[\frac{P}{\pi}\int_0^{\pi} dx\left(\frac{a_+}{\cos x - a_+} - \frac{a_-}{\cos x - a_-}\right)\right]. \tag{13c}$$

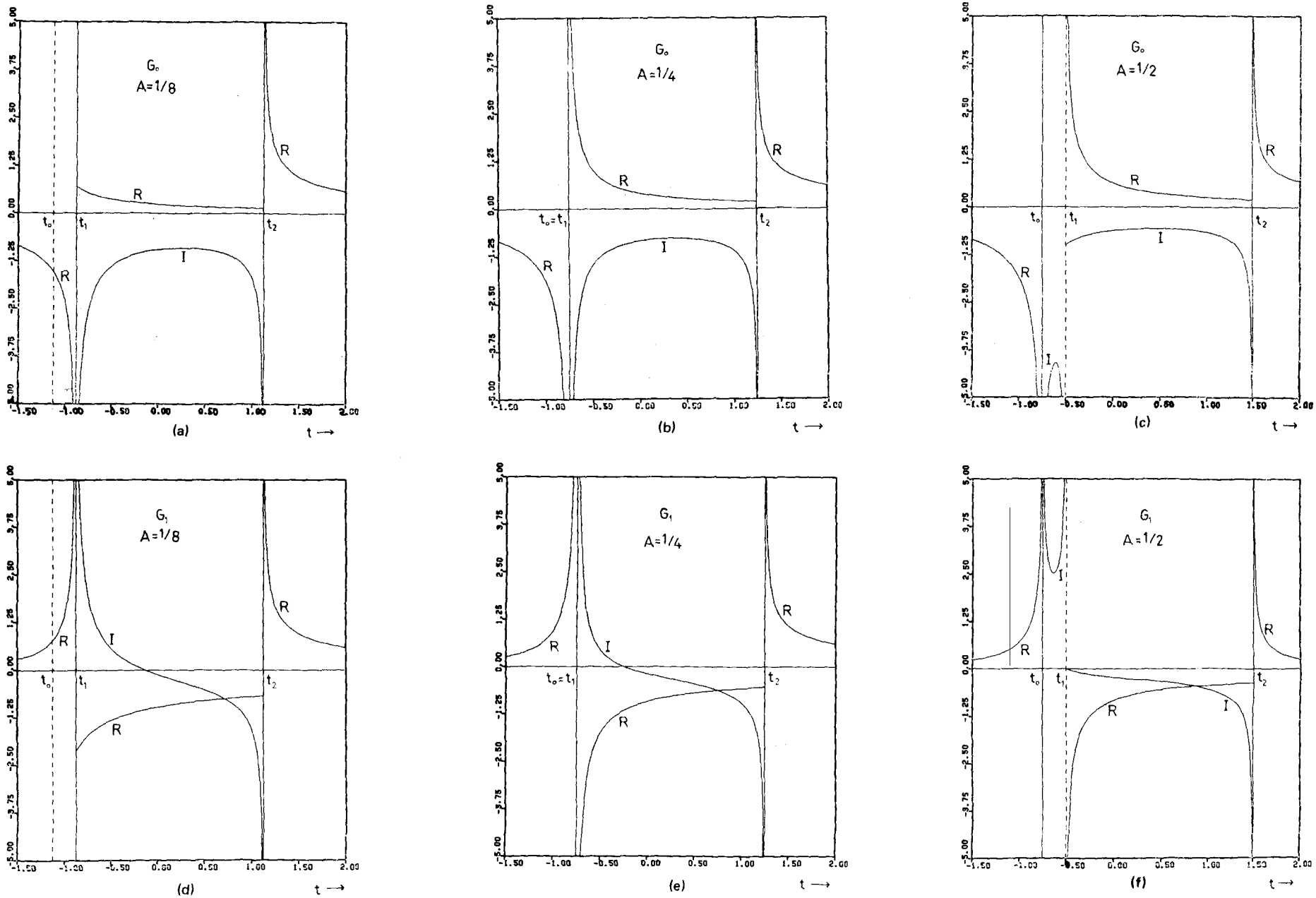


FIG. 3. Real (R) and imaginary (I) parts of G_0 and G_1.

Thus we only need to evaluate integrals of the form

$$I(a) = \frac{P}{\pi}\int_0^{\pi} \frac{dx}{\cos x - a}. \tag{14}$$

For $t \geqslant t_0$, a is real and I is a standard integral and has the well-known form

$$\begin{aligned} I(a) &= 0 \text{ for } |a| \leqslant 1 \\ &= -(a^2-1)^{-1/2} \text{ for } a > 1 \\ &= (a^2-1)^{-1/2} \text{ for } a < -1. \end{aligned} \tag{15}$$

For $t < t_0$, a is complex and the integral in Eq. (14) may be evaluated by a contour integration method. Making the substitution $y = \tan(x/2)$ in Eq. (14) gives

$$I(a) = -\frac{1}{\pi(1+a)}\int_{-\infty}^{\infty} \frac{dy}{y^2 - \gamma^2}, \tag{16}$$

where

$$\gamma^2 = (1-a)/(1+a). \tag{17}$$

We can rewrite Eq. (16) as a contour integral in the complex z plane as follows:

$$I(a) = -\frac{1}{\pi(1+a)}\oint_C \frac{dz}{z^2 - \gamma^2}, \tag{18}$$

where the contour C is shown in Fig. 4. The poles in the integrand are at $z = \pm\gamma$ and the residues are $\pm(2\gamma)^{-1}$ respectively. For the contour chosen, we only pick up the $z = \gamma$ pole and hence, using the residue theorem, we find

$$I(a) = \frac{1}{i(1+a)\gamma} = (a^2-1)^{-1/2}. \tag{19}$$

Substituting the above into Eq. (13) gives the following expressions for the real part of G_l (for $l = 0$ and 1 only):

$$\begin{aligned} t \geqslant t_2\text{: } \mathrm{Re}G_l &= D^{-1/2}[a_+^l (a_+^2 - 1)^{-1/2} \\ &\quad + a_-^l (a_-^2 - 1)^{-1/2}] \quad \text{for } l = 0,1, \end{aligned} \tag{20}$$

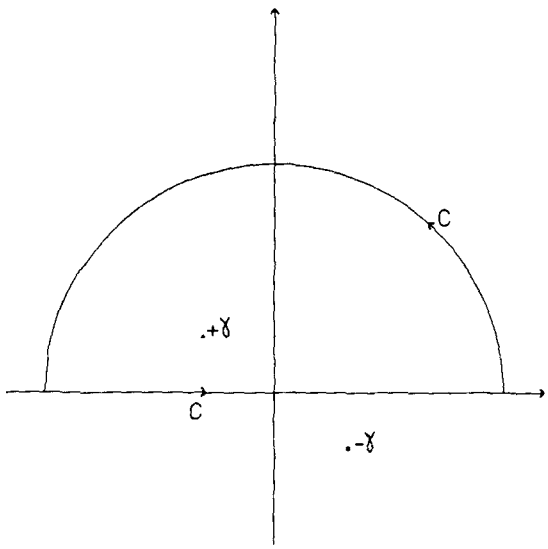


FIG. 4. Contour for the integral in Eq. (18).

$$t_1 \leqslant t \leqslant t_2\text{: } \mathrm{Re}G_l = D^{-1/2} a_-^l (a_-^2 - 1)^{-1/2} \text{ for } l = 0,1, \tag{21}$$

$$\begin{aligned} t_0 \leqslant t \leqslant t_1\text{: } \mathrm{Re}G_l &= D^{-1/2}[a_-^l (a_-^2 - 1)^{-1/2} \\ &\quad - a_+^l (a_+^2 - 1)^{-1/2}] \\ &\quad \text{for } A < 1/4,\ l = 0,1 \\ &= 0 \text{ for } A > 1/4,\ l = 0,1, \end{aligned} \tag{22}$$

$$\begin{aligned} t \leqslant t_0\text{: } \mathrm{Re}G_l &= D^{-1/2}[a_-^l (a_-^2 - 1)^{-1/2} \\ &\quad - a_+^l (a_+^2 - 1)^{-1/2}] \text{for } l = 0,1. \end{aligned} \tag{23}$$

We note that for $t < t_0$, $a_\pm$ are complex conjugates, $D^{1/2}$ is pure imaginary and hence the right-hand side of Eq. (23) is real. The variation of $\mathrm{Re}G_l$ is shown in Fig. 3 and again the diverging singularities are inverse square root ($\delta^{-1/2}$) except for the special case of $A = 1/4$ where the divergence at t_0 is $\delta^{-3/4}$.

IV. CONCLUSIONS

Equations (11), (12a) and (12b) give the imaginary parts of all the LGF's for the NNN linear chain, while Eqs. (20)–(23) give the real parts of the first two LGF's ($l = 0$ and 1) which provide the input for the recurrence relation, Eq. (2), from which higher orders can be obtained. We note that in the limit $A \to 0$ we have $D \to 1$, $a_+ \to t$ and both t_0 and $a_- \to -\infty$ so that Eqs. (11) and (20)–(23) reduce to the NN case. It is worth noting that in the NN case G_0 is purely imaginary (and symmetric) inside the band.

The present results have applications to a range of studies concerned with the behavior of excitations in anisotropic systems that have strong one-dimensional character from either linear or layered components. The present authors are applying these LGF's to study the problem of two spin excitations in the NNN chain for all possible pair wave vectors and for the entire spectrum[6] as an extension of Majumdar's work[7] which studied bound states at special wave vectors.

ACKNOWLEDGMENTS

The authors thank Dr. B. W. Southern for several discussions. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[1]See for example S. Katsura, T. Morita, S. Inawashiro, T. Horiguchi and Y. Abe, J. Math. Phys. **12**, 892 (1971), and references therein.
[2]G. S. Joyce, J. Phys. C **1**, 1510 (1971); **4**, L53 (1971); J. Phys. A **5**, L65 (1972); T. Horiguchi, J. Math. Phys. **13**, 1411 (1972); M. Inoue, J. Math. Phys. **15**, 704 (1974); **16**, 111 (1975); **16**, 809 (1975); T. Horiguchi and T. Morita, J. Phys. C **8**, L232 (1975).
[3]See for example G. Gilat, *Methods in Computational Physics*, edited by G. Gilat (Academic, New York, 1976), Vol. 15, p. 317 and references therein.
[4]P. D. Loly and M. Buchheit, Phys. Rev. B **5**, 1986 (1972); P. D. Loly, Phys. Rev. B **8**, 4405 (1973).
[5]L. van Hove, Phys. Rev. **89**, 1189 (1953).
[6]A. A. Bahurmuz and P. D. Loly, in preparation.
[7]C. K. Majumdar, J. Math. Phys. **10**, 177 (1969).

Aesthetic field theory—Extended particles in four-dimensional space–time

M. Muraskin
Physics Department, University of North Dakota, Grand Forks, North Dakota 58201

(Received 8 May 1980; accepted for publication 25 September 1980)

We continue our study of complex aesthetic field theory using a set of group theoretical data introduced previously. Our more detailed study shows significant differences with the corresponding real theory. We find that our theory allows for a four-dimensional particle system, where the particle system is viewed as a bound state of a three-dimensional maximum and minimum. The magnitude of the field is large in this "confluence" region. This property persists in time. We can think of this as a form of nonattenuation. We never see a large magnitude "free" maximum or minimum (confinement). A problem with the particle system is that the confluence region extends to a greater degree in z for large $|t|$ compared to the case at $t = 0$.

PACS numbers: 11.10.Cd, 11.10.St

I. INTRODUCTION

This article is a direct continuation of the paper "Aesthetic Field Theory: The Problem of Spatial Inversions".[1] A set of data $\Gamma^{\alpha}_{\beta\gamma}$ was chosen there, with the property of being invariant under three-dimensional rotations. Another set of data can be obtained from this set by means of a spatial inversion. Both sets of data were then used together in a complex version of aesthetic field theory.

In our previous paper[1] we observed that maps at $t = 0$, $z = 0$ showed results that do not look much different from what we have seen before in a real theory.[2] We have continued our study of this group theoretical data and found the similarities with real theory that we observed to be superficial. Further studies showed deep-seated differences with respect to particle structure.

The equations and notation used in this paper will be the same as in Ref. 1.

We shall review in the next section our previous work with respect to particle solutions of the aesthetic field theory.

II. PARTICLE BEHAVIOR IN AESTHETIC FIELD THEORY

In the past we have put forth considerable effort in investigating what the aesthetic field equations imply. The focus of attention has been in studying particle type solutions of these equations.

We first obtained particle solutions in Ref. 2. Here we saw a three-dimensional maximum and minimum. At $t = 0$ the maximum and minimum were well separated. However, as time went on the maximum and the minimum approached one another. As they grew closer together the magnitude of the maximum and the minimum increased markedly. There was a small region at the time (or is it time interval) of impact in which we could no longer follow the motion because of this. This small region of space in which the maximum and minimum were essentially on top of each other we have since called a "confluence" region.[3] After the scattering the maximum and minimum separated from one another and the magnitude of the field fell off sharply.

Although it was still possible to follow the particles around in time, nevertheless we can say that the particles underwent "attenuation."

We realize that in the world we live in particles like the electron show an incredible stability. They do not fade into the background as in the solution we found. Zabusky and Kruskal[4] showed that such a thing as nonattenuating particle solutions exist (solitons), arising from nonlinear field equations.

The question then is whether the aesthetic field equations are capable of particle solutions for which attenuation is not a problem (at least for a reasonable duration of time). (In our computer work we can only talk of nonattenuation for a reasonably long time interval. In our subsequent discussions this is what we shall mean when we talk of nonattenuation.)

We have found recently[3] that such stable particle solutions exist in complex null aesthetic field theory. Our particle system is a bound state of a maximum center and a minimum center. Unlike Ref. 2 the confluence region does not exist for a single time (or short duration), but instead moves along a path in time. The fields become quite large in magnitude in the confluence region. We say fields rather than field since all the Γ^{i}_{jk} have large magnitudes in this region. From a practical point of view we have not been able to find exactly how large the magnitudes become. The errors become too big as the size of the numbers grow. When we say "large" we mean (at least) several orders of magnitude greater than the environment, with the possibility that the numbers are many, many orders of magnitude greater than the surroundings. For example, coming into the confluence region with a 0.0003 grid, the computer printed overflow, which means numbers in excess of 10^{75}. Of course we cannot believe such numbers due to large errors. For this reason in Fig. 3 we drew a box around the confluence region. This procedure was also done in Ref. 3. Nonattenuation, for us, does not mean that the values of the field stays the same—it shall mean instead that the numbers remain large.

The drawback with our results in Ref. 3 was that such a particle system described above has only been found in a three-dimensional theory.

In addition to bound systems of two particles we found bound systems of 3 particles (particle refers to a maximum or minimum in the field) in Ref. 3. The two and three particle system that make up the bound state do

 022-2488/81/030569-03$1.00

not separate in time the way they did in Ref. 2. We shall call this effect "confinement."

In our three-dimensional theory[3] we found that the bound system moved, so far as we could tell, on a straight line (in a course sense). From our maps in Ref. 3 as well as in this paper we see many curved contour lines. It is not as if all lines are straight.

What happened in the complex null theory described in Ref. 3 when we tried to extend our work to four-dimensions? We found that the confluence region extended in z in such a way that we can find no end to the confluence region in our computer work. Either we are dealing with extremely large particles or particle interpretation is not possible in four dimensions. The possibility of the confluence region closing upon itself leading to a "topological" particle is intriguing. However, no evidence has been presented supporting such a solution.

In this paper we shall study the consequences of the group theoretical data presented previously [Eqs. (1) and (6) of Ref. 1] with respect to particle properties, in four-dimensional space–time.

III. DESCRIPTION OF THE SOLUTION

We have mapped A^1_{11} for convenience. At the confluence all the components have large magnitudes. We are dealing with a cooperative effect.

Figure 1 is a map of A^1_{11} at $t=0$, $z=-4$ (using 0.3 course grid). We see a planar maximum and a planar minimum. They are clearly separated. The magnitude of the maximum and minimum are not large (roughly 0.43 and 0.07, respectively). In Fig. 2 we have a map at $t=0$, $z=10$ with a course 0.5 grid. Again the maximum and minimum are separated and the value of the maximum and minimum are not great. However, we find as we progress up in z, starting with $z=-4$, that the planar maximum and the planar minimum become closer and closer to one another. Eventually they are "essentially" on top of each other. Here the magnitudes are extremely great. We find this to be the case in the region $x=3.64$, $y=-0.14$, $z=3.33$. We call this region a "confluence" region. In Fig. 3 we give a map of a "confluence" region at $t=-17.5$. The grid used here was 0.001. (The confluence region looks similar at different times so $t=-17.5$ was chosen for illustration.) In Ref. 3

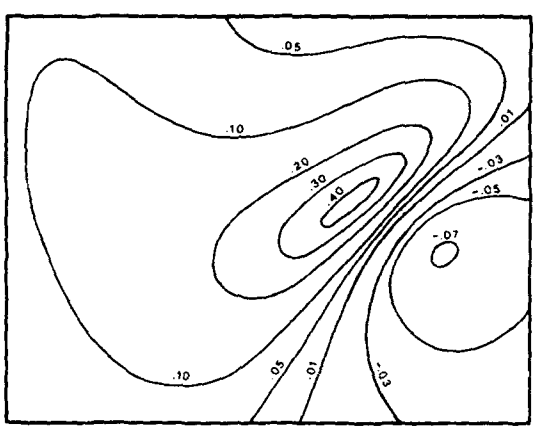


FIG. 1. Map of A^1_{11} around the origin at $t=0$, $z=-4$. The grid was 0.3.

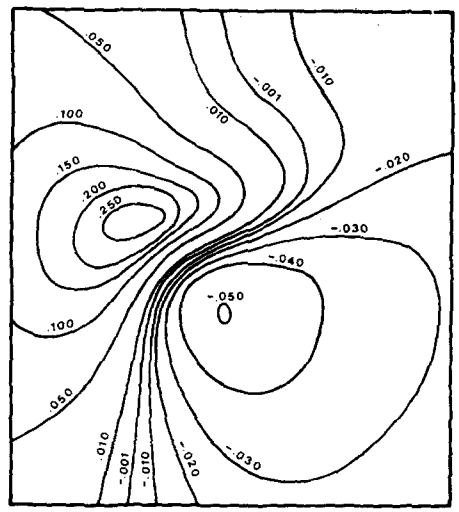


FIG. 2. Map of A^1_{11} around the origin at $t=0$, $z=10$. The grid was 0.5.

the confluence region extended in z without any end that we could find. In the current situation there is no doubt that the confluence region is bounded. There is a region in z for which the field has large magnitudes. The region, as determined with a large grid of 0.5, has dimensions $\Delta z < 3$.

Thus at $t=0$ we have a bounded particle system constructed from a three-dimensional maximum and minimum in close proximity. The question then is does this particle system persist in time? In Ref. 2 the maximum and minimum separated from one another and the magnitude diminished. However, in the present case we found that the magnitudes remain large as time went on. We found large magnitudes in the region around

$t=-10$	$t=-17.5$	$t=-104.5$
$x=11.5$	$x=17.44$	$x=86.34$
$y=0.92$	$y=1.78$	$y=11.76$
$z=10$	$z=15$	$z=73.33$

These points lie essentially on a four-dimensional straight line. However, we should caution against assigning a straight line for a world line since the large magnitudes extend in z as discussed previously. The large magnitudes are always in a region where large negative numbers are in close proximity to large positive numbers. We can think of our solution as a model

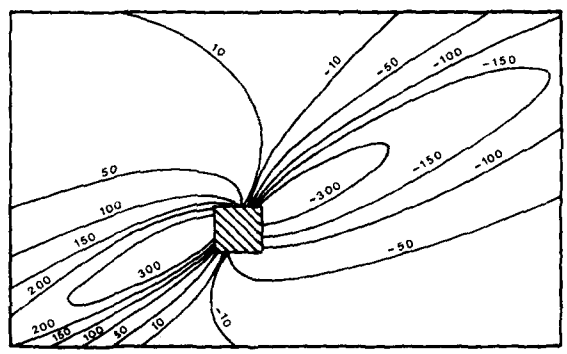


FIG. 3. Map of A^1_{11} around the confluence region at $t=-17.5$, $z=15$. The confluence region occurs within the box. The grid was 0.001.

for confinement. We have not seen a large magnitude "free" maximum or minimum. (We have yet to see even a small magnitude "free" maximum or minimum.) We can say we have a model for nonattenuation in the sense that the magnitude remains large at $t = -104.5$. Thus our bound system appears stable although we cannot say how long it will continue.

Are there any more particle systems like the one described above? We have some evidence to this effect. At $t = -17.5$ we notice again high positive numbers in close proximity to high negative numbers in the region of $x = 24$, $y = 5$, $z = 15$. This region was uncovered with a course grid of 0.5. As of now we have not confirmed any more potential particle systems. Computer time is a practical limitation in such a search.

There is a problem that should be noted. We find the confluence region extended in z much more at higher times than at $t = 0$. At $t = -17.5$ the confluence region extended in z such that $\Delta z < 16$ (a course determination). We found the spread of the confluence region is extremely small in x and y for a particular z.

IV. SUMMARY

The group theory data of this paper leads to a bound state system in four dimensions. The particle can be viewed as a bound state of a three-dimensional maximum and minimum. This picture persists in time. There are large magnitudes in the confluence region and only in the confluence region (where large negative numbers come in close proximity to large positive numbers). Large magnitudes have been found for $\Delta t \sim 100$ in our computer work. We shall call this effect nonattenuation.

The model we found for a particle is very different from, say, Ref. 4. There particles are not composite systems. In nature, the quark picture suggests that observed particles are composite, and the quark constituents do not exist freely. Our particle system has similarities to this quark picture. In Ref. 5 bound "kink-antikink" systems have been discussed in one spatial dimension.

Some questions still unanswered are: Can aesthetic field theory yield particle systems which are better localized? The variable (in time) extension of the confluence region could be a problem. Can the aesthetic field theory describe a multitude of particles? One or two is not yet a multitude. Are there curved trajectories in time? Can we find three particle bound states, as we did in null theory, in the case that the dimension of space-time is four?

The data (1) in Ref. 1 is invariant under three-dimensional rotations. Under reflection we obtain, from (1), the data (6), again found in Ref. 1. These two sets of data cannot be reached from one another by means of a continuous transformation. Thus, the question from a conceptual point of view is which data, (1) or (6) should be taken. There is no logical reason that favors one set over the other. The answer we have suggested is to take both sets, by allowing for the theory to be complex. A complex theory is consistent with the requirement that all derivatives as well as all tensors, be treated in a uniform manner so far as change is concerned. We have found that such hypotheses lead to an extended particle system in four dimensions.

Note added in proof: We use the same $f^{\alpha}{}_{i}$ and $h^{\alpha}{}_{i}$ as in Eqs. (21) and (22) of Ref. 1. However, Eq. (21) there should have read:

$$f^{0}{}_{1} = 0.44 \quad f^{0}{}_{2} = -0.16 \quad f^{0}{}_{3} = 0.39 .$$

[1] M. Muraskin, J. Math. Phys. 21, 1155 (1980).
[2] M. Muraskin and B. Ring, Found. Phys. 5, 513 (1975).
[3] M. Muraskin, Found. Phys. 10, 887 (1980).
[4] N. Zabusky and M. Kruskal, Phys. Rev. Lett. 15, 240 (1965).
[5] P. Caudrey, J. C. Eilbeck and J. Gibbon, Nuovo Cimento B 25, 497 (1975).

On the decoupling theorem in field theory

Edward B. Manoukian
Department of National Defence, Royal Military College of Canada, Kingston, Ontario K7L 2W3, Canada

(Received 15 July 1980; accepted for publication 22 August 1980)

An elementary proof of the decoupling theorem in field theory is given when all the masses in the theory are allowed to go to infinity. The proof is given directly in momentum space with the subtractions of the theory performed at the origin and containing nonzero mass particles. The theorem states that, in Euclidean space, if the masses of the theory are scaled by λ and λ is allowed to go to infinity then the corresponding Feynman amplitude vanishes not more slowly than λ^{-1}, i.e., $|\mathscr{A}|\leqslant\lambda^{-1}C$, for $\lambda\geqslant 1$.

PACS numbers: 11.10.Gh, 11.10.Jj, 11.10Ef, 02.30. — f

Several useful applications have been carried out in the literature (see e.g., Refs. 1,2) making use of the so-called decoupling theorem in field theory. When some of the masses in a theory, with nonzero mass particles, become large the proof of the theorem (in Euclidean space) is involved[3] due to the complex nature of the subtractions of renormalization and, to some extent, due to the complicated structure of the Feynman rules involved with higher spins and derivative couplings. An *elementary* and complete proof of this important theorem is given in this paper by working directly in momentum space with subtractions carried out at the origin and with nonzero mass particles for the cases when all the masses of the theory are allowed to go to infinity. The theorem states that, in Euclidean space, when the masses of the theory are scaled by a parameter λ and the latter is allowed to go to infinity, the renormalized Feynman amplitude vanishes. More precisely, the renormalized Feynman amplitude $\mathscr{A}$ does not vanish any more slowly than λ^{-1}, i.e., $|\mathscr{A}|\leqslant\lambda^{-1}C$, $\lambda\geqslant 1$, where C is a constant independent of λ. The simplicity of the proof we hope justifies this work.

A renormalized Feynman amplitude may be written in Euclidean space in the form:

$$\mathscr{A}(p\,M)=\int_{R^{4n}} dk R(p,k,M),$$
$$=M=(M_1,\ldots,M_s),\quad M_j>0, \tag{1}$$
$$p=(p_1^0,\ldots,p_m^3),k=(k_1^0,\ldots,k_n^3),$$
$$R(p,k,M)=\mathscr{G}(p,k,M)\prod_{j\in\mathscr{L}}(Q_j^2+M_j^2)^{-1},$$
$$Q_j=\sum_{i=1}^{n}a_{ji}k_i+\sum_{i=1}^{m}b_{ji}p_i\equiv K_j+P_j,$$

where M denotes the set of masses in the theory, $P(p,k,M)$ is a polynomial in its argument and $\mathscr{L}$ denotes the totality of all lines participating in the denominator of $R(p,k,M)$.

We write

$$P(p,k,M)=\sum_{\alpha}p^{\alpha}P_{\alpha}(k,M),$$
$$\alpha=(\alpha_{01},\ldots,\alpha_{3m}),$$
$$p^{\alpha}\equiv(p_1^0)^{\alpha_{01}}\cdots(p_m^3)^{\alpha_{3m}}, \tag{2}$$
$$d_{\mu j}\geqslant a_{\mu j}\geqslant 0,\quad j=1,\ldots,m;\quad \mu=0,1,2,3.$$

Let $p_{j0}^{\mu},\ldots,p_{jd_{\mu j}}^{\mu}$ be $(d_{\mu j}+1)$ distinct values for p_j^{μ}. Let

$$p_t^*=(p_{1t_{01}}^0,\ldots,p_{mt_{3m}}^3), \tag{3}$$

where $0\leqslant t_{\mu j}\leqslant d_{\mu j}$, then Lagrange's interpolating formula states[4,5] that we may find a constant $C_{\alpha}(p_t^*)$ depending on α and the fixed values in p_t^* such that

$$P_{\alpha}(k,M)=\sum_t C_{\alpha}(p_t^*)P(p_t^*,k,M), \tag{4}$$

where P and P_{α} are defined in (1) and (2), respectively, and the sums is overall $0\leqslant t_{\mu j}\leqslant d_{\mu j}$ with $j=1,\ldots,m$; $\mu=0,1,2,3$.

Let $D(p,k,M)=\Pi_{k\in\mathscr{L}}(Q_j^2+M_j^2)$. An elementary and useful inequality[4] is that

$$(Q_j^2+M_j^2)^{-1}A_j^{-1}\leqslant(K_j^2+M_j^2)^{-1}\leqslant(Q_j^2+M_j^2)^{-1}A_j, \tag{5}$$

where

$$A_j=1+|P_j|/M_j+P_j^2/M_j^2. \tag{6}$$

Using (4) and the right-hand side of the inequality (5), we may find a positive $G(p_t^*)$ depending on (p_t^*) such that

$$\left|\frac{P_{\alpha}(k,M)}{D(0,k,M)}\right|\leqslant\sum_t|C_{\alpha}(p_t^*)|\left|\frac{P(p_t^*,k,M)}{D(p_t^*,k,M)}\right|. \tag{7}$$

The inequality (7) states that the absolute convergence of $\int dk P(p_t^*,k,M)$ implies the absolute convergence of $\int dk P_{\alpha}(k,M)D^{-1}(0,k,M)$. This result is similar to the one in Lemma 6 of Ref. 4 except in the latter an expansion is made in powers of k rather than in p as we have done in (2).

By writing $\lambda M=(\lambda M_1,\ldots,\lambda M_s)$, we have

$$\mathscr{A}(p,\lambda,M)=(\lambda)^{d(G)}\int_{R^{4n}}dk R\left(\frac{p}{\lambda},k,M\right), \tag{8}$$

where $d(G)$ is the dimensionality of the graph G with which the renormalized Feynamn amplitude $\mathscr{A}$ is associated. From the left-hand side of the inequality (5) and Eq. (7) we have that $\int dk P_{\alpha}(k,M)D^{-1}(p/\lambda,k,M)$ is also absolutely convergent for all $\lambda>0$. By the Lebesgue dominated convergence theorem we may take the limit of $\lambda\to\infty$ inside the integral in (8) and we obtain from (2):

$$\lim_{\lambda\to\infty}\mathscr{A}(p,\lambda,M)$$
$$=\lim_{\lambda\to\infty}\sum_{\alpha}(\lambda)^{d(G)-|\alpha|}p^{\alpha}$$

 022-2488/81/030572-02$1.00

$$\times \int_{R^{4n}} dk P_\alpha(k, M) D^{-1}(p/\lambda, k, M)$$

$$= \lim_{\lambda \to \infty} \sum_\alpha (\lambda)^{d(G) - |\alpha|} p^\alpha$$

$$\times \int_{R^{4n}} dk P_\alpha(k, M) D^{-1}(0, k, M). \tag{9}$$

Now, quite generally

(i) if $d(G) < 0$, then $\alpha \geqslant 0$,

(ii) if $d(G) \geqslant 0$, then $\alpha \geqslant d(G) + 1$, by definition of the overall subtraction over G, and hence in all cases

$$\lim_{\lambda \to \infty} \mathscr{A}(p, \lambda, M) = 0 \tag{10}$$

Equation (9) and the inequality in (5) then gives the following estimate that

$$|\mathscr{A}(p, \lambda, M)| \leqslant \lambda^{-1} C, \quad \lambda \geqslant 1, \tag{11}$$

where

$$C = \prod_{j \in \mathscr{L}} \left(1 + \frac{|P_j|}{M_j} + \frac{P_j^2}{M_j^2}\right) \sum_\alpha |p^\alpha|$$

$$\times \int_{R^{4n}} dk \, |P_\alpha(k, M)| D^{-1}(0, k, M), \tag{12}$$

covering both cases $d(G) < 0$ and $d(G) \geqslant 0$.

[1] T. Appelquist and J. Carrazzonne, Phys. Rev. D 11, 2856 (1975).

[2] E. C. Poggio, H. R. Quinn, and J. B. Zuber, Phys. Rev. D **15**, 1630 (1977).

[3] J. Ambjørn, Commun. Math. Phys. **67**, 109 (1979). Unlike ours the proof in here is given in the so-called a-parameter representation of $\mathscr{A}$.

[4] Y. Hahn and W. Zimmerman, Commun. Math. Phys. **10**, 330 (1968).

[5] E. Issacson and H. Keller, *Analysis of Numerical Methods* (Wiley, New York, 1966).

On quantum solitons and their classical relatives: Spin $\frac{1}{2}$ approximation of the sine-Gordon system

Piotr Garbaczewski [a)]
Nordita, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

(Received 5 December 1979; accepted for publication 22 February 1980)

If gradient terms in the linear Bose chain Hamiltonian couple nearest neighbors only, then in the spin $\frac{1}{2}$ approximation they reduce to the spin $\frac{1}{2}$ x-y-z Hamiltonian. Hence, spin $\frac{1}{2}$ approximation reveals the Thirring model as the one, whose spectrum is completely included in the spectrum of the underlying Bose chain. Relation to the Coleman's equivalence is discussed.

PACS numbers: 11.10.Lm

1. INTRODUCTION

There are still missing points in the most recent studies of the famous Thirring-sine-Gordon model equivalence in $1+1$ dimensions. One knows here that the spin $\frac{1}{2}$ x-y-z Heisenberg chain provides an equivalent description of the massive Thirring model put on a linear (space direction) lattice. A continuum limit of the latter can be used to recover spectral properties of the quantum sine-Gordon system (Coleman's equivalence), within the appropriate limitations on the coupling constant values of the Bose model.[1-3]

However no one has satisfactorily investigated the question of a "classical limit" of the quantum sine-Gordon field, with the special emphasis on the relation of classical and quantum soliton fields. The only exception in this context were the semiclassical quantization methods of Ref. 4, and quite inconclusive Coleman's remarks in last sections of Ref. 1.

Second, suppose, we start from a classical sine-Gordon field energy density[5]

$$H(x,t)=\tfrac{1}{2}\left\{\left(\frac{\partial\phi}{\partial t}\right)^2-\left(\frac{\partial\phi}{\partial x}\right)^2+2m^2(1-\cos\phi)\right\}(x,t) \tag{1.1}$$

and approximate it on a linear lattice with spacing ϵ. Then

$$\begin{aligned}\int_{\mathbb{R}^1} H(x,t)\,dx &\to H\\ &=\sum_s\{[\pi_s^2+2m^2(1-\cos\phi_s)]-(\phi_s-\phi_{s+1})^2/\epsilon^2\}\\ &=\sum_s (H_s+V_{s,s+1})\end{aligned} \tag{1.2}$$

can be viewed to describe a linear chain of plane pendula subject to harmonic interactions among nearest neighbors.[6] The gradient terms $\{V_{s,s+1}\}_{s=0,\pm1,\dots}$ are here responsible for the emergence of nontrivial configurations in the pendular chain. In the "single site approximation" of H by $\Sigma_s H_s$, a quantization of the chain is immediate through a simple replacement of each classical pendulum in the chain by a respective quantum one. A corresponding Schrödinger problem,[7] involves a pair of variables $\{\pi=-i\hbar(\partial/\partial\phi),\phi\}$ which though infinitesimally canonical cannot be integrated to a representation of the CCR (canonical commutation relations) algebra, in the sense of Ref. 8.

As a consequence, there is no straightforward way of getting a quantum analog of gradient terms $\{V_{s,s+1}\}$, and probably it was the main reason why the lattice analog (1,2) of the classical sine-Gordon system, has never been explicitly related to the spin $\frac{1}{2}$ x-y-z Heisenberg chain. The latter can in principle be considered as a lattice ancestor of the quantum sine-Gordon system, and as such should somehow be related to the quantization of the classical lattice problem (1.2), which is a lattice descendant of the classical sine-Gordon system.

Our aim is to establish the underlying relation. Basic result of the paper can be summarized as follows: Quantum fields on a lattice are considered as functions of the single site level raising-lowering operators. *We prove that each linear Bose chain, whose gradient term in the Hamiltonian couples nearest neighbors only, in its spin $\frac{1}{2}$ approximation, is equivalent to the lattice Thirring model.* Specification of its coupling constant relies on the explicit form of the quantum field entering the gradient term. (For a particular case of the sine-Gordon field, each quantum soliton operator appears to give rise to its own Thirring problem.[8,9] A crucial point in this investigation is an interaction of a quantum system with a nonzero temperature reservoir, without which a spin $\frac{1}{2}$ approximation makes no sense.[10]

2. QUANTUM PENDULUM AND PENDULAR CHAIN

The quantum pendulum spectral problem is conventionally expressed in terms of the Mathieu equation:

$$(2q\cos 2z-d^2/dz^2)\psi(z)=a\psi(z), \tag{2.1}$$

with $q=m^2$, $\phi=2z$, $z\in[0,2\pi]$, $\psi=\psi(\phi)\in\mathscr{L}^2(0,4\pi)$ and a playing the role of the eigenvalue, compare, e.g., Ref. 7.

The spectrum of the quantum pendulum in the coupling constant range $q\in(0,\infty)$ is nondegenerate, and both eigenfunctions and eigenvalues exhibit a q-dependence. The Mathieu (eigen)functions:

$$\begin{aligned}&ce_{2n}(z\pm\pi)=ce_{2n}(z),\quad se_{2n+2}(z\pm\pi)=se_{2n+2}(z),\\ &ce_{2n+1}(z\pm\pi)=-ce_{2n+1}(z),\\ &se_{2n+1}(z\pm\pi)=-se_{2n+1}(z),\quad n=0,1,2,\cdots,\end{aligned} \tag{2.2}$$

[a)]Permanent Address: Institute of Theoretical Physics, University of Wroclaw, Cybulskiego 36, Poland.

 022-2488/81/030574-07$1.00

$$\frac{1}{\pi}\int_0^{2\pi} ce_k(z)ce_l(z)\,dz = \delta_{kl} = \frac{1}{\pi}\int_0^{2\pi} se_k(z)se_l(z)\,dz,$$
$$\int_0^{2\pi} ce_k(z)se_l(z)\,dz = 0, \tag{2.3}$$

form a complete orthonormal system in $h = \mathscr{L}^2(0,4\pi)$ and hence

$$h = \bigoplus_{n=0}^{\infty} h_n = \bigoplus_{n=0}^{\infty} (h_n^{ce} \oplus h_n^{se}) = h^{ce} \oplus h^{se}. \tag{2.4}$$

In the limit $q \to 0$, $E_0(q)$ falls down to its minimum E_0, while $E_k^{ce} \to E_k^{se} \to E_k(0), E_k(0) < E_{k+1}(0)$ for $k > 0$. In the opposite limit $q \to \infty$, the spectrum $\{E_k(q)\}_{k=0,1,\cdots}$ goes to that of the doubly degenerate harmonic oscillator with the identifications:

$$E_{2n}^{ce} \to E_{2n+1}^{ce} \to E_{2n}^{\infty}, \quad E_{2n+1}^{se} \to E_{2n+2}^{se} \to E_{2n+1}^{\infty},$$

for all $n = 0,1,\cdots$.

By introducing:

$$e_{4n} = (1/\sqrt{\pi})ce_{2n}, \quad e_{4n+1} = (1/\sqrt{\pi})ce_{2n+1},$$
$$e_{4n+2} = (1/\sqrt{\pi})se_{2n+1}, \quad e_{4n+3} = (1/\sqrt{\pi})se_{2n+2}, \tag{2.5}$$

we can define densely in h the pair of the raising and lowering operators a^*, a, for quantum pendulum

$$a^* = \sum_{k=0}^{\infty} \sqrt{k+1}\, e_{k+1} \otimes e_k,$$
$$a = \sum_{k=1}^{\infty} \sqrt{k}\, e_{k-1} \otimes e_k, \tag{2.6}$$

which generate in h a Fock representation of the CCR algebra

$$[a,a^*]_- = \mathbb{1} = \sum_n e_n \otimes e_n, \quad ae_0 = 0.$$

In terms of $\{e_n\}$ the quantum pendulum Hamiltonian becomes immediately diagonalized, which allows an expression of H in terms of the true (integrable) generators a^*, a of the CCR algebra

$$H = \sum_{n=0}^{\infty} E_n e_n \otimes e_n = \sum_n E_n a_n^*{:}\exp(-a^*a){:}a^n = H(a^*,a), \tag{2.7}$$

where we have exploited the fact that: $\exp(-a^*a)$: is a projection on the ground state e_0 in h.

On the other hand, one easily finds that the operators

$$\mathbb{S}^+ = a^* \frac{\cos^2(\pi a^*a/2)}{(a^*a+1)^{1/2}},$$
$$\mathbb{S}^- = \frac{\cos^2(\pi a^*a/2)}{(a^*a+1)^{1/2}} a, \tag{2.8}$$

generate in h a reducible representation of CAR algebra

$$[\mathbb{S}^-,\mathbb{S}^+]_+ = \mathbb{1} = \sum_n e_n \otimes e_n,$$
$$\mathbb{S}^- e_{2n} = 0, \quad \mathbb{S}^+ e_{2n} = e_{2n+1}, \quad \forall n = 0,1,\cdots, \tag{2.9}$$
$$(\mathbb{S}^-)^2 = 0 = (\mathbb{S}^+)^2,$$

which becomes reduced on each two-dimensional sector $h_n^{ce(se)}$ of h. In particular, if to denote P_0 a projection on h_0^{ce} in h, then:

$$\mathbb{S}_0^+ = P_0 \mathbb{S}^+ P_0 = a^*{:}\exp(-a^*a){:},$$
$$\mathbb{S}_0^- = P_0 \mathbb{S}^- P_0 = {:}\exp(-a^*a){:}a, \tag{2.10}$$

are identities on h_0^{ce}, and the spin $\frac{1}{2}$ operator $\mathbf{S}$, with $S^+ = \mathbb{S}_0^+, S^- = \mathbb{S}_0^-, S^2 = -\frac{1}{2} + \mathbb{S}_0^+ \mathbb{S}_0^-$, emerges at once. Let us now consider a linear chain of elementary quantum systems. It is characterized by a countable set $\{a_s^*,a_s\}_{s=0,\pm 1,\cdots}$ of the CCR algebra generators, which form a reducible representation in the general tensor product space

$$\mathscr{H} = \prod_s^{\otimes} (h)_s, \tag{2.11}$$

where with each site of the chain, we associate a copy of the quantum pendulum Hilbert space. An irreducible, Fock component of this CCR algebra arises in the proper subspace IDPS(Ω) of $\mathscr{H}$, where

$$\Omega = \prod_s^{\otimes} (e_0)_s, \quad [a_s,a_t^*]_- = \delta_{st}\mathbb{1},$$
$$[a_s,a_t]_- = 0 = [a_s^*,a_t^*]_-, \quad a_s\Omega = 0 \forall s. \tag{2.12}$$

By comparison with (2.7) we find that a correctly quantized form of the lattice sine-Gordon Hamiltonian, should be given as a nonlinear function of generators $\{a_s^*a_s\}$:

$$H \to \hat{H} = H(a^*,a) = \sum_s \left\{ \sum_k E_k a_s^{*k}{:}\exp(-a_s^*a_s){:}a_s^k + V_{s,s+1}(a^*,a) \right\}, \tag{2.13}$$

where, in principle, the gradient term should read

$$\hat{V}_{s,s+n} = V_{s,s+1}(a^*,a) = [(1/\epsilon)(\hat{\phi}_{s+1} - \hat{\phi}_s)]^2,$$
$$\hat{\phi}_s = \phi_s(a^*,a). \tag{2.14}$$

In Ref. 10, we have given a description of the quantum sine-Gordon field in terms of (single-site) canonical generators $\{a_s^*,a_s\}_{s=0,\pm 1,\cdots}$ and shown that each single classical soliton field ϕ, after quantization should give rise to its associated irreducibility sector IDPS(ϕ) of $\mathscr{H}$. On the other hand, in Ref. 10, we have formulated a model independent description of criteria under which expectation values of observables associated with a finite part of the lattice Bose system, can be made to converge to these of the associated Fermi (or spin $\frac{1}{2}$) system. A leading idea behind, was that of a "spin $\frac{1}{2}$ approximation concept" for quantum Bose systems in thermal bath. We have shown that if a projection

$$P_0 = \prod_s P_0^s = \prod_s \{{:}\exp(-a_s^*a_s){:} + a_s^*{:}\exp(-a_s^*a_s){:}a_s\} \tag{2.15}$$

on the lowest two levels of each single lattice degree of freedom, happens to be a spectral projection for a Bose system

$$[P_0,\hat{H}_B]_- = 0, \tag{2.16}$$

then for a finite fraction of lattice sites, one has

$$P_0 H_B(a^*,a) P_0 = H_F(\mathbb{S}^+,\mathbb{S}^-) = \hat{H}_F \tag{2.17}$$

and a resolution of the spectral problem for the spin $\frac{1}{2}$ lattice, partly (within $P_0\mathscr{H} = \mathscr{H}_0$) resolves a spectral problem for $\hat{H}_B$. *In connection with the time development, notice that for all vectors* $|a\rangle, |b\rangle \in P_0\mathscr{H} = \mathscr{H}_0$ *there holds*

$$\langle a|\exp(i\hat{H}_B t)|b\rangle = \langle a|\exp(i\hat{H}_F t)|b\rangle, \tag{2.18}$$

see, e.g. Ref. 10. *If* $|a\rangle \in \mathscr{H}_0$, *but* $|b\rangle \in \mathscr{H}$ *the identity* (18) *holds as well.* Because in Ref. 8 we have established a connection between classical solitons and irreducibility domains for the CCR algebra generated by $\{a_s^*, a_s\}_{s=0,\pm 1,\ldots}$ quantum solitons can be specified by identifying the corresponding IDPS-generating vectors. Hence, in the infinite volume limit, the spectral problem for $\hat{H}_F$, should indeed give rise to a classification of quantum solitons.

In the present papers we are concentrated on the relation between the classical sine-Gordon equation and the equivalent lattice approximations of the quantum sine-Gordon system: the spin $\frac{1}{2}$ *x-y-z* and Thirring model.

Studies of the Fermi contents of quantum solitons will not be studied here

3. SPIN $\frac{1}{2}$ APPROXIMATION OF GRADIENTS IN THE LINEAR BOSE CHAIN, AS SPIN $\frac{1}{2}$ *x-y-z* HEISENBERG PROBLEM

(i) Take an elementary quantum system (single site Schrödinger problem in the linear chain) in its two-level (spin $\frac{1}{2}$) approximation. Within a Hilbert space h it means that we preserve two lowest levels $\{e_0, e_1\}$ of the general spectral solution $\{e_k\}_{k=0,1,\ldots}$. We characterize them by indicating an orientation of the spin $\frac{1}{2}$ arrow: $S^2 e_0 = (-\frac{1}{2})e_0$, $S^2 e_1 = (\frac{1}{2})e_1$, or $e_0 = |0\rangle \equiv \downarrow$, $e_1 = |1\rangle \equiv \uparrow$, for each fixed excitation possibility.

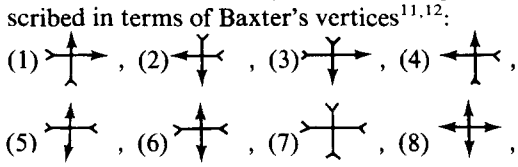
The set of all mappings between the excitation levels of the two-level problem can be described in terms of four arrow diagrams:

$$\frac{\uparrow}{\uparrow}, \frac{\uparrow}{\downarrow}, \frac{\downarrow}{\uparrow}, \frac{\downarrow}{\downarrow}.$$

Let us now consider the two chain neighbors, both in the spin $\frac{1}{2}$ approximation. Then a propagation of excitations along a linear chain can be completely given in terms of 16 arrow diagrams, describing elementary site-to-site energy exchanges $(s,\uparrow\downarrow) \longleftrightarrow (s+1,\uparrow\downarrow)$:

$$\begin{aligned}
&(1)\ \frac{\uparrow\ \uparrow}{\uparrow\ \uparrow}, \frac{\downarrow\ \downarrow}{\downarrow\ \downarrow}, \quad (2)\ \frac{\downarrow\ \uparrow}{\uparrow\ \downarrow}, \frac{\uparrow\ \downarrow}{\downarrow\ \uparrow},\\
&(2)\ \frac{\uparrow\ \downarrow}{\uparrow\ \downarrow}, \frac{\downarrow\ \uparrow}{\downarrow\ \uparrow}, \quad (4)\ \frac{\uparrow\ \uparrow}{\downarrow\ \downarrow}, \frac{\downarrow\ \downarrow}{\uparrow\ \uparrow},\\
&(5)\ \frac{\uparrow\ \uparrow}{\uparrow\ \downarrow}, \frac{\downarrow\ \downarrow}{\downarrow\ \uparrow}, \quad (6)\ \frac{\uparrow\ \downarrow}{\downarrow\ \downarrow}, \frac{\downarrow\ \uparrow}{\uparrow\ \uparrow},\\
&(7)\ \frac{\uparrow\ \downarrow}{\uparrow\ \uparrow}, \frac{\downarrow\ \uparrow}{\downarrow\ \downarrow}, \quad (8)\ \frac{\uparrow\ \uparrow}{\downarrow\ \uparrow}, \frac{\downarrow\ \downarrow}{\uparrow\ \downarrow}.
\end{aligned} \tag{3.1}$$

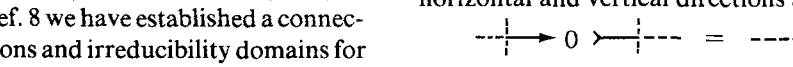
Each of these elementary transition diagrams can be described in terms of Baxter's vertices[11,12]:



(3.2)

provided, we adopt the Kadanoff-Wegner convention[13]: the right or upward arrow in the Baxter's diagram corresponds to the case when the two adjacent (with an arrow in between) spins are parallel, while a leftward or downward arrow is assigned to antiparallel spins.

We shall be composing transition diagrams in both horizontal and vertical directions according to the rule.



(3.3)

and basically interesting for us will be the transitions between spin-up-spin-down configurations of a linear chain of the identical two-level systems, which can be described in terms of so-called "transition matrices":

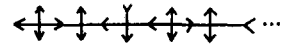
$$T(|\text{conf } 1\rangle) = |\text{conf } 2\rangle, \tag{3.4}$$

like e.g., the mapping

$$\begin{array}{c}
\cdots\ s\ s+1\ s+2\ s+3\ s+4\ s+5\ \cdots \\
\cdots\ |\uparrow|\uparrow|\uparrow|\downarrow|\downarrow|\downarrow|\ \cdots
\end{array}
\begin{array}{c}
\equiv \\
\equiv
\end{array}
\begin{array}{l}
\cdots|1\rangle|1\rangle|1\rangle|0\rangle|0\rangle|0\rangle\cdots \\
\cdots e_1\otimes e_1\otimes e_1\otimes e_0\otimes e_0\otimes e_0\cdots
\end{array} \tag{3.5}$$

$$\begin{array}{c}
\cdots\ s\ s+1\ s+2\ s+3\ s+4\ s+5\ \cdots \\
\cdots\ |\downarrow|\uparrow|\downarrow|\uparrow|\downarrow|\uparrow|\cdots
\end{array}
\begin{array}{c}
\equiv \\
\equiv
\end{array}
\begin{array}{l}
\cdots e_0\otimes e_1\otimes e_0\otimes e_1\otimes e_0\otimes e_1 \\
\cdots|0\rangle|1\rangle|0\rangle|1\rangle|0\rangle|1\rangle\cdots,
\end{array}$$

which in the language of Baxter's diagrams reads



(3.6)

(ii) One immediately finds that whenever the two configurations of the infinitely long spin $\frac{1}{2}$ chain differ in the infinite number of single site entries, then the respective vectors $|\text{conf } 1\rangle$, $|\text{conf } 2\rangle$ are orthogonal within the direct product space $\mathscr{H}_0 = \Pi_s{}^{\otimes}(h_0)_s$, $(h_0)_s = P_0^s(h)_s$, P_0^s $= :\exp(-a_s^* a_s): + a_s^* :\exp(-a_s^* a_s): a_s$ projecting on the linear span of e_0, e_1 in $(h)_s$.

These two vectors give thus rise to the two unitarily inequivalent irreducibility sectors for the spin $\frac{1}{2}$ algebra

$$\begin{aligned}
&\{\mathfrak{S}_s^+, \mathfrak{S}_s^-\}_{s=0,\pm 1,\ldots},\\
&[\mathfrak{S}_s^\pm, \mathfrak{S}_t^\pm]_- = 0, \quad s \neq t,\\
&[\mathfrak{S}_s^+, \mathfrak{S}_s^-]_+ = P_0^s,
\end{aligned} \tag{3.7}$$

IDPS ($|\text{conf}1,2\rangle$), respectively. If not orthogonal, the two configurations are represented by product vectors within the same irreducibility sector.

Suppose we have chosen some configuration, i.e., $|\text{conf}\rangle$, and hence IDPS ($|\text{conf}\rangle$). Mappings between configurations within IDPS ($|\text{conf}\rangle$) influence at most a finite (though arbitrary) number of spins $\frac{1}{2}$ in the chain, and can be represented by operators acting in IDPS ($|\text{conf}\rangle$).

To investigate the structure of such configuration-to-configuration mappings within any fixed IDPS ($|\text{conf}\rangle$), we can exploit the original Baxter's study of transfer matrices for the eight-vertex model,[11,12] vertex configurations being just the Baxter's diagrams mentioned before.

(iii) To proceed along the Baxter's lines, we must admit a thermal coupling of the linear chain with the environment. (See, e.g., for the notion of a "field-reservoir interaction" of ours.[14]) Then we must assign to the vertices, the appropriate Boltzmann weights

$$\omega_1 = \omega_2 = a = \exp(-\beta\epsilon_1), \quad \omega_3 = \omega_4 = \exp(-\beta\epsilon_2), \tag{3.8}$$

$\omega_5 = \omega_6 = c = \exp(-\beta\epsilon_3)$, $\omega_7 = \omega_8 = d = \exp(-\beta\epsilon_4)$,

where energies $\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4$, we shall for a while leave unspecified, β being the inverse temperature of our open system.

Let us now form a related set of parameters:

$$w_1 = \tfrac{1}{2}(c+d), \quad w_2 = \tfrac{1}{2}(c-d),$$
$$w_3 = \tfrac{1}{2}(a-b), \quad w_4 = \tfrac{1}{2}(a+b), \tag{3.9}$$

and notice that transitions between different configurations of the N-particle segment of the linear chain, which are realized in a sequence of N steps, can be represented by the Baxter's "transfer matrix,"[11,13,3] on the Baxter's $N \times N$, eight-vertex lattice, provided the toroidal boundary conditions are taken into account. One knows that an operator representative T of the transfer matrix in IDPS ($|\text{conf}\rangle$) commutes with the spin $\frac{1}{2}$ x-y-z model Hamiltonian,[15,11,12] so that the most general Hamiltonian, responsible for quantum fluctuations around a fixed configuration $|\text{conf}\rangle$ of the chain is one of the form

$$H_{xyz} = -\sum_s \sum_a J_a S_s^a S_{s+1}^a, \tag{3.10}$$

where $a = x,y,z$.

An explicit form of the underlying H_{xyz} operator can be deduced for its N-site version (N arbitrary)[11]:

$$H_{xyz} = -J_z(\mathrm{sn}\zeta)\left\{\tfrac{1}{2}\sum_{s=1}^{N}\sum_a p_a' \mathbb{S}_s^a \mathbb{S}_{s+1}^a\right\}, \tag{3.11}$$

with

$$\mathbb{S}^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbb{S}^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \mathbb{S}^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$
$$J_z = J_x/J_y = \mathrm{cn}(2\zeta)/\mathrm{dn}(2\zeta), \tag{3.12}$$
$$\mathrm{cn}(2\zeta) = J_x/J_z,$$

and

$$\begin{aligned}
&P_1 = \tfrac{1}{2}(w_1 - w_2 - w_3 + w_4) \Rightarrow p_1' = \mathrm{cn}(2\zeta)/\mathrm{sn}(2\zeta),\\
&P_2 = \tfrac{1}{2}(-w_1 + w_2 - w_3 + w_4) \Rightarrow p_2' = \mathrm{dn}(2\zeta)/\mathrm{sn}(2\zeta),\\
&P_3 = \tfrac{1}{2}(-w_1 - w_2 + w_3 + w_4) \Rightarrow p_3' = 1/\mathrm{sn}(2\zeta),\\
&P_4 = \tfrac{1}{2}(w_1 + w_2 + w_3 + w_4) \Rightarrow p_4'\\
&\quad = (\mathrm{cn}(2\zeta) - \mathrm{dn}(2\zeta) - 1)/\mathrm{sn}(2\zeta),
\end{aligned} \tag{3.13}$$

$\mathrm{sn}(V,l)$, $\mathrm{cn}(V,l)$, $\mathrm{dn}(Vl)$, being the elliptic functions of the modulus l

$$\begin{aligned}
&l = [(s_3 - s_2)(s_4 - s_1)/(s_4 - s_2)(s_3 - s_1)]^{1/2}\\
&\quad \times [(J_z^2 - J_y^2)/J_z^2 - J_x^2)]^{1/2},\\
&\mathrm{sn}^2(v,l) + \mathrm{cn}^2(v,l) = 1,\\
&l^2\,\mathrm{sn}^2(v,l) + \mathrm{dn}^2(v,l) = 1,
\end{aligned} \tag{3.14}$$

where the particular value ζ of $V \in \mathbb{R}^+$ is chosen to be fixed by

$$\mathrm{sn}(\zeta,l) = [(s_3 - s_1)/(s_4 - s_1)]^{1/2}. \tag{3.15}$$

The four parameters $\{s_i\}_{i=1,2,3,4}$ are related to the initial Baxter's parameters $\{w_i\}_{i=1,2,3,4}$ through formulas:

$$w_j^2 = p(\xi - s_j), \tag{3.16}$$

where

$$\begin{aligned}
\xi = \xi(v) = [&(s_3 - s_1)s_4 - s_3(s_4 - s_1)\,\mathrm{sn}^2(v,l)]/\\
&\cdot[s_3 - s_1 - (s_4 - s_1)\,\mathrm{sn}^2(v,l)]
\end{aligned} \tag{3.17}$$

and p is one more free parameter of the theory.

(iv) At $\{w_j\}$ fixed, both p, and $\{s_j\}$ exhibit the V-dependence and the nonuniqueness of their choice can be removed by assuming that we evaluate them as the initial $V = 0$ data corresponding to $\{w_j\}_{j=1,2,3,4}$ and at a fixed value of the constant ζ. Then a connection (3.16) between $\{w_j\}$ and $\{s_j\}$ is unique.

Let us here emphasize an important Baxter's observation[11] that the two transfer operators $T_{\{s\}}$, $T_{\{s'\}}$ commute if $\{s_j\} = \{s_j'\}$. It implies that by varying a single parameter $\zeta \in \mathbb{R}^+$ of the theory we have classified all noncommuting transfer operators: $T_{\{s\}} := T_\zeta$.

Commuting transfer operators are associated with the same spin $\frac{1}{2}$ x-y-z Hamiltonian, hence a related one-parameter family of Heisenberg Hamiltonians $H_{xyz}(\zeta)$ emerges, each one determining its own spin $\frac{1}{2}$ algebra irreducibility sector IDPS(ζ). Notice that by fixing ζ, we have fixed l and:

$$\begin{aligned}
&J_x = \mathrm{cn}^2(2\zeta)/\mathrm{dn}(2\zeta),\\
&J_y = \mathrm{cn}(2\zeta)/\mathrm{dn}^2(2\zeta),\\
&J_z = \mathrm{cn}(2\zeta)/\mathrm{dn}(2\zeta),
\end{aligned} \tag{3.18}$$

i.e., the coupling constants of the H_{xyz} problem.

All this means that: If the Boltzmann weights a,b,c,d, of the eight-vertex problem, are once established and fixed, there exists a one-parameter $\zeta \in \mathbb{R}^+$ family of inequivalent spin $\frac{1}{2}$ x-y-z Heisenberg problems responsible for propagation of excitations along a linear chain of spins $\frac{1}{2}$ (or a linear Bose chain in the spin $\frac{1}{2}$ approximation)

$$s_j = s_j(\zeta) = f_j(\zeta,w), \quad j = 1,2,3,4, \tag{3.19}$$

and $\{w_i = w_i(a,b,c,d)\}_{i=1,2,3,4}$ everthing at a fixed inverse temperature β of the reservoir.

The nonuniqueness can be removed if $\{a,b,c,d,\zeta\}$ arise as a five-valued function of a single common parameter λ: $a = a(\lambda)$, $b = b(\lambda)$, $c = c(\lambda)$, $d = d(\lambda)$, $\zeta = \zeta(\lambda)$, as then $s_j = s_j(\lambda)$ is uniquely determined by giving the value of λ, at β fixed.

4. DETERMINATION OF BOLTZMANN WEIGHTS

(i) Within the Kadanoff–Wegner parametrization, the set a,b,c,d of Boltzmann factors can be deduced in terms of Ising parameters $\mathbb{S}_{jk} = \pm 1$, according to the general prescription:

$$\begin{aligned}
\exp(&K^*\mathbb{S}_{jk}\mathbb{S}_{j+1,k+1} + K^*\mathbb{S}_{j+1,k}\mathbb{S}_{j,k+1}\\
&+ \lambda\mathbb{S}_{j+1,k+1}\mathbb{S}_{j+1,k}\mathbb{S}_{j,k+1}\mathbb{S}_{jk}),
\end{aligned} \tag{4.1}$$

where j enumerates neighboring transition diagrams (rows of the Baxter's square lattice), while k enumerates the neighboring spins in the linear chain (columns of the Baxter's lattice). For each fixed (j,k)th transition diagram we have

$$\begin{aligned} -\beta\epsilon_1 &= K^+ + K^- + \lambda, \\ -\beta\epsilon_2 &= \lambda - (K^+ + K^-), \\ -\beta\epsilon_3 &= K^+ - K^- + \lambda, \\ -\beta\epsilon_4 &= -(K^+ - K^-) - \lambda, \end{aligned} \tag{4.2}$$

so that a complete partition function of the Baxter's lattice

$$\sum_{(\mathfrak{S}=\pm 1)} \prod_{j,k} \exp(K^+ \mathfrak{S}_{jk}\mathfrak{S}_{j+1,k+1} + K^- \mathfrak{S}_{j+1,k}\mathfrak{S}_{j,k+1} + \lambda \mathfrak{S}_{jk}\mathfrak{S}_{j+1,k+1}\mathfrak{S}_{j+1,k}\mathfrak{S}_{j,k+1}) \tag{4.3}$$

describes the two interpenetrating (crossed bonds) Ising lattices with Ising variables $\mathfrak{S}_{jk}, \mathfrak{S}_{j+1,k}$ attached at each (j,k)th site, and an interaction between the lattices arising due to each set of four nearest-neighbor spins, see e.g., Refs. 13, 16, and 17. Notice that the (j,k)th diagram:

$$\begin{array}{c|c} (j+1,k) & (j+1,k+1) \\ \hline (j,k) & (j,k+1) \end{array} \tag{4.4}$$

gives account of the crossed mappings between spin-up and spin-down states of the two neighboring spins $\frac{1}{2}$ in the linear chain.

Depending on elementary exchange energies during site-to-site interactions of the nearest chain neighbors, the spin $\frac{1}{2}$ approximation of the gradient terms occuring on the Bose Hamiltonian, gives rise to:

(a) single Ising system if either $K^+ \neq 0$,

$K^- = \lambda = 0$, or $K^- \neq 0, K^+ = \lambda = 0$;

(b) two independently living Ising systems if

$K^+ \neq 0, K^- \neq 0$, but $\lambda = 0$;

(c) the general spin $\frac{1}{2}$ x-y-z Heisenberg system in the either case.

It needs suitable limitations on the exchange energy values, like, e.g.:

(a)
$\epsilon_1 = -\epsilon_2 \Rightarrow K^+ = (\beta/2)(\epsilon_2 - \epsilon_3), K^- = (\beta/2)(\epsilon_2 + \epsilon_3)$, i.e., we need either $\epsilon_2 = \epsilon_3$ or $\epsilon_2 = -\epsilon_3$;

(b) $\epsilon_1 = -\epsilon_2$, but $\epsilon_2 \neq \pm \epsilon_3$.

(ii) Notice that a two-particle Hilbert space $(h_0)_s \otimes (h_0)_{s+1}$ is four-dimensional. With P_0 projecting on h_0 in h, we find that $V^0_{s,s+1} = P_0^s P_0^{s+1} V_{s,s+1} P_0^s P_0^{s+1}$ acts in $(h_0)_s \otimes (h_0)_{s+1}$ invariantly and can immediately be diagonalized, thus leading to the four real eigenvalues, which we identify with the exchange energies $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$, respectively.

The particular assignments of energy values of the transition diagrams, according to (3.1), (3.2), (3.8) rely on the form of respective eigenvectors. Let us denote:

$$\begin{aligned} |1\rangle &= e_0 \otimes e_0, \quad |2\rangle = \frac{1}{\sqrt{2}}(e_0 \otimes e_1 + e_1 \otimes e_0), \\ |3\rangle &= \frac{1}{\sqrt{2}}(e_0 \otimes e_1 - e_1 \otimes e_0), \quad |4\rangle = e_1 \otimes e_1. \end{aligned} \tag{4.5}$$

Let $\{\alpha_j^k\}$ be a complex 4×4 matrix:

$$\sum_k \bar{\alpha}_i^k \alpha_j^k = \delta_{ij}, \quad \sum_i \bar{\alpha}_i^k \alpha_i^l = \delta_{kl} \tag{4.6}$$

and let the four orthonormal vectors

$$|\alpha, j\rangle = \sum_i \alpha_j^i |i\rangle \tag{4.7}$$

be the eigenvectors of $V^0_{s,s+1}$:

$$V^0_{s,s+1} |\alpha, j\rangle = \epsilon_j |\alpha, j\rangle. \tag{4.8}$$

Then an expectation value

$$\epsilon_j = \langle \alpha, j | V^0_{s,s+1} | \alpha, j \rangle = \sum_{k,l} \bar{\alpha}_j^k \alpha_j^l \langle k | V^0_{s,s+1} | l \rangle \tag{4.9}$$

explicitly reveals which transition diagrams give a nonzero counterpart to ϵ_j. An identification is here immediate: $\langle k | V^0_{s,s+1} | l \rangle$ corresponds to the arrow diagram describing a transition from a configuration associated with $|k\rangle$, to this associated with $|l\rangle$.

(iii) Notice that $\langle k | V^0_{s,s+1} | l \rangle$ makes use of the single site basis for the elementary Schrödinger problem. If the basis exhibits any parametric dependence (which is the case for examples of the anharmonic oscillator or quantum pendulum), then all $\langle k | V^0_{s,s+1} | l \rangle$'s do exhibit also, and hence $\epsilon_j = \epsilon_j(\lambda) \Rightarrow w_j = w_j(\lambda)$ arises immediately (provided there is a single parameter, like the coupling constant involved). Compare, e.g., concluding remarks of the previous section.

Let us emphasize that to find Boltzmann weights, one needs to have an explicit operator expression for $V_{s,s+1}$ in terms of the single site raising-lowering operators: $\hat{V} = V(a^*, a)$.

Remark: For the particular case of the ϕ^4_2 theory in $1+1$ dimensions:

$$H = \sum_{j=1}^N \left[-\tfrac{1}{2}\frac{\partial^2}{\partial \phi_j^2} + \frac{\tau}{2}\phi_j^2 + \tfrac{1}{4}\phi_j^4 + \frac{c}{2}(\phi_{j+1} - \phi_j)^2 \right], \tag{4.10}$$

with $\phi \in \mathbb{R}^1$, the spin $\frac{1}{2}$ approximation arises in the single site anharmonic oscillator basis, and due to the simplest possible form of the gradient term

$$\phi_{j+1} - \phi_j = \frac{1}{\sqrt{2}}[(a_{j+1}^* + a_{j+1}) - (a_j^* + a_j)],$$

i.e., $\phi_j = (1/\sqrt{2})(a_j^* + a_j)$ one gets an immediate expression of $P_0 H P_0$ in terms of Fermi variables c^*, c (which in turn are associated with spin $\frac{1}{2}$ variables via the Jordan–Wigner formulas), according to Ref. 18:

$$P_0 H P_0 = \sum_{j=1}^N \{\epsilon c_j^* c_j + \Delta[1 - (c_j^* - c_j)(c_{j+1}^* + c_{j+1})]\}, \tag{4.11}$$

where $\Delta = c|\langle 1|\phi|0\rangle| = c\langle 0|\phi^2|0\rangle$, $\epsilon = E_1 - E_0$ and $|0\rangle$, $|1\rangle$ are vectors of the single-site basis.

By comparing with Ref. 19, one immediately recognizes that for large systems the gradient term $\Delta[1 - (c_j^* - c_j)(c_{j+1}^* + c_{j+1})]$ up to an irrelevant constant Δ, coincides with the one-dimensional Ising term $-J\Sigma_j s_j^x s_{j+1}^x$ provided $\Delta = J/4$.

A dependence of J on coupling constants $\{\tau, c\}$ of the ϕ_2^4 theory is here manifest, as entering J via the $\langle 0|\phi^2|0\rangle = |\langle 1|\phi|0\rangle|$ factor in the expression for Δ.

5. QUANTUM SINE-GORDON SYSTEM AS THE MASSIVE THIRRING MODEL: COMMENTS ON COLEMAN'S EQUIVALENCE

(i) Recall that by virtue of Sec. 4(ii), for the particular case of the quantum sine-Gordon chain, the (m,β) paramatrization enters the Heisenberg model coupling constants **J** by evaluating at β fixed, the explicit values of $V^0_{s,s+1}$ eigenvalues $\{\epsilon_i\}_{i=1,2,3,4}$ The single site problem is quantum pendulum here with m being the coupling constant, and hence indeed $J = J(m,\beta)$.

By rescaling the variables H,ϕ of (1.2) according to

$$\phi = \beta_C \phi', \quad H = \beta_C^2 H', \tag{5.1}$$

we arrive at the Coleman's form,[1] of the sine-Gordon energy density

$$H' = \left\{\left(\frac{\partial\phi'}{\partial t}\right)^2 - c\left(\frac{\partial\phi'}{\partial x}\right)^2\right\}\frac{m^2}{\beta_C^2}(1 - \cos\beta_C\phi'), \tag{5.2}$$

where $c = -1$ in the Coleman's case, and β_C is the Coleman's coupling constant β of Ref. 1.

The single-site quantum pendulum problem arising in connection with the rescaled pendulary chain, is related to the previous one by

$$E'_k = E_k/\beta_C^2, \quad \forall k = 0,1,\cdots, \tag{5.3}$$

and just the E'_0, E'_1 eigenvalues are used in the spin $\frac{1}{2}$ approximation procedure of previous sections.

(ii) It has been proved by Luther,[2] that in case of a very weak anisotropy of the x-y-z model

$$J_x \cong J_y \Rightarrow \mathrm{sn}(2\zeta)\cdot\mathrm{dn}(2\zeta) \cong 1, \tag{5.4}$$

one should make an identification

$$J_z = \beta_C^2/8\pi, \tag{5.5}$$

with the restriction that $\beta_C^2 \in [0,8\pi]$ as within this interval only a continuum limit can be taken for the lattice problem, and shown to lead to the quantum sine-Gordon system of the form (5.2). A massive Thirring model appears here as a mediating step both on the lattice and continuous levels.

By assuming $J_z = J_z(m,\beta)$ we have a unique connection between the (m,β) sine-Gordon chain and the **(J)** Heisenberg problem, compare, e.g., Sec. 3(iv). But then, we must have

$$\beta_C^2 = \beta_C^2(m,\beta) \tag{5.6}$$

and hence in the Coleman's framework varying β_C means: (a) varying β at m fixed, (b) varying m at β fixed, (c) varying both m and β simultaneously.

In fact, the case (a) only was considered in Ref. 1, as then one can "forget" about $\beta = 1/kT$.

(iii) Take $\beta_C^2(m,\beta) \in \mathbb{R}^+$ and assume that at m fixed, $\beta_C^2(m,\beta)$ grows monotonically when $T \to 0$ (i.e., $\beta \to \infty$), and decreases in the opposite extreme, of $T \to \infty$ (i.e., $\beta \to 0$). In terms of the Coulomb gas of charges $\pm q$,[20] at thermal equilibrium, $\beta_C = (4\pi\beta)^{1/2}q, q = q(m)$ and there is a natural critical temperature β_0 corresponding to the Coleman's bound $\beta_C^2 = 8\pi$. Namely for $\beta_C^2 \in [0,8\pi)$ the Coulomb gas lives in its plasma phase, while for $\beta_C^2 \in [8\pi,\infty)$ a dipole gas occurs, and then the system must be kept on the lattice as ultraviolet cutoff cannot be consistently removed from the theory.

(iv) The very same Coulomb gas picture arises while varying m at β fixed. In that case, there is a critical value of $m = m_0$, i.e., $q_0 = q(m_0)$ at which a transition from plasma to dipoles occurs. A correct variability interval for the plasma phase is here $1/m \in [0,1/m_0)$, while $1/m \in [1/m_0,\infty)$ for the dipole phase. In this connection let us recall an old Lenard's result,[21] see, e.g., also Refs. 22 and 23 that in the zero-space dimension quantum pendulum stands for an equivalent of the Coulomb gas problem where the dipole and plasma cases (no phase transition here!) appear at the opposite extremes of m: $m = 0$ and $m = \infty$, respectively.

(v) It suggest the way to understand the Coleman's equivalence in terms of the single-site (i.e., quantum pendulum) data. Let us notice that for large m the quantum pendulum spectral problem admits an equivalent description in terms of the anharmonic oscillator, spectral problem which we choose in the form of[24]:

$$H_\nu = \tfrac{1}{2}[p^2 + \nu(q^2 - (2\nu)^{-1})^2], \quad \nu \to 0. \tag{5.7}$$

In the (weak) limit $\nu \to 0$, the corresponding spectral problem is doubly degenerate and gives

$$\lim_{\nu\to 0} E_\nu^{2j} = \lim_{\nu\to 0} E_\nu^{2j+1} = \sqrt{2}(j + \tfrac{1}{2}) = \epsilon_j. \tag{5.8}$$

For quantum pendulum, the $m \to \infty$ limit gives

$$E_{2n}^{ce} \to E_{2n+1}^{ce} \to (2n + \tfrac{1}{2})\epsilon = (\epsilon/\sqrt{2})\epsilon_{2n},$$
$$E_{2n+1}^{se} \to E_{2n+2}^{se} \to \{(2n+1) + \tfrac{1}{2}\}\epsilon = (\epsilon/\sqrt{2})\epsilon_{2n+1}, \tag{5.9}$$

where ϵ is a fixed positive factor, insensitive to the $m \to \infty$ limit.[25,9]

Whenever to rescale H according to Ref. 24, with an appropriate redefinition of the original constants of Ref. 24:

$$H_\nu = H'_\nu[\beta'(g-1)^{1/2}]^{-1} \cong H'_\nu\beta_C^2,$$
$$g = m/\beta'^3, \quad \nu = g(g-1)^{-3/2}, \quad \beta' = \beta'(\beta), \tag{5.10}$$

then, the $\nu \to 0$ limit is just the same as $m \to \infty$ limit at fixed β. Obviously the $m \to \infty$ limit does make sense for β_C^2 itself, as

$$\beta_C^2 = \beta_C^2(\beta,m) = \frac{\beta'^{1/2}}{(m - \beta'^3)^{1/2}} \tag{5.11}$$

and it does exist for the rescaled Hamiltonian $H'_\nu = H_\nu/\beta_C^2$. Namely, if $\nu \to 0$, then

$$M_\nu = (E_\nu^1)' - (E_\nu^0)' \leqslant A \exp(-Bg^{1/2}), \tag{5.12}$$

where A, B are fixed, g-independent constants. Then $\beta_C^2 = 8\pi$ would imply

$$m = m_0(\beta) = \beta'^3 + \beta'/(8\pi)^2, \tag{5.13}$$

i.e., to each value of β, provided β is small enough [to insure that $m_0(\beta)$ is sufficiently large], we find its own corresponding $m_0(\beta)$.

The small β-large $m_0(\beta)$ limitation means that we need a fixed upper bound M_ν^0:

$$A \exp(-Bg^{1/2}) \leqslant M_\nu^0 \ll 1, \tag{5.14}$$

which induces an appropriate balance of β and $m_0(\beta)$ values.

Let then, at $\beta_C^2 = 8\pi$, $A \exp(-Bg^{1/2}) = M_\nu^0 \ll 1$. It fixes the range of variability for β's at m fixed (or m's at β fixed), within which the bound M_ν^0 is not passed by the energy interval value M_ν.

(vi) By recalling the properties of quantum pendulum when varying m at β fixed, we find that with the lowering of m a separation between lowest two levels of quantum pendulum $\Delta = E^1 - E^0$ increases, while a separation $\Delta' = E^2 - E^1$ between next lowest decreases. Hence a bound M_ν^0 can be consistently interpreted as this value of Δ, beginning from which a two-level approximation of the pendulary chain fails, and a three-level one (at least) should be taken into account.

Remark 1: In terms of the Coulomb gas in zero-space dimension, the dipole gas appears when a spin 0 approximation is reliable (lowest eigenvalue contribution to the partition function is of interest). The plasma appears when a spin $\frac{1}{2}$ approximation becomes reliable (double degeneracy of the lowest eigenvalue occurs).

While working in 1 +1 dimensions, the spin $\frac{1}{2}$ approximation is the lowest one, exhibiting a nontrivial gradient structure, and the nontrivial dipole one, is just the spin 1 (or 2,3,4,···) approximation of the chain as then next eigenvalue contributions to the partition function are taken into account.

Remark 2: In connection with the form of gradient terms in the spin 1 approximation of the linear chain, let us notice that SU(3) is the largest symmetry group for the three-level system [while SU(2) was for a two-level one], and hence the most general form of the nearest neighbor gradient term should be that of the current–current interaction type, with the number eight of coupling constants involved. The structure of the gradient term may vary depending on the explicit choice of the model. For example, a three-level approximation if applied to a system of coupled two-dimensional oscillators in the plane rotor approximation,[26] involves gradient terms of the form $\cos(\phi_{j+1} - \phi_j)$. Their image in the spin 1 approximation is simply $J(S_{j+1}^x S_j^x + S_{j+1}^y S_j^y)$, i.e., the spin 1 x-y model coupling

Remark 3: Coming back to the spin $\frac{1}{2}$ approximation framework, let us notice that the minimal form of gradient terms of lattice Bose systems, which in the two-level approximation, give rise to the Heisenberg chain, has been derived in Ref. 27.

Some other aspects of the two-level approximation trick in connection with the Bose contents of spinor fields can be found in Refs. 28 and 29. The whole program of my investigations on the spin $\frac{1}{2}$ approximation concept was initiated in Ref. 30.

ACKNOWLEDGMENT

I would like to express my gratitude to Professor A. Luther for an invitation to Nordita, where this work has been done. Warm thanks to Alan for his numerous suggestions.

[1] S. Coleman, Phys. Rev. D **11**, 2088 (1975).
[2] A. Luther, Phys. Rev. B **14**, 2153 (1976).
[3] M. Lüscher, Nucl. Phys. B **117**, 475 (1976).
[4] R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. D **11**, 3424 (1975).
[5] P. J. Caudrey, J. C. Eilbeck, and J. D. Gibbon, Nuovo Cimento B **25**, 478 (1975).
[6] P. Garbaczewski and Z. Popowicz, Bull. Pol. Acad. Sci. **26**, 361 (1978).
[7] P. Garbaczewski, Phys. Lett. A **71**, 9 (1979).
[8] G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Theory* (Wiley, New York, 1972).
[9] P. Garbaczewski and G. Vitiello, Nuovo Cimento A **44**, 108 (1978).
[10] P. Garbaczewski, "On quantum solitons and their classical relatives," University of Alberta preprint, Edmonton, 1979.
[11] R. Baxter, Ann. Phys. (N.Y.) **70**, 193 (1972).
[12] R. Baxter, Ann. Phys. (N.Y.) **76**, 1 (1973).
[13] L. P. Kadanoff and F. J. Wegner, Phys. Rev. B **4**, 3989 (1971).
[14] P. Garbaczewski, "A concept of spin $\frac{1}{2}$ approximation in the quantum theory of lattice Bose systems," J. Math. Phys. **21**, 2670 (1980).
[15] H. B. Sutherland, J. Math. Phys. **11**, 3183 (1970).
[16] F. W. Wu, Phys. Rev. B **4**, 2312 (1975).
[17] A. Luther and I. Peschel, Phys. Rev. B **12**, 3908 (1975).
[18] D. J. Scalapino and B. Stoeckly, Phys. Rev. D **14**, 3376 (1976).
[19] P. Pfeuty, Ann. Phys. (N.Y.) **57**, 79 (1970).
[20] S. Samuel, Phys. Rev. D **18**, 1916 (1978).
[21] A. Lenard, J. Math. Phys. **2**, 682 (1961).
[22] J. Frölich, Commun. Math. Phys. **47**, 269 (1976).
[23] J. Zittartz, Z. Phys. B **32**, 113 (1978).
[24] D. Isaacson, Commun. Math. Phys. **53**, 257 (1977).
[25] N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Oxford U.P., Oxford, England, 1951).
[26] B. Stoeckly and D. J. Scalapino, Phys. Rev. B **11**, 205 (1975).
[27] P. Garbaczewski, Phys. Rep. C **36**, No. 2 (1978).
[28] P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[29] P. Garbaczewski, Phys. Lett. A **73**, 280 (1979).
[30] P. Garbaczewski, Acta Phys. Austriaca Suppl. **18**, 434 (1977).

Perturbation methods for the (almost) zero modes and the Green's function in an instanton gas background

Jarmo Hietarinta [a)]
Research Institute for Theoretical Physics, University of Helsinki, Helsinki, Finland

Dennis Nemeschansky [b)]
Department of Technical Physics, Helsinki University of Technology, Espoo, Finland

(Received 18 September 1980; accepted for publication 31 October 1980)

We formulate an approximation method which can be used to compute the almost zero modes and the generalized Green function in the background of an approximate instanton gas solution. The main problem of choosing the starting point is solved by requiring that the zeroth-order operator H_0 annihilates the original isolated single-instanton zero modes. The zeroth-order approximation to H_0^{-1}, on the other hand, is constructed from the free Green function with local bumps provided by the one-instanton Green function.

PACS numbers: 11.10.Np

I. INTRODUCTION

Considerable progress has been made in constructing Green functions and zero modes (ZM) in the background of exact multiinstanton solutions.[1,2] In dilute gas calculations one also uses background configurations that are "approximate solutions."[3] These approximate solutions are typically linear superpositions of instantons and anti-instantons of various sizes and orientations located far from each other. Since the configuration is only an approximate solution the lowest energy modes are not zero modes but almost-zero modes (AZM), whose eigenvalues approach zero as the separations of the instantons and anti-instantons increase. In this paper we present approximation methods for constructing Green functions and the almost-zero modes and their eigenvalues for the case when the background field configuration is an approximate solution.

The difficulty in constructing a perturbative expansion for the present problem is that it is not easy to choose a known zeroth-order starting point. The parameter that determines the convergence is the distance between the instantons: In the limit of infinite separation the local instanton determines the behavior of the Green function and the zero mode completely in its neighborhood. However, this starting point with infinite separations is not very convenient in practice and we will follow a different route.

In Sec. II we will choose our zeroth-order operator H_0 so that the single instanton zero modes $\{\varphi_i^{(0)}\}$ will stay as zero modes of H_0. This means that the original operator H will differ from H_0 only in the subspace spanned by the zero modes located around the various instantons and anti-instantons. For this H_0 we then construct the modified Green function $\bar{G}_0$ defined by

$$H_0(x)\bar{G}_0(x,y) = \delta(x-y) - P_0(x,y), \tag{1.1}$$

where P_0 is the projection operator to the space spanned by the zero modes. The zeroth-order approximation to $\bar{G}_0$ is constructed from the free Green function G_F and the single instanton Green functions $\bar{G}_k$

$$\bar{G}_0 = (1-P_0)\Big(G_F + \sum_k (\bar{G}_k - G_F)\Big)(1-P_0) + \cdots. \tag{1.2}$$

The AZM $\psi_i^{(0)}$ of H, and their eigenvalues $\lambda_i^{(0)}$ are then computed in Sec. III starting from $\varphi_i^{(0)}$ and using $\bar{G}_0$. Finally in Sec. IV we find the generalized inverse $\bar{G}$ of H,

$$H(x)\bar{G}(x,y) = \delta(x-y) - \sum_i \psi_i^{(0)}(x)\psi_i^{(0)}(y), \tag{1.3}$$

using the method of Sec. II and $\psi_i^{(0)}$ and $\lambda_i^{(0)}$ from Sec. III.

The expansions that we obtain converge rapidly if the gas is dilute. The speed of convergence depends for example on how fast the tails of $\varphi_k^{(0)}(x)$ and $\bar{G}_k(x,y) - G_F(x,y)$ approach zero outside the k th instanton. For the kink gas the relevant parameter that describes the convergence vanishes exponentially as the average distance increases, while for Yang–Mills instantons it vanishes like an inverse power.

II. THE GREEN'S FUNCTION WITH OLD ZERO MODES

Let us assume that the operator, whose inverse and AZM we want to calculate, can be written as

$$H = D + V(\phi), \tag{2.1}$$

where D contains derivative parts and is independent of the background field configuration ϕ and where $V(\phi(\infty)) = 0$. For example for the kink system defined by

$$S[\phi] = \int_{-\infty}^{\infty} dt\, [\tfrac{1}{2}(\partial_t\phi)^2 + \tilde{V}(\phi)], \tag{2.2a}$$

we get

$$D = -\partial_t^2 + m^2, \tag{2.2b}$$

$$V(\phi) = \tilde{V}''(\phi) - m^2, \tag{2.2c}$$

$$m^2 = \tilde{V}''(\phi(\infty)). \tag{2.2d}$$

For the fermionic operator in quantum chromodynamics we

[a)] Present address: Department of Physical Sciences, University of Turku, Turku, Finland.
[b)] Present address: Physics Department, Princeton University, Princeton, N. J. 08540.

 022-2488/81/030581-04$1.00

have[2]

$$H = -\gamma^{\mu}(\mathbb{1}\partial_{\mu} - ig\tfrac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_{\mu}), \tag{2.3}$$

where the division (2.1) is obvious. (Here we have suppressed Dirac and group indices.) In the latter example the potential is linear in the background field $\mathbf{A}_{\mu}$, which will bring some simplifications.

For a single instanton or anti-instanton background the zero modes $\varphi^{(0)}_{k,\alpha}, \forall\alpha$ of $H(\phi^k)$ are assumed known,

$$[D + V(\phi^k)]|\varphi^{(0)}_{k,\alpha}\rangle = 0, \tag{2.4}$$

where K labels the instanton. It is possible to have several ZM, labeled by α, for each instanton.

Let us next take a background field configuration $\bar{\phi}$ that is a superposition of single instanton solutions. (In practice we will only need the property that near the l th instanton $\bar{\phi}\simeq\phi^l$.) We will now construct H_0 so that it will annihilate all of the single instanton ZM's $\varphi^{(0)}_{k,\alpha}\ \forall k,\alpha$, but is otherwise unchanged. H_0 will differ from $H(\bar{\phi})$ only in the potential. Let us write

$$V(\bar{\phi}) = V_0(\bar{\phi}) + W(\bar{\phi}), \tag{2.5a}$$

$$H_0 = D + V_0(\bar{\phi}), \tag{2.5b}$$

then V_0 should be such that

$$H_0|\varphi^{(0)}_{k,\alpha}\rangle = 0, \quad \forall k,\alpha. \tag{2.6}$$

Let us denote by P^0_k and P^0 the projection operators to the subspaces spanned by $\{|\varphi^{(0)}_k\rangle\forall\alpha\}$ and $\{|\varphi^{(0)}_{k,\alpha}\rangle\forall k,\alpha\}$, respectively. For P^0_k we have trivially

$$P^0_k = \sum_{\alpha} |\varphi^{(0)}_{k,\alpha}\rangle\langle\varphi^{(0)}_{k,\alpha}|, \tag{2.7a}$$

but we cannot write $P_0 = \Sigma_k P^0_k$, because the states $|\varphi^{(0)}_{k,\alpha}\rangle$ are not necessarily orthogonal for different k. Rather let us define

$$\tilde{P}^0_k = \sum_{\alpha}\sum_{l,\beta} |\varphi^{(0)}_{k,\alpha}\rangle\langle\varphi^{(0)}_{k,\alpha}|\varphi^{(0)}_{l,\beta}\rangle^{-1}\langle\varphi^{(0)}_{l,\beta}|. \tag{2.7b}$$

Then we have

$$P_0 = \sum_k \tilde{P}^0_k. \tag{2.7c}$$

The $\tilde{P}^0_k$'s have the useful property that they separate the zero modes of different instantons when operating to the right, i.e.

$$\tilde{P}^0_k|\varphi^{(0)}_{l,\beta}\rangle = \delta_{kl}|\varphi^{(0)}_{l,\beta}\rangle. \tag{2.7d}$$

The potential V_0 may now be written as

$$V_0 = (1 - P_0)V(1 - P_0) + \sum_k [(1 - P_0)\tilde{V}_k\tilde{P}^0_k + \tilde{P}^{0\dagger}_k\tilde{V}_k(1 - P_0)] + \sum_k\sum_l \tilde{P}^{0\dagger}_k\tilde{V}_{kl}\tilde{P}^0_l, \tag{2.8}$$

where the V's should be determined so that (2.6) holds. We find

$$\tilde{V}_k = \tilde{V}_{lk} = V(\phi^k). \tag{2.9}$$

The remainder W involves terms like $\Sigma_k\tilde{P}^{0\dagger}_k[V(\bar{\phi}) - V(\phi^k)]$. This term vanishes as the instanton separations increase, because $\tilde{P}^0_k$'s and V's vanish far away from instantons, and near a particular l th instanton $\bar{\phi}\sim\phi^l$, and therefore only tail effects remain.

We will also assume that the free Green function and single instanton Green function are known (See Ref. 1 for Y–M instantons and Ref. 4 for kinks). They are respectively solutions of

$$D(x)G_F(x, y) = \delta(x - y), \tag{2.10a}$$

and

$$[D(x) + V(\phi^k(x))]\bar{G}_k(x, y) = \delta(x - y) - P^0_k(x, y). \tag{2.10b}$$

Let $\bar{G}_0$ be the generalized inverse of H_0, i.e., the solution of

$$H_0(x)\bar{G}_0(x, y) = \delta(x - y) - P_0(x, y). \tag{2.11}$$

To solve $\bar{G}_0$ from this we write

$$\bar{G}_0 = \tilde{G}_0 + T_0, \tag{2.12a}$$

$$\tilde{G}_0 = (1 - P_0)\Big(G_F + \sum_k(\bar{G}_k - G_F)\Big)(1 - P_0), \tag{2.12b}$$

where T_0 also annihilates P_0. (The 0 subscripts refer to the ZM-space). The choice of $\tilde{G}_0$ comes from the observation that the instantons only bring local corrections to the Green function. Near the l th instanton $\bar{G}_k - G_F\simeq 0, \forall k\neq l$ and therefore $\tilde{G}_0\simeq\bar{G}_l$. Cross terms where x and y in $\tilde{G}_0(x, y)$ are near different instantons are small because the instantons are far apart. Substituting (2.12) to (2.11) gives

$$H_0T_0 = J_0, \tag{2.13}$$

where

$$J_0 = -H_0\tilde{G}_0 + 1 - P_0 = (1 - P_0)\Big\{\sum_k V(\phi^k)\bar{G}_k - V_0\Big[G_F + \sum_k(\bar{G}_k - G_F)\Big]\Big\} \times(1 - P_0). \tag{2.14}$$

From (2.13) and (2.12) we get an integral equation for $\bar{G}_0$:

$$\bar{G}_0 = \tilde{G}_0 + \bar{G}_0J_0, \tag{2.15}$$

so that

$$\bar{G}_0 = \tilde{G}_0 + \sum_{n=1}^{\infty}\tilde{G}_0J_0^n. \tag{2.16}$$

For the convergence this expression we need $\|J_0\| < 1$. But as the instanton separation increase we infact have $\|J_0\|\to 0$, as can be seen from (2.14): At a point far away from all instantons we have $V\simeq 0$, and near the l th instanton both terms in the curly brackets approach $V(\phi^l)\bar{G}_l$ and cancel.

In this way we have constructed the inverse of our zeroth order (= "unperturbed") operator H_0. The essential feature was to keep at this stage the ZM's intact, and modify H_0 only in the subspace $\{\varphi^{(0)}\}$. Another approach could be to find an $\tilde{H}_0$ so that $\tilde{G}_0$ of (2.12b) is an exact (generalized) inverse of $\tilde{H}_0$. This would lead to a more complicated W in (2.5a) but avoid a series expansion (2.16) at this stage.

III. THE ALMOST-ZERO MODES

To find the almost-zero modes we divide H as in Sec. II:

$$H = H_0 + W, \tag{3.1}$$

$$W = \sum_k \{(1 - P_0)[V(\bar{\phi}) - V(\phi^k)] \times \tilde{P}_k^0 + \tilde{P}_k^{0t}[V(\bar{\phi}) - V(\phi^k)](1 - P_0)\} + \sum_{k,l} \tilde{P}_k^{0t}[V(\bar{\phi}) - V(\phi^k)]\tilde{P}_e^0. \quad (3.2)$$

The degenerate eigenstates of H_0 are the single instanton ZM's $\{|\varphi_{k,\alpha}^{(0)}\rangle, \forall k,\alpha\}$. As the separations between instantons increase, $W \to 0$ and the eigenstates $|\varphi_{k,\alpha}^{(0)}\rangle$ become exact.

According to usual results of degenerate perturbation theory[5] the first order results are found by diagonalizing the matrix

$$\langle \varphi_{k,\alpha}^{(0)}|W|\varphi_{l,\beta}^{(0)}\rangle = \langle \varphi_{k,\alpha}^{(0)}|H|\varphi_{l,\beta}^{(0)}\rangle. \quad (3.3)$$

The diagonal elements are the new eigenvalues. Since the instantons are in arbitrary locations the degeneracy is usually broken completely already in first-order. Let us denote the diagonalizing eigenvectors by $\hat{\psi}_i^0$. (If the original degeneracy was two-fold, $\hat{\psi}_1^0 \propto \varphi_1^{(0)} + \varphi_2^{(0)}$, $\hat{\psi}_2^0 \propto \varphi_1^{(0)} - \varphi_2^{(0)}$.)

For the calculation of higher order terms we write the AZM as

$$\psi_i^{(0)} = \hat{\psi}_i^0 + \sum_{j \neq i} \alpha_{ij}\hat{\psi}_j^0 + \chi_i, \quad (3.4)$$

where $P_0\chi_i = 0$. To first order $\alpha_{ij} = 0$ and $\chi_i = 0$. Substituting this in the eigenvalue equation,

$$H\psi_i = E_i\psi_i, \quad (3.5)$$

gives

$$H_0\chi_i = E_i\psi_i^{(0)} - W\psi_i^{(0)}. \quad (3.6)$$

Since the inverse of H_0 only exits in the space orthogonal to the old ZM's we must project from the left with P_0, which gives the condition

$$\langle \hat{\psi}_k^0|E_i - W|\psi_i^{(0)}\rangle = 0, \quad \forall k,i. \quad (3.7)$$

For $k = i$ we get

$$E_i = \epsilon_i^1 + \langle \hat{\psi}_i^0|W|\chi_i\rangle, \quad (3.8)$$

and for $k \neq i$

$$\alpha_{ik} = \frac{\langle \hat{\psi}_k^0|W|\chi_i\rangle}{E_i - \epsilon_k^1}, \quad (3.9)$$

where

$$\epsilon_j^1 = \langle \hat{\psi}_j^0|W|\hat{\psi}_j^0\rangle. \quad (3.10)$$

This leaves

$$(H_0 - E_i)\chi_i = \Delta_i,$$

$$\Delta_i = -(1 - P_0)W\left(\hat{\psi}_i^0 + \sum_{j \neq i}\alpha_{ij}\hat{\psi}_j^0\right), \quad (3.11)$$

where we have also used the fact that $(1 - P_0)W(1 - P_0) = 0$. From (3.11) we can iterate for χ_i and get

$$\chi_i = \sum_{n=0}^{\infty} (E_i\bar{G}_0)^n\bar{G}_0\Delta_i, \quad (3.12)$$

Equations (3.8)–(3.10), (3.12) determine the eigenfunctions and eigenvalues order by order. The first order results were given above. To the next order

$$\chi_i^{(1)} = -\tilde{G}_0W|\hat{\psi}_i^{(0)}\rangle, \quad (3.13a)$$

$$E_i^{(2)} = \epsilon_i^1 - \langle \hat{\psi}_i^0|W\tilde{G}_0W|\tilde{\psi}_i^0\rangle, \quad (3.13b)$$

$$\alpha_{ik}^{(1)} = -\langle \hat{\psi}_k^0|W\tilde{G}_0W|\hat{\psi}_i^0\rangle/(\epsilon_i^1 - \epsilon_k^1). \quad (3.13c)$$

To this order we may take $\tilde{G}_0$ instead of $\bar{G}_0$, since the difference is one order higher.

The convergence of the expansion (3.12) depends on the eigenvalues of the almost-zero modes. But since they go to zero as the instantons get separated, the expansion will always converge for large enough separation.

IV. GREEN FUNCTION FOR THE FULL OPERATOR

In Sec. II we showed how the generalized Green function can be computed for the operator H_0 with zero modes $\{\varphi_i^{(0)}\}$. For the full H we should now solve

$$H(x)\bar{G}(x,y) = \delta(x - y) - \sum_i \psi_i^{(0)}(x)\psi_i^{(0)}(y). \quad (4.1)$$

Here $\{\psi_i^{(0)}\}$ are the normalized almost-zero modes of H:

$$H(x)\psi_i^{(0)}(x) = \lambda_i\psi_i^{(0)}(x), \quad (4.2)$$

where $\lambda_i \to 0$ as the instanton separations increase. Although the ψ's are not exact zero modes of H it is still necessary to project that subspace out, because we want a well-behaved inverse. Note that the norm of $\bar{G}$ is for most purposes $\simeq 1/\min_i\{|\lambda_i|\}$, and blows up for dilute gas unless the AZM are projected out.

Let us denote by P the projection operator to the space spanned by the AZM's. Following the method of Sec. III we may calculate the AZM's and that way obtain P. One can also use direct expansion methods for P, e.g.,[5]

$$P = P_0 + P_0W\bar{G}_0 + \bar{G}_0WP_0 + \cdots. \quad (4.3)$$

Anyway, for the purposes of this section we may assume that P is known. To calculate $\bar{G}$ we may again start with a similar zeroth-order approximation as in Sec. II, namely

$$\bar{G} = \tilde{G} + T, \quad (4.4a)$$

$$\tilde{G} = (1 - P)\left(G_F + \sum_k (\bar{G}_k - G_F)\right)(1 - P). \quad (4.4b)$$

This differs from (2.12) only in that the projection operators are different, e.g., T now annihilates $\psi_i^{(0)}$'s rather than $\varphi_{k,\alpha}^{(0)}$'s. Substituting (4.4) to (4.1) gives

$$HT = J, \quad (4.5a)$$

where

$$J = (1 - P)\left\{\sum_k V(\phi^k)\bar{G}_k - V\left[G_F + \sum_k(\bar{G}_k - G_F)\right]\right\} \times (1 - P) + (1 - P)\sum_k P_k^0(1 - P). \quad (4.5b)$$

The result is again very similar to (2.14) except for the term $(1 - P)\Sigma_k P_k^0(1 - P)$. This term does not vanish, for although the spaces spanned by the original zero modes and the almost-zero modes have the same dimension they are not identical.

The terms in J are known, small, and annihilate P. The Eqs. (4.3) and (4.4) allow then for a iterative determination of $\bar{G}$ from

$$\bar{G} = \tilde{G} + \bar{G}J, \quad (4.6)$$

$$\bar{G} = \tilde{G} + \sum_{n=1}^{\infty} \tilde{G} J^n. \tag{4.7}$$

The result has the same form as in Sec. II, except that now all operators have different subspace projected out. The convergence properties are the same as before.

V. DISCUSSION

Calculations with dilute instanton gas are made subtle in many places due to the existence of (almost) zero modes. In this paper we have studied the problem of computing the generalized inverse of an operator H in an instanton gas background, assuming that the problem has been solved for a single instanton background. There are two problems in computing H^{-1}: 1) The inverse is well-behaved only in a certain (at the moment unknown) subspace perpendicular to the AZM, 2) There is no simple way to isolate the "unperturbed" H_0.

In this paper we have chosen H_0 so that it annihilates the individual zero modes $\{\varphi^{(0)}_{k,\alpha} \forall k, \alpha\}$ of all instantons, i.e., $H_0 P_0 = 0$, but is otherwise identical to H. In this way we know the subspace where H_0^{-1} exists. The zeroth-order approximation to H_0^{-1} is obtained from the free Green function with bumps at the location of each instanton supplied by the local exact Green function:

$$H_0^{-1} = (1 - P_0)\Big[G_F + \sum_k (\bar{G}_k - G_F)\Big](1 - P_0) + T_0. \tag{2.12'}$$

T_0 is the small perturbation that was computed in Sec. II.

After H_0^{-1} is known it is straightforward to compute the AZM of H, this was done in Sec. III. Once the AZM's are computed we also know the subspace where H^{-1} is well defined, and then H^{-1} can be computed according to Sec. IV starting with

$$H^{-1} = (1 - P)\Big[G_F + \sum_k (\bar{G}_k - G_F)\Big](1 - P) + T. \tag{4.4'}$$

Following the above method one can now compute the needed almost-zero modes and Green functions in a dilute gas background. The expansion method presented here and its speed of convergence could be tested with exact multi-instanton solutions, where the AZM's and Green functions are known in closed form.[1,2]

[1] L. S. Brown, R. D. Carlitz, D. B. Creamer, and C. Lee, Phys. Lett. B **70**, 180 (1977); *ibid.* B **71**, 103 (1977); Phys. Rev. D **17**, 1583 (1978); E. Corrigan, P. Goddard, D. B. Fairlie, and S. Templeton, Nucl. Phys. B **140**, 31 (1978); N. H. Christ, E. Weinberg, and N. K. Stanton, Phys. Rev. D **18**, 2013 (1978); E. Corrigan, P. Goddard, and S. Templeton, Nucl. Phys. B **151**, 93 (1979).

[2] B. Grossman, Phys. Lett. A **61**, 86 (1977); H. Osborn, Nucl. Phys. B **140**, 45 (1978).

[3] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, Phys. Lett. B **63**, 334 (1976); S. Coleman, *The uses of instantons*, 1977 Erice lectures.

[4] J. Hietarinta, Nucl. Phys. B **164**, 343 (1980).

[5] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1965), Vol. II, Chap. XVI.

Linearly superposable and finite-action classical Yang–Mills fields

C. H. Oh and Rosy Teh
School of Physics, University of Science of Malaysia, Penang, Malaysia

(Received 11 December 1979; accepted for publication 15 February 1980)

We show that in the Minkowski space, a self-dual gauge field can be linearly superposed with either another self-dual gauge field or a non-self-dual gauge field to give a new solution of the Yang–Mills equation. From these new solutions, Euclidean solutions are constructed. Some of these new solutions are valid in any dimensions of space–time.

PACS numbers: 11.10.Np, 11.10.Qr

I. INTRODUCTION

Recently there has been much interest in searching for new classical solutions of the Yang–Mills (YM) equations[1] and their properties.[2] The technique involved is usually to employ an ansatz for the gauge field A_μ^a and the most well-known one is the Corrigan–Fairlie–'t Hooft–Wilczek ansatz[3]

$$A_\mu(x) = g\frac{\sigma^a}{2i}A_\mu^a(x) = i\sigma_{\mu\nu}\,\partial^\nu \ln\phi(x), \tag{1}$$

where g is the gauge field coupling constant, $\sigma^{ij} = \frac{1}{2}\epsilon^{ijk}\sigma_k$, and in the Minkowski space, $\sigma^{j0} = i\sigma^j/2$ whereas in the Euclidean space $\sigma^{j4} = \sigma^j/2$. Here σ^j are the Pauli matrices. With this ansatz, the SU(2) YM equations reduce to a nonlinear equation for the function $\phi(x)$,

$$\Box\phi(x) + \lambda\phi^3(x) = 0, \tag{2}$$

and the self-anti-duality condition gives rise to

$$\Box\phi(x) = 0. \tag{3}$$

Thus the single instanton solution[4] corresponds to

$$\phi(x) = 4/(2 + \lambda r^2), \quad r^2 = x_\mu x^\mu, \tag{4}$$

while a single meron[5] is given by

$$\phi(x) = (\lambda)^{-1/2}\frac{1}{r}, \tag{5}$$

where λ is a constant. In Ref. 6 the meron solution is recovered from the real part of a complex self-dual solution in the Minkowski space and in Ref. 7, specific examples of self-dual complex solutions in the Minkowski space are constructed which can be summed up to yield the meron solution. These results suggest that the superposition principle may be possible for classical nonabelian YM fields. In fact, in Ref. 8 it has been pointed out that functions of the form $\phi(u)$, with $u = px + e, p^2 = p_\mu p^\mu = 0$, and e being a constant, will lead to self-dual gauge fields A_μ which can be linearly superposed to produce other self-dual solutions. Of course trivial linear superposition can always be achieved for the YM fields by writing an ansatz of the form $A_\mu^a(x) = g^a B_\mu(x)$. With g^a being a constant vector, one automatically gets rid of the nonlinear term in the field strength $F_{\mu\nu}^a$ and reduces the nonabelian equation to the linear field equation. Hence, solutions with such an ansatz are linearly superposable. This type of solutions has been discussed in Refs. 9 and 10.

In this paper we exhibit self-dual gauge field configurations in the Minkowski space which can be linearly superposed with self-dual or non-self-dual fields to give nontrivial solutions to the YM equations. Some of these solutions so obtained by superposition can be analytically continued to the Euclidean space. In this way, we obtain a family of solutions which can be made regular everywhere in the Euclidean space and as a result, their topological charge, action, as well as energy–momentum densities vanish. These solutions, however, possesses nonzero complex electric and magnetic field strengths. Examples of complex zero-action solutions in the Euclidean space have been considered before.[9] In Ref. 9, however, an ansatz of the form

$$A_\mu = a_\mu(x)\,M, \tag{6}$$

with M being a constant nilpotent matrix, is employed which essentially linearizes the YM equations, and, consequently, the resulting solutions are not truly non-Abelian. As is well known, in the Euclidean space the self-dual solutions, e.g. the instanton,[4] possess finite actions[11] and it has been conjectured that all finite action solutions of the YM equations must be either self-dual or self-anti-dual.[2] Some of the solutions obtained here in Sec. III seem to confirm the conjecture.

In Sec. II, linearly superimposable YM fields in the Minkowski space are displayed and some of their properties are given. In Sec. III we show that by making the constant in the solutions in the Euclidean space–time to be complex, nonsingular finite action solutions are derived. Some solutions obtained in Secs. II and III are valid in any dimensional space–time and the corresponding static ones are hence written down in Sec. IV. These static solutions are complex and singular along a straight line. They can be regarded as the generalization of the cylindrical solutions given in Ref. 13. The complex static gauge field solution can be understood by converting it into an exact solution for the real static SU(2) gauge field coupled minimally to the triplet Higgs field when the self-interaction potential of the later vanishes.[14] We end in Sec. V with brief remarks.

II. LINEARLY SUPERPOSABLE SOLUTIONS

The self-duality condition (3) admits any function of a single variable $u = px + e$, with $p^2 = p_\mu p^\mu = 0$ and e being a constant, as a solution in the Minkowski space–time. The metric $g_{\mu\nu}$ is $(-\,+\,+\,+)$. Consider now two self-dual solutions $\phi(x) = \phi_1(u)$, $u = px + e$, and $\phi(x) = \phi_2(v)$, $v = qx + f (q^2 = 0, f = \text{constant})$. If we demand the product $\phi(x) = \phi_1(u)\,\phi_2(v)$ be a solution of the YM equation (2), then the function form of ϕ_1 and ϕ_2 is the square root of the argument, i.e., $\phi_1 = u^{-1/2}$ and $\phi_2 = v^{-1/2}$ with $pq \neq 0$ in order

 022-2488/81/030585-05$1.00

for $\phi = \phi_1\phi_2$ be a solution of (2). Obviously, if $pq = 0$, $\phi = \phi_1\phi_2$ is again a self-dual solution. Thus, starting from two self-dual configurations, $A_\mu = i\sigma_{\mu\nu}\,\partial^\nu \ln\phi_1$ and $A_\mu = i\sigma_{\mu\nu}\,\partial^\nu \ln\phi_2$, the linear sum $A_\mu = (i\sigma_{\mu\nu}\,\partial^\nu \ln\phi_1 + i\sigma_{\mu\nu}\,\partial^\nu \ln\phi_2)$ is again a solution which may be self-dual or non-self-dual depending on whether pq vanishes or not. Note that

$$\phi = \phi_1^2 = \frac{1}{u} = \frac{1}{px+e}, \quad p^2 \neq 0, \tag{7}$$

also leads to a self-dual solution if $2p^2 = -\lambda$. This solution and its elliptic generalizations have been discussed in Ref. 8.

The non-self-dual solution

$$\phi = \phi_1\phi_2 = (uv)^{-1/2}, \quad \lambda = -\tfrac{1}{2}pq, \quad p^2 = q^2 = 0, \tag{8}$$

can be generalized. One finds that

$$\phi = (uv)^{-1/2}\,E(w), \tag{9}$$

with $w = A\ln(u/v)$, A = arbitrary constant, satisfies the YM equation (2) provided the function $E(w)$ is a solution of the following differential equation,

$$E'^2 + aE^2 + \tfrac{1}{2}bE^4 = c, \tag{10}$$

where $E' = dE/dw$, c is a constant, and for the solution (9), $a = -1/(4A^2)$, $b = -\lambda/(2A^2pq)$. This means that $E(w)$ can be any one of the 12 Jacobi elliptic functions.[15] Note that for solution (9) the constraint $\lambda = -pq/2$ is not required. The solution (8) has singularities on planes defined by $u = 0$ and $v = 0$. The generalized solution (9) will not introduce additional singularities if we choose for $E(w)$ the following elliptic functions,

$$E_1 = \mathrm{dn}(w,k), \quad E_2 = \mathrm{nd}(w,k), \quad 0 \leqslant k \leqslant 1. \tag{11}$$

The corresponding gauge fields can be written as

$$A_\mu = i\frac{\sigma_{\mu\nu}}{E_{1,2}}\left\{\left(\mp k^2 A\,\frac{\mathrm{sn}(w)\,\mathrm{cn}(w)}{\mathrm{dn}(w)} - \frac{1}{2}\right)\frac{p^\nu}{u} - \left[\mp k^2 A\,\frac{\mathrm{sn}(w)\,\mathrm{cn}(w)}{\mathrm{dn}(w)} + \frac{1}{2}\right]\frac{q^\nu}{v}\right\}. \tag{12}$$

The generalized solution (9) with $E(w)$ given by (11) reduces to the original solution (8) when the parameter k of the elliptic functions vanishes. For $k = 1$, $\mathrm{dn}(w) = 1/\cosh w$, and $\mathrm{nd}(w) = \cosh w$. All the above solutions are valid in any dimensions of space–time.

For the ansatz (1), the energy-momentum tensor is given by the expression

$$\theta_{\mu\nu}(x) = (\lambda/g^2)[4\partial_\mu\phi\partial_\nu\phi - 2\phi\partial_\mu\partial_\nu\phi - g_{\mu\nu}(\lambda\phi^4/2 + \partial_\alpha\phi\partial^\alpha\phi)]. \tag{13}$$

One finds for the generalized solution (9), the energy density is

$$\theta_{00} = \frac{\lambda}{g^2}\frac{2cA^2}{uv}\left[2\left(\frac{p_0}{u} - \frac{q_0}{v}\right)^2 - \frac{pq}{uv}\right], \tag{14}$$

where p_0 is the time component of p_μ, and the momentum density is

$$\theta_{0i} = \frac{\lambda}{g^2}\frac{4cA^2}{uv}\left(\frac{p_i}{u} - \frac{q_i}{v}\right)\left(\frac{p_0}{u} - \frac{q_0}{v}\right). \tag{15}$$

The energy and momentum densities for solution (8) are also respectively given by expressions (14) and (15) except one must now put $8cA^2 = -1$. It is interesting to note that although the solution ϕ_1 and ϕ_2 lead to self-dual gauge fields with zero energy–momentum tensor,[16] the sum of these gauge fields possesses nonvanishing real energy and momentum densities. Furthermore, we have computed the field strengths F_{0i} and F_{ij} associated with solutions (8) and (9). They are complex and singular and their spatial directions are fixed. The singularities of the above solutions need the presence of external sources to sustain themselves and cannot be gauge-transformed away.

The self-dual solution ϕ_1 can also be combined with the non-self-dual meron solution to give another non-self-dual solution. Translating the position of the meron from the origin to the position $x_\mu = -a_\mu$ gives the solution,

$$\phi = [(x+a)^2]^{-1/2}. \tag{16}$$

The product of expression (16) with the solution ϕ_1, i.e.,

$$\phi = [u(x+a)^2]^{-1/2}, \quad p^2 = 0, \quad \lambda = -pa, \tag{17}$$

satisfies the nonlinear equation (2). Note that for $a_\mu = 0$, the resulting solution (17) is self-dual. Thus adding the self-dual gauge field $A_\mu = i\sigma_{\mu\nu}\,\partial^\nu \ln\phi_1$ to the single meron solution results in a non-self-dual configuration or a self-dual configuration depending on whether a_μ is nonvanishing or not. Unlike the previous case, the solution (17) is valid only in four-dimensional space–time. One can again generalize solution (17) to

$$\phi = [u(x+a)^2]^{-1/2}\,E(w), \quad p^2 = 0, \tag{18}$$

with $w = A\ln[(x+a)^2/u]$ and the $E(w)$ is the Jacobi elliptic function with $a = -1/(4A^2)$, $b = -\lambda/(4pqA^2)$. As before, the generalized solution (18) will not give additional singularities only if $E(w)$ is given by the choice (11).

The energy and momentum densities associated with generalized solution (18) are evaluated by Eq. (13) and one obtains

$$\theta_{00} = -\frac{\lambda}{g^2}\frac{8A^2c}{u(x+a)^2}\left(\frac{(pa/2) + 2p_0(x+a)_0}{u(x+a)^2} - \frac{p_0^2}{2u^2} - \frac{2(x+a)_0^2}{(x+a)^4}\right), \tag{19a}$$

and

$$\theta_{0i} = -\frac{\lambda}{g^2}\frac{8A^2c}{u(x+a)^2}\left(\frac{p_0(x+a)_i + p_i(x+a)_0}{u(x+a)^2} - \frac{p_0p_i}{2u^2} - \frac{2(x+a)_0(x+a)_i}{(x+a)^4}\right). \tag{19b}$$

For the solution (17), the expressions for the energy and momentum densities are respectively the same as Eqs. (19a,b), except that $8A^2c = -1$. Although solution (17) and its generalization (18) give rise to complex gauge field and the field strengths, their energy and momentum densities are real.

III. EUCLIDEAN SOLUTIONS

The solutions discussed in Sec. II, other than the solution (7), are valid only in the Minkowski space-time as they

become trivial in the Euclidean space-time due to the requirement $p^2 = q^2 = 0$.

The self-dual solutions (7), parametrized by the four-vector p_μ and the constant e, are valid in the Euclidean space-time and they have some interesting properties. One can also directly obtain them from Eq. (2) by regarding ϕ as a function of one variable u. They are singular on a plane defined by

$$p_\mu x_\mu = -e.$$

If we now make the constant e complex and restrict p_μ and the coordinates x_μ to assume real values only, then solutions (7) are regular everywhere. Their energy and momentum densities vanish. The topological charge density for the ansatz (1) can be written as[17]

$$D(\phi) = \pm \frac{1}{2g^2} \Box\Box \ln\phi , \tag{20}$$

whilst the Lagrangian density is given by

$$L = \pm D(\phi) - 3\lambda^2\phi^4/2g^2 . \tag{21}$$

For solutions (7) with e being a complex constant, the topological charge density is found to be

$$D(\phi) = \pm \frac{3}{2g^2}\left(\frac{p^2}{u^2}\right)^2 .$$

This leads to zero topological charge and zero action. Solutions (7) are, however, not gauge-transform of a pure gauge since the electric and magnetic fields are nonvanishing and given by

$$E_i = F_{4i} = -i\sigma_{4i}\, p^2/u^2 ,$$

and

$$B_i = \tfrac{1}{2}\epsilon_{ijk}\, F^{jk} = -\tfrac{1}{2}i\,\epsilon_{ijk}\sigma^{jk}\, p^2/u^2 .$$

As the constant e is complex, these field strengths are necessarily complex. Thus with e being a complex constant, we have demonstrated a complex self-dual[18] gauge field in the real Euclidean space-time which is regular and possesses zero energy, zero Pontryagin index, and zero action. This type of solutions have been discussed in Ref. 9. Solutions (7) are, however, different from the voidon of this reference since the nonlinear term of $F_{\mu\nu}$, which is absent in Ref. 9, is retained for solutions (7), and the $F_{\mu\nu}$ here is not proportional to a constant nilpotent matrix.

Although the solutions in Sec. II are trivial in the Euclidean space–time, one can, by manipulations, derive from the non-self-dual solution (8) the following expression,

$$\phi(x) = (\bar{u}^2 + \bar{p}^2 x_4^2)^{-1/2},$$
$$(\bar{u} = (p_i x^i + e), \quad \bar{p}^2 = p_i p^i , \tag{22}$$

which is a solution for Eq. (2) provided $\lambda = -\bar{p}^2$. This solution holds in any dimensions of the Euclidean space–time and by virtue of the ansatz (1), it leads to real gauge fields A^a_μ,

$$A_\mu(x) = (-i)\frac{\sigma_{\mu j}\, p^j \bar{u} + \sigma_{\mu 4}\, p_4^2\, x_4}{\bar{u}^2 + \bar{p}^2 x_4^2} . \tag{23}$$

As long as the Euclidean time $x_4 \neq 0$, the gauge field is regular everywhere. However, at $x_4 = 0$, A_μ is singular on the plane defined by

$$x_i\, p^i = -e . \tag{24}$$

It appears that this singularity of A_μ is caused by an external source which is switched on only at the Euclidean time $x_4 = 0$. One can shift the singularity occurring at $x_4 = 0$ to $x_4 = f$(const) by writing

$$\phi(x) = [\bar{u}^2 + \bar{p}^2(x_4 - f)^2]^{-1/2}, \quad \lambda = -\bar{p}^2, \tag{25}$$

which is also a solution of Eq. (2). Furthermore, in two-dimensional space–time with coordinates (x,x_4), expression (22) takes the form

$$\phi = [(x+e)^2 + x_4^2]^{-1/2}, \quad \lambda = -1, \tag{26}$$

which is regular everywhere except at a point.

As before, solution (22) can be generalized by incorporating the Jacobi elliptic functions. Thus we find that

$$\phi(x) = (\bar{u}^2 + \bar{p}^2 x_4^2)^{-1/2} E(w),$$
$$w = A \tan^{-1}\left(\frac{\bar{p}x_4}{\bar{u}}\right), \tag{27}$$

satisfies the non-self-dual equation (2) if we require $E(w)$ to be one of the elliptic functions with $a = 1/A^2$ and $b = \lambda/(A^2\bar{p}^2)$. Choosing $E(w)$ be given by Eq. (11), the generalized solution (27) introduces no extra singularities.

For the generalized solution (27), one can evaluate the Euclidean energy and momentum densities and one gets

$$\theta_{44} = \frac{\lambda}{g^2}\frac{cA^2(3\bar{p}^2\bar{u}^2 - \bar{p}^2 x_4^2)\,\bar{p}^2}{(\bar{p}^2 x_4^2 + \bar{u}^2)^3} , \tag{28}$$

and

$$\theta_{4i} = \frac{\lambda}{g^2}\frac{4cA^2\bar{p}^2 x_4 \bar{u} p_i}{(\bar{p}^2 x_4^2 + \bar{u}^2)^3} , \tag{29}$$

where the constant c is defined in Eq. (10). The energy and momentum densities for solution (22) are also respectively given by Eqs. (28) and (29) with $2cA^2 = 1$. For the solution (22), $D(\phi)$ vanishes everywhere except possibly at the singularity plane defined by Eq. (24). The action diverges due to the singularity, but the total energy, as evaluated from

$$\xi = \int d^3x\,\theta_{44} \quad (x_4 = 0), \tag{30}$$

is zero.

IV. STATIC SOLUTIONS

As pointed out earlier, the solutions in Sec. III and some solutions in Sec. II are valid in any dimensions of space–time. It is then straightforward to write down the static solutions. For the ansatz (1) with the function $\phi(x_i)$ being time-independent, the static YM field equation is

$$\nabla^2\phi + \lambda\phi^3 = 0. \tag{31}$$

Expressions (8) and (9) are respectively the solutions of Eq. (31), provided we replace u by $\bar{u} = p_i x^i + e$, v by $\bar{v} = q_i x^i + f$, w by $\bar{w} = A\ln(\bar{u}/\bar{v})$, and λ by $\bar{\lambda} = -\frac{1}{2}p_i q^i$. However, these solutions are meaningless because of the constraints $p_i p^i = q_i q^i = 0$. For expression (20), the corresponding solutions of Eq. (31) which lead to complex static gauge fields are either

$$\phi = \frac{1}{\tilde{u}}, \tag{32}$$

or

$$\phi = (\tilde{u}^2 + \tilde{p}^2 x_3^2)^{-1/2}, \quad \lambda = -\tilde{p}^2, \tag{33}$$

where $\tilde{u} = p_A x^A + e$, $\tilde{p}^2 = p_A p^A$, and the index A runs from 1 to 2. Solution (32) and its elliptic generalization have been obtained in Ref. 8 and can be regarded as leading to a static gauge field due to the presence of a plane of sources. The elliptic generalization of (33) can be deduced from expression (25) and one has

$$\phi = (\tilde{u}^2 + \tilde{p}^2 x_3^2)^{-1/2} E(\tilde{w}),$$
$$\tilde{w} = A \tan^{-1}(\tilde{p} x_3 / \tilde{u}). \tag{34}$$

Solution (33) is regular everywhere except along a line on the x^1–x^2 plane defined by

$$\tilde{u} = 0, \quad x^3 = 0. \tag{35}$$

This line of singularity indicates the presence of external sources as can be shown by using the method of Ref. 19. The elliptic generalization provides no additional singularities if $E(\tilde{w})$ is given by the choice (11).

The energy density associated with solution (33) is negative,

$$\theta_{00} = -\tfrac{1}{2}(\lambda/g)^2(\tilde{u}^2 + \tilde{p}^2 x_3^2)^{-2}, \tag{36}$$

and it increases as x_i increases. The momentum density θ_{0i} for the static gauge field is zero as $\phi(x)$ here is time-independent. For the electric and magnetic fields we find, after some calculations,

$$E_A = i\tilde{p}^2 \phi^2 [\phi^2 x_3 p_A (x_3 p_B \sigma_{0B} - \tilde{u}\sigma_{03}) - \sigma_A],$$
$$E_3 = -i\tilde{p}^2 \phi^4 x_3 (\tilde{u} p_B \sigma_{0B} + \tilde{p}^3 x_3 \sigma_{03}), \tag{37}$$

and

$$B_A = \frac{i}{2} \phi^4 \tilde{p}^2 x_3 p_A (\sigma_3 \tilde{u} - x_3 \sigma_B p_B),$$
$$B_3 = \frac{i}{2} \phi^4 \tilde{p}^2 \tilde{u} (x_3 p_B \sigma_B - \tilde{u}\sigma_3). \tag{38}$$

In Ref. 13, we have found that

$$\phi = (-\lambda x_A x^A)^{-1/2}, \tag{39}$$

in a static solution of the YM equation. By setting $p_1 = 1$, $p_2 = e = 0$, solution (33) reduces to expression (39). Hence we can regard (33) as a generalization of (39).

The complex static gauge field as derived from expression (33) can be converted into the real SU(2) static gauge field coupled minimally to the triplet Higgs field Φ^a, when the self-interaction potential of the later vanishes. This is done by setting[14]

$$gA_i^a = \epsilon_{iab} \partial^b \ln\phi,$$
$$gA_0^a = \sinh\gamma \partial^a \ln\phi,$$
$$g\Phi^a = \cosh\gamma \partial^a \ln\phi,$$

where γ is a real constant. Following 't Hooft,[20] one can then define the electromagnetic field $\mathcal{F}_{\mu\nu}$. As for the solution (39),[13] the magnetic field, $(1/2)\epsilon_{ijk}\mathcal{F}^{jk}$, vanishes but not the electric field $\mathcal{F}_{0i}$ as

$$\mathcal{F}_{0i} = \frac{\tilde{p}\sinh\gamma}{g[\tilde{u}^2 + \tilde{p}^2 x_3^2]^{3/2}} (\tilde{u} p_B \delta_{Bi} + \tilde{p}^2 x_3 \delta_{3i}).$$

Thus the singularity of (33) arises from external sources which are "electric" in nature. In contrast with solutions (33) and (39), static solutions known to us in the literature lead to a nonvanishing magnetic field.

V. COMMENTS

We now make some brief remarks on the solutions obtained in this paper.

(I) The linear superposition principle is, of course, in general, not valid for the YM theory. However in Sec. II we have constructed examples for which a self-dual gauge field can be linearly superposed with either a self-dual or a non-self-dual gauge field to give new solutions. As these solutions are obtained without linearizing the YM equation, they are truly nonabelian. What characterizes these solutions to be linearly superposable is not clear to us. We merely observe that self-duality[18] may play a role here, since linear superposition of two non-self-dual gauge fields seems unlikely to yield a solution, except in the case of the two-meron solution.

(II) If the arbitrary constant e is allowed to be complex, solution (7) leads to a Euclidean self-dual gauge field with zero action, zero Pontryagin index, and zero total energy. Complex solutions with zero action may play a role in the semiclassical approximation of the Feynman functional integral.

(III) Solutions (7), (8), and (20) are valid in any dimensions of space–time. In Ref. 5, the single instanton and a pair of merons in the six-dimensional coordinates, ξ^μ, are respectively written

$$\phi = (2/\lambda)^{1/2}(\xi_\mu p^\mu)^{-1}, \quad p_\mu p^\mu = 1, \tag{40}$$

and

$$\phi = (2\lambda)^{-1/2} \left[\frac{p_\mu q^\mu}{(p_\mu \xi^\mu)(q_\mu \xi^\mu)}\right]^{-1/2}, \quad p^2 = q^2 = 0. \tag{41}$$

Apart from a factor "i", these expressions are respectively the same as solutions (7) and (8). Does this mean that one can interpret solutions (7) and (8) respectively as the six-dimensional instanton and a pair of merons? One can answer this question only if one can calculate their corresponding Pontryagin index in the six-dimensional space–time. This we have not done.

We note that in a recent preprint,[21] Kovacs and Lo also discuss the solution (7).

[1] C.N. Yang and R.L. Mills, Phys. Rev. **96**, 191 (1954).
[2] For a review, see R. Jackiw, C. Nohl, and C. Rebbi, "Classical and Semiclassical solutions of Yang-Mills theory," MIT preprint, CTP 675 (1977).
[3] E. Corrigan and D.B. Fairlie, Phys. Lett. B **67**, 69 (1977); G. 't Hooft (unpublished); F. Wilczek, in *Quark Confinement and Field Theory*, edited by D.R. Stump and D.H. Weingarten (Wiley, New York, 1977).
[4] A. Belavin, A. Polyakov, A. Schwartz, and Y. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[5] V. de Alfaro, S. Fubini, and G. Furlen, Phys. Lett. B **65**, 163 (1976).
[6] C. Rebbi, Phys. Rev. D **17**, 483 (1978).

[7]J. Cervero, Lett. Nuovo Cimento **23**, 19 (1978).
[8]C.H. Oh, and Rosy Teh, Phys. Lett. B **87**, 83 (1979).
[9]L. Dolan, Phys. Lett. B **76**, 105 (1978).
[10]F. Melia and S.Y. Lo, Phys. Lett. B **77**, 71 (1978).
[11]Any function $\phi(x) = \Sigma_i \tanh P_i(\eta^i x)$, with $\eta^i_\mu \eta^{i\mu} = 0$ and P_i being even polynomials in $\eta^i x$ such that $P_i \gtrless 0$ for all i and x_μ, leads to a self-dual gauge field configuration with finite action. See Ref. 12 here.
[12]W. Bernreuther, Phys. Rev. D **16**, 3609 (1977).
[13]C.H. Oh and R. Teh, Phys. Rev. D. **21**, 531 (1980).
[14]C.H. Oh, Phys. Rev. D **18**, 4817 (1978).
[15]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[16]C.H. Oh and S. Murugesu, Prog. Theor. Phys. **59**, 671 (1978).
[17]R. Jackiw and C. Rebbi, Phys. Rev. D **14**, 517 (1976); A. Actor, "Elliptic function methods in ϕ^4 theory and Yang-Mills theory," preprint, Universität Dortmund, 1978.
[18]We use "non-self-dual" to mean "non-self-dual" as well as "non-self-anti-dual"; similary for "self-dual".
[19]C.H. Oh, Phys. Rev. D **19**, 1248 (1979).
[20]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[21]E. Kovacs and S.Y. Lo, "Two general classes of self-dual Minkowski propagating solutions in Yang-Mills gauge theory," Preprint UM-P-78/54.

An axiomatic framework for supersymmetries

Demetris P. K. Ghikas [a)]
Department of Mathematics, Bedford College, University of London, Regent's Park, London NWI 4NS, England

(Received 12 July 1978; accepted for publication 15 February 1980)

A framework is proposed in which supersymmetries can be defined in the language of axiomatic quantum field theory. This framework contains a nuclear *-algebra into which the Borchers' algebra of a general supermultiplet is embedded through a map which, in some sense, generalizes the concept of a superfield. The algebra of supersymmetry is represented on the constructed nuclear *-algebra as an algebra of graded derivations. Since a graded derivation cannot be integrated to a usual automorphism group (only to a formal group), it is assumed that conditions at the Lie algebraic level are strong enough to produce supersymmetric behavior. Thus it is conjectured that if a functional of the nuclear *-algebra is annihilated by the algebra of supersymmetry, and if this functional is related to a state of the Borchers' algebra through the embedding map, this state through the GNS construction gives rise to a supersymmetric Wightman theory. This supersymmetric condition produces an infinite number of correlations among the n-point functions. The 2-point Wightman functions for a general supermultiplet are completely analyzed and it is found that their behavior is similar to the perturbative results. Finally it is proved that the free fields satisfy these supersymmetric conditions.

PACS numbers: 11.30.Pb, 11.10.Cd, 02.10. – v

I. INTRODUCTION

Supersymmetry has been introduced as a graded extension of the Lie-algebraic symmetries.[1] Accepting supersymmetry as a symmetry one is expecting some sort of correlations to manifest themselves at some level of the resulting theory. But one is also facing the problem of explicitly deriving these correlations in a consistent framework.[1,2]

In theoretical physics a framework is provided by the Lagrangian theories which through the superspace-superfield constructions can lead to very concrete perturbative results.[1] Unfortunately, the concepts introduced in these approaches, like Grassmann parameters, formal groups, and indefinite metric spaces,[2-4] are very obscure and very far removed from the usual understanding of symmetry. Formulated in mathematically precise terms this understanding takes the form of a duality: "mathematical regularities of the framework $\longleftrightarrow$ physical correlations." The purpose of the present work is to study the possibility for such a duality of symmetries for the case of supersymmetries. The introduced new mathematical structures can in turn be used for more general dualities concerning new ideas of symmetry.

The safest, conceptually, domain of mathematical formulations for the duality of symmetries, is the axiomatic quantum field theory. The two formulations of AQFT the C*-algebraic[5] and the Wightman theory[6,7] provide us with important concepts which can point towards the proper generalizations needed for supersymmetry. Since Wightman theory is more closely related to the conventional quantum field theory, we state the problem as follows:

Can we specify conditions which must be satisfied by a Wightman theory so that, if such a theory exists, it is supersymmetric, namely, that the dynamical correlations (Wightman functions) are characteristic of a supersymmetric theory (supersymmetric in the conventional sense)?

We note here that there exists a work[8] where supersymmetry is treated in a Wightman framework. The central issue in this reference is whether a graded-Lie algebra can be thought of as an extension of the usual Lie algebras of the known symmetries and also what are the possible extensions. But there exist no investigations about the dynamical implications of such a symmetry, that is studies concerning the existence of supersymmetric theories. If such theories exist the dynamical correlations are consequences of conditions of symmetry.

Such conditions of symmetry are understood in a satisfactory level in the C*-algebraic frameworks. For this reason we are going to use the Borchers' algebra formulation of Wightman theory and generalize to it "most naturally" the results already established in the C*-algebraic frameworks. In the latter and for usual symmetries (no grading) one has the following situation:

The Lie algebra $\mathcal{L}$ acts on the C*-algebra $\mathfrak{A}$ as an algebra of derivations. These derivations may be bounded or unbounded. One can integrate $\mathcal{L}$ into a group $\mathcal{G}$ of automorphisms of $\mathfrak{A}$ (for unbounded derivations certain conditions should be satisfied first).[9] Then an invariant under $\mathcal{G}$ state of $\mathfrak{A}$, through the GNS construction, gives a covariant representation of $\mathfrak{A}$ on a Hilbert space. In the same time $\mathcal{L}$ annihilates this state.

In order to obtain a Wightman theory in a direct way one has to start from the Borchers' algebra $\underline{\mathcal{L}}$, which is not a C* but a nuclear *-algebra.[7] For $\underline{\mathcal{L}}$ one can, possibly,[10] generalize the theory of symmetries arising from usual Lie algebras and it is expected that the only difficulties will come from the more difficult topological structure of $\underline{\mathcal{L}}$, being a nuclear space and not a Banach space. But even if one had a complete theory for derivations on $\underline{\mathcal{L}}$ it is not at all obvious how to consider a graded-Lie algebra (supersymmetry) as an

Present address: Laboratory of Nuclear Technology, Polytechnic Faculty, University of Patras, Panepistimiopolis, Patras, Greece.

 022-2488/81/030590-10$1.00

algebra of derivations acting on $\underline{\mathscr{S}}$. We want to enlarge $\underline{\mathscr{S}}$ in such a way so that this is possible.

We propose the following scheme:

(1) Construct a nuclear *-algebra, call it $\underline{\underline{TAS}}$, on which the algebra $\underline{\mathscr{L}}$ of Supersymmetry acts as a graded-Lie algebra of derivations.

(2) Construct a map $\underline{\mathscr{T}} : \underline{\mathscr{S}} \to \underline{\underline{TAS}}$ from the Borchers' algebra $\underline{\mathscr{S}}$, corresponding to a given supermultiplet, into $\underline{\underline{TAS}}$, as a natural generalization of the superfield concept.

(3) Let $(\underline{\underline{TAS}})'$ and $\underline{\mathscr{S}}'$ be the topological dual spaces of $\underline{\underline{TAS}}$ and $\underline{\mathscr{S}}$. Let $\underline{F} \in (\underline{\underline{TAS}})'$ be such that there exists a Wightman state $\underline{W} \in \underline{\mathscr{S}}'$ so that

$$\underline{W}(\underline{f}) = \underline{F}(\underline{\mathscr{T}}\,\underline{f}), \quad \forall\, \underline{f} \in \underline{\mathscr{S}}. \tag{1}$$

If $\underline{Q}$ is any element of the graded-Lie algebra $\underline{\mathscr{L}}$, let $\underline{F}$ satisfy additionally the condition:

$$\underline{F}(\underline{\underline{Q}}\,\underline{\underline{\Phi}}) = 0, \quad \forall \underline{\underline{\Phi}} \in \underline{\underline{TAS}}. \tag{2}$$

For a usual Lie algebra (2) would be one of the necessary conditions for the integrability of $\underline{\mathscr{L}}$ to a group of automorphisms. Here we cannot integrate $\underline{\mathscr{L}}$, at least for algebraic reasons. Then we have to make the following conjecture: If $\underline{W}$ is a Wightman state arising from an $\underline{F}$ through (1) and this $\underline{F}$ satisfies (2) then $\underline{W}$ gives a Wightman theory (by the GNS construction) with correlations characteristic of supersymmetry.

In the present work we construct the algebraic skeleton of the above scheme, without touching deep topological problems, or questions of existence. Then we use (1) and (2) to derive correlations for the 2-point Wightman functions, where we use in a decisive way the ansatz that the 2-point components of $\underline{F}$ have the structure of the Källen–Lehmann representation (KL). We obtain correlations very similar to those in the perturbative calculations.

For a general n-point function we need some sort of a generalized KL representation, and thus we did not do any calculations for $n \geqslant 3$. But we were able to prove that for any n the free fields satisfy (1) and (2) and thus we conclude that they are supersymmetric in our sense. We consider it as supporting the consistency of our framework.

We will omit nearly all details of the construction in order not to overburden the reader with technicalities which can be found in Ref. 11. We collect the necessary formulas and technical points in the Appendix along with the conventions and identities used in the work. The various propositions are stated without proofs since they are not difficult and can also be found in Ref. 11.

II. THE ALGEBRAIC AND TOPOLOGICAL STRUCTURES OF THE THEORY

A. The Borchers' algebra of a scalar superfield

The framework we are going to construct can be thought of as a prototype for the treatment of any graded-Lie algebra and in the case of supersymmetry any supermultiplet. Nevertheless for reasons of concreteness and simplicity and also for comparison we are going to deal with the scalar supermultiplet of Salam and Strathdee (general superfield). Being interested only in a scalar supermultiplet, we consider as the irreducible set of fields, which will generate our Wightman theory, the set of independent fields appearing as coefficients in the θ-parameter expansion of a classical scalar superfield.

Now, taking as more primitive object the Borchers' algebra that generates a Wightman theory, the former must be an algebra of test functions that incorporates the appropriate representations of SL(2,$\mathbb{C}$). This means that, when the Borchers' algebra is represented on a Hilbert space as an algebra of generators, there is a well-defined prescription to obtain fields with definite transformation properties under SL(2,$\mathbb{C}$).

This is achieved as follows. Start from a 4-dimensional vector space V and consider its basis $\{\theta_\alpha, \alpha = 1,...,4\}$ to be a Majorana representation of SL(2,$\mathbb{C}$). Construct the Grassmann algebra $\Lambda(V)$.[12] Then the representation of SL(2,$\mathbb{C}$) can be uniquely extended to $\Lambda(V)$,[13] and one can define covariants formed out of the Grassmann parameters $\{\theta_\alpha\}$. Let us equip the vector space $[\Lambda(V)]$ with its Euclidean topology and let $\mathscr{S}(\mathbb{R}^4,[\Lambda(V)])$ denote the space of $\mathbb{C}^\infty$ functions $f(x)$ from $\mathbb{R}^4$ to $[\Lambda(V)]$ such that for all pairs of polynomials P,Q in four variables with complex coefficients $P(x)Q(\partial/\partial x)f(x)$ remains in a bounded subset of $[\Lambda(V)]$ as x varies over $\mathbb{R}^4$. We equip $\mathscr{S}(\mathbb{R}^4,[\Lambda(V)])$ with the topology of uniform convergence of the functions $P(x)Q(\partial/\partial x)f(x)$ on the whole space $\mathbb{R}^4$ for all possible P and Q. Every element of $\mathscr{S}(\mathbb{R}^4,[\Lambda(V)])$ is of the form

$$\Phi(x_\mu,\theta_\alpha) = A(x) + \bar{\theta}_\alpha \psi_\alpha(x) + ... + \tfrac{1}{32}(\bar{\theta}\theta)^2 D(x). \tag{3}$$

It is straightforward to define the classical scalar superfield as the elements (3), which satisfy

$$\Phi(x_\mu,\theta_\alpha) = \Phi(x'_\mu,\theta'_\alpha),$$

for appropriate transformations $x \to x'$, $\theta \to \theta'$. Let $(E_A,E_\psi,...,E_D)$ be vector spaces associated with the representations appearing in (3) and let $\{e_\tau^T\}$, $T = A,\psi...,D$, be the corresponding basis. Consider the following definitions:

$$\mathscr{E}_{T_1...T_n} = E_{T_1} \otimes \cdots \otimes E_{T_n},$$

$$\mathscr{E}_n = \sum_{T_1...T_n} \mathscr{E}_{T_1...T_n} \quad \text{(algebraic direct sum)},$$

$$\underline{\mathscr{E}}\,\underline{\mathscr{S}}_n = \mathscr{E}_n \hat{\otimes} \mathscr{S}(\mathbb{R}^4) \cong \mathscr{S}(\mathbb{R}^{4n},\mathscr{E}_n)$$

$$\underline{\mathscr{E}}\,\underline{\mathscr{S}} = \sum_n \oplus\, \underline{\mathscr{E}}\,\underline{\mathscr{S}}_n, \quad \underline{\mathscr{E}}\,\underline{\mathscr{S}}_0 = \mathbb{C} \quad \text{(topological direct sum)}.$$

Equipping each E_{T_i} with its Euclidean topology, $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ becomes a nuclear topological vector space.[14,15,11] The elements of $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ are terminating sequences of functions with values in tensor products of finite dimensional vector spaces.

Let $\underline{\underline{f}} \in \underline{\mathscr{E}}\,\underline{\mathscr{S}}$. We have

$$\underline{\underline{f}} = \{\underline{f}_0 \underline{f}_1,...\underline{f}_n,...\}, \quad \underline{f}_0 \in \mathbb{C}, \tag{4}$$

$$\underline{f}_n = \sum_{\substack{T_1...T_n \\ \tau_1...\tau_n}} e_{\tau_1}^{T_1} \otimes ... \otimes e_{\tau_n}^{T_n} \otimes {}^n f^{\bar{\tau}_1...\bar{\tau}_n}_{\tau_1...\tau_n}, \tag{5}$$

where τ_i is a tensor/spinor index corresponding to the represent. T_i and $\bar{\tau}_i$ means that this index transforms under the represent. $\bar{T}_i$ which is the adjoint of T_i for the spinor and coincides with T_i otherwise.

We want to turn $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ into an algebra. The vector space structure is introduced as in the usual Borchers' Algebra.[11] The product and the star operations are introduced similar-

ly. The precise definitions appear in the Appendix. We have the following *Proposition* 1: The space $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$ equipped with the operations defined above is a nuclear *-algebra.

For the topological dual $(\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}})'$ we can prove easily that its elements are of the form

$$\underline{\underline{W}} = \{\underline{W}_0, \underline{W}_1, ..., \underline{W}_n, ...\} \quad \text{(infinite sequence)},$$

where $\underline{W}_n = \{\underline{W}_n^{T_1...T_n}\}$ are finite sequences and

$$\underline{W}_n^{T_1...T_n} = \sum_{\tau_1...\tau_n} \vec{e}^{*\tau_1}_{T_1} \otimes ... \otimes \vec{e}^{*\tau_n}_{T_n} \otimes {}^nW^{T_1...T_n}_{\tau_1...\tau_n},$$

with $\{\vec{e}^{*\tau_i}_{T_i}\}$ basis of $E^*_{T_i}$ dual to $\{e^{T_i}_{\tau_i}\}$ and ${}^nW^{T_1...T_n}_{\tau_1...\tau_n} \in \underline{\underline{\mathscr{S}}}'(\mathbb{R}^{4n})$. Then from the duality of the bases we have

$$\underline{\underline{W}}(\underline{\underline{f}}) = \sum_n \sum_{\substack{T_1...T_n \\ \tau_1...\tau_n}} {}^nW^{T_1...T_n}_{\tau_1...\tau_n}({}^nf^{\bar{\tau}_1...\bar{\tau}_n}_{\bar{T}_1...\bar{T}_n}), \tag{6}$$

with $\underline{\underline{f}} \in \underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$, $\underline{\underline{W}} \in (\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}})'$.

Given a state $\underline{\underline{W}}$ of $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$ we can construct a representation of $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$ on a Hilbert space $\mathscr{H}_W$ with a straightforward extension of the procedure followed in the Borchers' algebra $\underline{\underline{\mathscr{S}}}$.[11]

It can be seen that the representation $\underline{\underline{\pi}}_W : \underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}} \to \mathscr{L}(\mathscr{H}_W)$ defined by

$$\underline{\underline{\pi}}_W(\underline{\underline{f}})[\underline{\underline{g}}] = [\underline{\underline{f}} \times \underline{\underline{g}}],$$

has the property

$$\underline{\underline{\pi}}_W(\underline{\underline{f}}) = \sum_n \sum_{\substack{T_1...T_n \\ \tau_1...\tau_n}} {}^n\pi^{T_1...T_n}_{\tau_1...\tau_n}(f^{\bar{\tau}_1...\bar{\tau}_n}_{\bar{T}_1...\bar{T}_n}), \tag{7}$$

with ${}^n\pi \in \underline{\underline{\mathscr{S}}}'(\mathbb{R}^{4n}, \mathscr{L}(\mathscr{H}_W))$, the space of tensors of rank n with components operator valued distributions acting on $\mathscr{H}_W$. Also

$$([\underline{\underline{g}}], \underline{\underline{\pi}}_W(\underline{\underline{f}})[\underline{\underline{h}}]) = \underline{\underline{W}}(\underline{\underline{g}}^* \times \underline{\underline{f}} \times \underline{\underline{h}}). \tag{8}$$

Finally we point out that the locality ideal,[7] must be generalized to accommodate the locality concept for fermions. See Ref. 11 for details.

B. An indexing algebra

As it has been stated in the Introduction the scheme we propose contains a big algebra on which supersymmetry can be represented as an algebra of graded derivations. At the same time its elements must contain the necessary test functions to reproduce the component fields of the theory. Thus the elements must be hybrid consisting of test functions and Grassmann elements. But as we are going to see shortly, they must contain an extra bit of information, which is needed in order to have the correct representation of supersymmetry, on one hand, and appropriate generality and consistency on the other. This third bit of information comes from the elements of an extra algebra, the indexing algebra, which with the Grassmann algebra $\Lambda(V)$ and the Borchers' algebra $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$ are fused to form the big algebra $\underline{\underline{\underline{TAS}}}$. A typical element of $\underline{\underline{\underline{TAS}}}$ is going to be of the form

$$M_1 \otimes ... \otimes M_k \otimes \alpha \otimes f^k_n(x_1, ..., x_n), \tag{9}$$

where $M_i \in \Lambda(V)$, $f^k_n \in \underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$, and α is an element of the indexing algebra. We proceed to motivate the introduction and the structure of the latter.

Let the product of n classical superfields with argument$(x_1, ..., x_n)$. Its form is $\Phi(x_1, ..., x_n, \theta_\alpha)$ for some function Φ. Consider an infinitesimal supersymmetric transformation

$$x^\mu_1 \to x^\mu_1 + \epsilon\gamma^\mu\theta, ..., x^\mu_n \to x^\mu_n + \epsilon\gamma^\mu\theta, \quad \theta_\alpha \to \theta_\alpha + \epsilon_\alpha.$$

If we expand $\Phi(x^\mu_1 + \epsilon\gamma^\mu\theta, ..., x^\mu_n + \epsilon\gamma^\mu\theta, \theta_\alpha + \epsilon_\alpha)$ in powers of ϵ we see that always appear the sum $\Sigma_{i=1}^n \partial/\partial x^\mu_i$. It will be seen below, that from translation invariance, the symmetric states of the big algebra $\underline{\underline{\underline{TAS}}}$ will annihilate the terms involving this sum. But we want to have the representation of Q_α as

$$C_{\alpha\beta}\frac{\partial}{\partial\theta_\beta} - i(\gamma^\mu)_{\alpha\beta}\theta_\beta\frac{\partial}{\partial x^\mu},$$

or at least modify it without loosing the part containing the derivations $\partial/\partial x^\mu_i$. We can solve this problem if we define Q_α in such a way that apart from the sum $\Sigma_{i=1}^n \partial/\partial x^\mu_i$ all possible partial sums appear like

$$\sum_{i=r_{k-1}+1}^{r_{k-1}+r_k} \partial/\partial x^\mu_i,$$

where r_k is defined by $r_1 + ... + r_l = n$, $0 \leqslant r_k \leqslant n$, $l \leqslant n$.

But as it turns out we must have these sums associated with the factors M_i in (9) in a specific way. To produce this appearance of the partial sums the action of Q_α must be guided by the element α in (9). α must be associated with a set of integers which are indexing the subset of the set of the arguments $(x_1, ..., x_n)$ with respect to which the derivations must be taken.

But due to the operations of product and conjugation there happens that the various subsets of arguments intermingle in a complicated way. If we want to have possible some theorems on desired properties, we must take into account this fact and equip the elements of the indexing algebra with this faculty as well. Some details of the construction of the indexing algebra of order k, $\underline{\underline{\mathscr{A}}}^k$, can be found in the Appendix. It is a *-algebra (associative).

C. A possible generalization of the Borchers' algebra

In this paragraph we finally introduce the big algebra $\underline{\underline{\underline{TAS}}}$ as a possible generalization of the algebra $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$.

We define the vector space over $\mathbb{C}$:

$$(\underline{\underline{\underline{TAS}}})^k_{n'...n} = \underbrace{[\Lambda(V)] \otimes ... \otimes [\Lambda(V)]}_{k \text{ factors}} \otimes \left[\underline{\underline{\mathscr{A}}}^k_{n'...n}\right] \otimes \underset{n = n' + ... + n}{\mathscr{S}(\mathbb{R}^{4n})}. \tag{10}$$

Both $\Lambda(V)$ and $\left[\underline{\underline{\mathscr{A}}}^k_{n'...n}\right]$ are finite dimensional vector spaces. See the Appendix for the definition of $\underline{\underline{\mathscr{A}}}^k_{n'...n}$. We equip them with their Euclidean topology so that they are Hausdorff and complete. Then we define the complete vector space

$$(\underline{\underline{\underline{\bar{T}\bar{A}\bar{S}}}})^k_{n'...n} = \mathscr{S}(\mathbb{R}^{4n}, [\Lambda(V)] \otimes ... \otimes [\Lambda(V)] \otimes \left[\underline{\underline{\mathscr{A}}}^k_{n'...n}\right]). \tag{11}$$

Let

$$\underline{\underline{\underline{TAS}}} = \sum_{\substack{k=0,1,2,... \\ N=1,2,...}} \oplus (\underline{\underline{\underline{\bar{T}\bar{A}\bar{S}}}})^k_{n'...n}, \quad (\underline{\underline{\underline{\bar{T}\bar{A}\bar{S}}}})^0_0 = \mathbb{C} \tag{12}$$

(topological direct sum), $k \leqslant n^j, j = 1,...,N$. The elements of $\underline{TAS}$ are terminating sequences

$$\underline{\Phi} = \{\{\underline{\Phi}_0, \underline{\Phi}_1, ..., \underline{\Phi}_n, \cdots\}\}, \tag{13}$$

$$\underline{\Phi}_n = \{\underline{\Phi}^1_n, ..., \underline{\Phi}^k_n, ..., \underline{\Phi}^n_n\}, \tag{14}$$

$$\underline{\Phi}^k_n = \sum_{\substack{1 \leqslant N \leqslant n \\ k \leqslant n^j, j = 1,...,N, \ n^1 + ... + n^N = N.}} \underline{\Phi}^k_{n^1...n^N}, \tag{15}$$

and $\underline{\Phi}^k_{n^1...n^N}$ is linear combination of elements of the form

$$M_1 \otimes ... \otimes M_k \otimes \alpha\{n\} \otimes f^{\{n,\tau\}},$$

where M_i and $\alpha\{n\}$ are basis elements of the vector spaces $[\Lambda(V)]$ and $[\mathscr{A}^k_{n^1...n^N}]$ correspondingly. The M_i are covariant monomials in $\Lambda(V)$. The coefficient functions $f^{\{n,\tau\}}$ are in $\mathscr{S}'(\mathbb{R}^{4n})$, and transform under Lorentz transformation as tensors of rank k. In the Appendix some explanation for the notation can be found.

We define below the operations with $\underline{\Phi}, \underline{\Psi} \in \underline{TAS}$, $\lambda \in \mathbb{C}$.

Sum and multiplication by scalar

$$\underline{\Phi} + \lambda\underline{\Psi} = \{\{\underline{\Phi}_0 + \lambda\underline{\Psi}_0, ..., \underline{\Phi}_n + \lambda\underline{\Psi}_n, ...\}\}, \tag{16}$$

$$\underline{\Phi}_n + \lambda\underline{\Psi}_n = \{\underline{\Phi}^1_n + \lambda\underline{\Psi}^1_n, ..., \underline{\Phi}^n_n + \lambda\underline{\Psi}^n_n\}. \tag{17}$$

Product

$$(\underline{\Phi} \times \underline{\Psi})_n = \{(\underline{\Phi} \times \underline{\Psi})^1_n, ..., (\underline{\Phi} \times \underline{\Psi})^n_n\}, \tag{18}$$

$$(\underline{\Phi} \times \underline{\Psi})^k_n = \sum_{p=k}^{n-k} \underline{\Phi}^k_p \times \underline{\Psi}^k_{n-p} \tag{19}$$

$$= \sum_{p=k}^{n-k} \sum_{1 \leqslant N \leqslant p} \sum_{1 \leqslant M \leqslant n-p} \underline{\Phi}^k_{n^1...n^N} \times \underline{\Psi}^k_{m^1...m^M} \tag{20}$$

$k \leqslant n^s, \quad s = 1,...,N, \quad k \leqslant m^t, \quad t = 1,...,M,$

$n^1 + ... + n^N = p, \quad m^1 + ... + m^M = n - p,$

$$\underline{\Phi}^k_{n^1...n^N} \times \underline{\Psi}^k_{m^1...m^M}$$

$$= \sum_{\{\ \}} (-1)^\lambda \times M_1 N_1 \otimes ... \otimes M_k N_k \otimes (\alpha\{n\} \times \alpha'\{m\}) \otimes (f^{\{n,\tau\}} \times g^{\{m,\sigma\}}). \tag{21}$$

The explicit from is very lengthy. MN is the product of M,N in $\Lambda(V)$, $\alpha \times \alpha'$ the product of α,α' in $\mathscr{A}^k$ and $f \times g$ the Borchers' product. The above is a direct generalization of the product in the algebras which are tensor products of two graded algebras.[12] In the Appendix the notation is explained a little.

We define the star operation in $\underline{TAS}$ as follows:

$$\underline{\Phi}^* = \{\{(\underline{\Phi}^*)_0, (\underline{\Phi}^*)_1, ..., (\underline{\Phi}^*)_n, ...\}\},$$

$$(\underline{\Phi}^*)_0 = \bar{\underline{\Phi}}_0 \tag{22}$$

$$(\underline{\Phi}^*)_n = \{(\underline{\Phi}^*)^1_n, ..., (\underline{\Phi}^*)^n_n\}, \quad (\underline{\Phi}^*)^k_n = (\underline{\Phi}^k_n)^*. \tag{23}$$

Let $\underline{\Phi}^k_n$ be of the form

$$\underline{\Phi}^k_n = M_1 \otimes ... \otimes M_k \otimes \alpha \otimes f. \tag{24}$$

We define

$$(\underline{\Phi}^k_n)^* = M^*_k \otimes ... \otimes M^*_1 \otimes \alpha^* \otimes f^*. \tag{25}$$

We have propositions 2 and 3 as follows.

Proposition 2: The space $\underline{TAS}$ equipped with the operations defined above is a *-algebra (associative).

Proposition 3: $\underline{TAS}$ is a nuclear TVS. For the topological dual one can easily see that we have[11]:

$$(\underline{TAS})' = \times_n \sum_{1 \leqslant N \leqslant n} \oplus [(\bar{\underline{TA}}\bar{\underline{S}})^k_{n^1...n^N}]', \tag{26}$$

$$[(\bar{\underline{TA}}\bar{\underline{S}})^k_{n^1...n^N}]' = \mathscr{S}'(\mathbb{R}^{4n}, [\Lambda(V)]^* \otimes ... \otimes [\Lambda(V)]^*, [\mathscr{A}^k_{n^1...n^N}]^*), \tag{27}$$

where $[\Lambda(V)]^*$ and $[\mathscr{A}^k_{n^1...n^N}]^*$ are the algebraic duals of $[\Lambda(V)]$ and $[\mathscr{A}^k_{n^1...n^N}]$ respectively and $\mathscr{S}'(\mathbb{R}^{4n})$ the topological dual of $\mathscr{S}(\mathbb{R}^{4n})$.

The elements of $(\underline{TAS})'$ are infinite sequences,

$$\underline{F}\{\{F_0, \underline{F}_1, ..., \underline{F}_n, ...\}\}, \quad F_0 \in C \tag{28}$$

$$\underline{F}_n = \{\underline{F}^1_n, ..., \underline{F}^k_n, ..., \underline{F}^n_n\}, \tag{29}$$

and we have for $\underline{\Phi} \in \underline{TAS}$

$$\underline{F}(\underline{\Phi}) = \sum_n \underline{F}_n(\underline{\Phi}_n) = \sum_n \sum_{k=1}^n \underline{F}^k_n(\underline{\Phi}^k_n)$$

(a finite sum since $\underline{\Phi}$ is a terminating sequence), where

$$\underline{F}^k_n \in \mathscr{S}'(\mathbb{R}^{4n}, [\Lambda(V)]^* \otimes ... \otimes [\Lambda(V)]^* \otimes [\mathscr{A}^k_{n^1...n^N}]). \tag{30}$$

$$\underline{\Phi}^k_n \in \mathscr{S}(\mathbb{R}^{4n}, [\Lambda(V)] \otimes ... \otimes [\Lambda(V)] \otimes [\mathscr{A}^k_{n^1...n^N}]). \tag{31}$$

It can be seen, using the dual bases in $[\Lambda(V)]^*$, $[\mathscr{A}^k_{n^1...n^N}]^*$ that

$$\underline{F}^k_n(\underline{\Phi}^k_n) = \sum\sum \omega(f)$$

where $\omega \in \mathscr{S}'(\mathbb{R}^{4n})$, and all the tensor indices can be easily added.

Now we come to a crucial point. We want to map $\underline{\mathscr{C}}\,\underline{\mathscr{S}}$ into $\underline{TAS}$ in such a way that the images in $\underline{TAS}$ of the elements of $\underline{\mathscr{C}}\,\underline{\mathscr{S}}$ are the most likely generalizations of the superfield concept.

Let $\mathscr{C}^*_n$ be the dual of $\mathscr{C}_n$ and let

$$\underline{\mathscr{C}}\,\underline{\mathscr{C}}_n = \mathscr{C}^*_n \otimes [\Lambda(V)], \quad \underline{\mathscr{C}}\,\underline{\mathscr{C}} = \Sigma_n \oplus \underline{\mathscr{C}}\,\underline{\mathscr{C}}_n, \tag{32}$$

$\underline{\mathscr{C}}\,\underline{\mathscr{C}}$ is a vector space over $\mathbb{C}$. We turn it into an algebra by defining the product:

$$\underline{\mathscr{C}}\,\underline{\mathscr{C}}_n \times \underline{\mathscr{C}}\,\underline{\mathscr{C}}_m \to \underline{\mathscr{C}}\,\underline{\mathscr{A}}_{n+m},$$

$$(\bar{e}^*_{T_1} \otimes ... \otimes \bar{e}^*_{T_n} \otimes M^{\{T\}}) \times (\bar{e}^*_{S_1} \otimes ... \otimes \bar{e}^*_{S_m} \otimes N^{\{S\}})$$

$$= \bar{e}^*_{T_1} \otimes ... \otimes \bar{e}^*_{S_m} \otimes M^{\{T\}} N^{\{S\}}, \tag{33}$$

where $\{T\}, \{S\}$ are sets of tensor indices and MN is the product of M,N in $\Lambda(V)$.

Let the element $\underline{j} \in \underline{\mathscr{C}}\,\underline{\mathscr{C}}$:

$$\underline{j} = e^*_A \otimes 1 \oplus \bar{e}^{*\alpha}_\psi \otimes \theta_\alpha + ... + \tfrac{1}{32} e^*_D \otimes (\bar{\theta}\theta)^2, \tag{34}$$

and $\underline{j}^{n_i} = \underline{j} \times \underline{j} \times ... \underline{j}$, n_i factors. For every (n,k) positive integers $1 \leqslant k \leqslant n$ and a partition of n, $n_1 + ... + n_k = n$, let the map (see Appendix for the notation):

$$\mathscr{T}^{(n,k)}_{n_1...n_k} : \underline{\mathscr{C}}\,\underline{\mathscr{S}}_n \to (\underline{TAS})^k_n \quad \text{(i.e., } N = 1, n^1 = n), \tag{35}$$

$$\mathscr{T}^{(n,k)}_{n_1...n_k}(\bar{e}^{T_1}_{\tau_1} \otimes ... \otimes \bar{e}^{T_n}_{\tau_n} \otimes f^{\tau_1...\tau_n}_{T_1...T_n})(x_1,...,x_n)$$

$$= \sum_{Q(i_1,...,i_n)} (-1)^{\pi(Q)} \left[\underline{j}^{n_1}(\bar{e}^{T_{i_1}}_{\tau_{i_1}} \otimes ... \otimes \bar{e}^{T_{i_{n_1}}}_{\tau_{i_{n_1}}}) \otimes ... \otimes \left[\underline{j}^{n_k}(e^{T_{i_{n_k+1}}}_{\tau_{i_{n_k+1}}} \otimes ... \otimes \bar{e}^{T_{i_k}}_{\tau_{i_k}})\right]\right.$$

$$\otimes {}^{k}\underline{\alpha}^{(n,n,k)}_{\begin{bmatrix} r_1 \dots r_n \\ i_1 \dots i_n \end{bmatrix}(n_1,n_k)} \otimes f^{\tau_1 \dots \tau_n}_{T_1 \dots T_n}(x_1,\dots,x_n), \tag{36}$$

where $Q(i_1,\dots,i_n)$ means summation over all the permutations of $(1,\dots,n)$, and $\pi(Q)$ = the number of times two spinor representation are interchanged in order to obtain the configuration $\{T_{i_1},\dots,T_{i_n}\}$ from $\{T_1,\dots,T_n\}$,
$r_1 = r_2 = \dots = r_n = 1$, and

$$\underline{j}^{n_l}(\vec{e}^{T_1}_{\tau_1} \otimes \dots \otimes \vec{e}^{T_{n_l}}_{\tau_{n_l}})$$

etc. are defined as follows:
Since $\underline{j}^{n_l}$ is a sum of terms of the form

$$\vec{e}^{*T_{j_1}}_{T_{j_1}} \otimes \dots \otimes \vec{e}^{*T_{j_{n_l}}}_{T_{j_{n_l}}} \otimes M^{T}_{\tau},$$

we define

$$(\vec{e}^{*T_{j_1}}_{T_{j_1}} \otimes \dots \otimes \vec{e}^{*T_{j_{n_l}}}_{T_{j_{n_l}}} \otimes M^{T}_{\tau})(e^{T_1}_{\tau_1} \otimes \dots \otimes e^{T_{n_l}}_{\tau_{n_l}})$$
$$= \delta^{\tau_{j_1}}_{\tau_1}\delta^{T_1}_{T_{j_1}} \dots \delta^{\tau_{j_{n_l}}}_{\tau_{n_l}}\delta^{T_{n_l}}_{T_{j_{n_l}}} M^{T}_{\tau}. \tag{37}$$

Thus $\underline{j}^{n_l}$ are maps $\mathscr{E}_{n_j} \to [\Lambda(V)]$.

Let P^n_k be the number of partitions of n into k parts $n_1 + \dots + n_k = n$ such that $0 \leqslant n_s \leqslant n$, $s = 1,\dots,k$. We extent the map (36) to $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ with the definitions:

$$\underline{\mathscr{T}}^{k}_{n} = \frac{1}{p^n_k} \sum_{\substack{n_1,\dots,n_k \\ n_1+\dots+n_k=n}} \underline{\mathscr{T}}^{(n,k)}_{n_1,\dots,n_k}, \tag{38}$$

$$\underline{\mathscr{T}}_{n} = \left\{ \frac{1}{n}\underline{\mathscr{T}}^{1}_{n},\dots,\frac{1}{n}\underline{\mathscr{T}}^{n}_{n} \right\}, \tag{39}$$

and the infinite sequence

$$\underline{\underline{\mathscr{T}}} = \{1,\underline{\mathscr{T}}_1,\dots,\underline{\mathscr{T}}_n,\dots\}. \tag{40}$$

This defines a linear map of graded modules:

$$\underline{\underline{\mathscr{T}}}:\underline{\mathscr{E}}\,\underline{\mathscr{S}} \to \underline{TAS},$$
$$\underline{\underline{\mathscr{T}}}(\underline{f}) = \{\{\dots,\underline{\mathscr{T}}_n(\underline{f}_n),\dots\}\}, \tag{41}$$
$$\underline{\mathscr{T}}_n(\underline{f}_n) = \left\{\dots,\frac{1}{n}\underline{\mathscr{T}}^{k}_{n}(\underline{f}_n),\dots\right\}. \tag{42}$$

III. REPRESENTATION OF SUPERSYMMETRY AS AN ALGEBRA OF GRADED DERIVATIONS ON TAS. PHYSICAL INTERPRETATION

A. Representation of supersymmetry

We need some definitions.

Definition 1: Let $\underline{\Phi}^{k}_{n}$ be an element of the form (24). We call the number

$$\nu(\underline{\Phi}^{k}_{n}) = \nu(M_1) + \dots + \nu(M_k), \tag{43}$$

the degree of $\underline{\Phi}^{k}_{n}$ in $\underline{TAS}$, where $\nu(M_i) = 0,1$ is the degree of M_i in $\Lambda(V)$.

According to this definition it can be easily seen that $\underline{TAS}$ is a semigraded algebra. Let us recall two known definitions from the theory of semigraded algebras.[12]

Definition 2: Let $E = E_+ + E_-$ be a semigraded algebra. The map

$$J:E \to E, \quad x \to x', \quad x = x_+ + x_-, \quad x' = x_+ - x_-,$$

is called the *main involution* of E and it is an involutive automorphism of E with the property

$$J(x \times y) = J(x) \times J(y). \tag{44}$$

Definition 3: Let ϕ be an homomorphism of a semigraded algebra E into E'. A ϕ-derivation of E into E' is a linear mapping $D:E \to E'$, homogeneous of some given degree ν, satisfying for every $x,y \in E$

$$D(xy) = D(x)\phi(y) + \phi(J^{\nu}(x))D(y), \tag{45}$$

where

$$J^{\nu} = J, \quad \nu \text{ odd},$$
$$= 1, \quad \text{even}.$$

If $E = E'$ then D is called a *graded derivation*.

Now we want to construct a representation of the spinorial charges Q_α in such a way that they act as derivations on $\underline{TAS}$. It will turn out that they will act as graded derivations, while the rest of the generators of supersymmetry will act as usual derivations.

Every element of $\mathscr{A}^k$ is a linear combination of k-tuples of symbols. Let us write compactly for a k-tuple

$$\underline{\alpha}\{n\} = \{{}^{1}\underline{\alpha}(n,k),\dots,{}^{s}\underline{\alpha}(n,k),\dots,{}^{k}\underline{\alpha}(n,k)\}$$

Let ${}^{s}\underline{\alpha}(n,k)$ represent the symbol $\{n,1\}|m_1,\dots,m_p\}$. For every s let us define the map ind_s from the set of basis elements to $(Z^+)^p$:

$$\mathrm{ind}_s(\underline{\alpha}\{n\}) = \mathrm{ind}_s({}^{s}\underline{\alpha}(n,k)) \equiv \{m_1,\dots,m_p\}. \tag{46}$$

Let $\mathscr{L}(\mathscr{S}(\mathbb{R}^{4n}))$ denote the vector space of the linear operators on $\mathscr{S}(\mathbb{R}^{4n})$. We define the linear maps

$$D^{s}_{\mu}:[\mathscr{A}^{k}_{n^1,\dots,n^N}] \to \mathscr{L}(\mathscr{S}(\mathbb{R}^{4n})), \quad n = n^1 + \dots + n^N,$$
$$\underline{\alpha}\{n\} \to D^{s}_{\mu}(\underline{\alpha}\{n\}), \quad s = 1,\dots,k, \quad \mu = 0,1,2,3,$$
$$(D^{s}_{\mu}(\underline{\alpha}\{n\}))(f)(x_1,\dots,x_n) = \left(\sum_i \frac{\partial}{\partial x^{\mu}_i} f\right)(x_1,\dots,x_n),$$
$$t \in \mathrm{ind}_s(\underline{\alpha}\{n\}), \quad t \neq 0. \tag{47}$$

Let $\{\underline{\alpha}\{n\}\}$ and $\{\underline{\xi}\{n\}\}$ be the bases of $[\underline{\mathscr{A}}^{k}_{n^1,\dots,n^N}]$ and $[\underline{\mathscr{A}}^{k}_{n^1,\dots,n^N}]^*$, respectively, $n = n^1 + \dots + n^N$. We define the linear maps

$$Q^{(n,k)s}_{\alpha}:(\underline{\bar{T}\bar{A}\bar{S}})^{k}_{n} \to (\underline{\bar{T}\bar{A}\bar{S}})^{k}_{n}, \quad \alpha = 1,2,3,4,$$
$$Q^{(n,k)s}_{\alpha}(\underline{\Phi}^{k}_{n})$$
$$= \sum_{\substack{\underline{\alpha} \in \{\underline{\alpha}\{n\}\} \\ \underline{\xi} = \underline{\alpha}^*}} \{\mathbb{C}_{\alpha\beta}[I_\Lambda \otimes \dots \otimes I_\Lambda \otimes \underbrace{D_\beta}_{s\text{th position}} \otimes I_\Lambda \otimes$$
$$\dots \otimes I_\Lambda \otimes (\underline{\alpha} \circ \underline{\xi}) \otimes I_s]$$
$$- i(\gamma^{\mu})_{\alpha\beta}[I_\Lambda \otimes \dots \otimes I_\Lambda \otimes \theta_\beta \otimes I_\Lambda \otimes \dots$$
$$\otimes I_\Lambda \otimes (\underline{\alpha} \circ \underline{\xi}) \otimes D^{s}_{\mu}(\underline{\alpha})](\underline{\Phi}^{k}_{n}), \tag{48}$$

where $D_\beta = \partial/\partial\theta_\beta$, and I_Λ and I_s are the identity maps in $[\Lambda(V)]$, and $\mathscr{S}(\mathbb{R}^{4n})$.

$$g^{(n,k)}_{s}:(\underline{\bar{T}\bar{A}\bar{S}})^{k}_{n} \to (\underline{\bar{T}\bar{A}\bar{S}})^{k}_{n}, \quad s = 1,\dots,k,$$
$$g^{(n,k)}_{s}(M_1 \otimes \dots \otimes M_k \otimes \underline{\alpha}\{n\} \otimes f^{\{n,\tau\}})$$
$$= (-1)^{\nu(M_1)+\dots+\nu(M_{s-1})}(M_1 \otimes \dots \otimes M_k \otimes \underline{\alpha}\{n\} \otimes f^{\{n,\tau\}}),$$

where $\nu(M_i)$ is the degree of M_i in $\Lambda(V)$. We define

$$Q^{(n,k)}_{\alpha} = \sum_{s=1}^{k} g^{(n,k)}_{s} \circ Q^{(n,k)s}_{\alpha}. \tag{49}$$

Let

$$\underline{Q}_\alpha^{(n)} = \{\underline{Q}_\alpha^{(n,1)},\ldots,\underline{Q}_\alpha^{(n,k)}\}$$

$$\underline{Q}_\alpha = \{\{\underline{Q}_\alpha^{(1)},\ldots,\underline{Q}_\alpha^{(n)},\cdots\}\} \quad \text{infinite sequence).} \tag{50}$$

Then $\underline{Q}_\alpha$ acts on $\underline{TAS}$ as follows:

$$\underline{Q}_\alpha(\underline{\Phi}) = \{\{0,\underline{Q}_\alpha^{(1)}\underline{\Phi}_1,\ldots,\underline{Q}_\alpha^{(n)}\underline{\Phi}_n,\cdots\}\}, \tag{51}$$

$$\underline{Q}_\alpha^{(n)}\underline{\Phi}_n = \{\underline{Q}_\alpha^{(n,1)}\underline{\Phi}^1_n,\ldots,\underline{Q}_\alpha^{(n,n)}\underline{\Phi}^n_n\}. \tag{52}$$

$\underline{Q}_\alpha$ is defined everywhere so that its domain $D(\underline{Q}_\alpha)$ is all of $\underline{TAS}$. We have, therefore, Proposition 4.

Proposition 4: $\underline{Q}_\alpha$ acts on $\underline{TAS}$ as a graded derivation of odd degree.

We want to examine the star properties of $\underline{Q}_\alpha$. In a star algebra a derivation δ is called symmetric if

$\delta(A^*) = (\delta(A))^*$, $A \in D(\delta)$ = the domain of δ and $*$ denoting the star operation.

Due to the graded behavior of $\underline{Q}_\alpha$ we must generalize this definition and for that purpose we introduce the following concept.

Definition 4: Let E be a semigraded $*$-algebra and J its main involution. A graded derivation δ on E is called *graded-symmetric* of degree ν when it satisfies the property

$$\delta(x^*) = (-J)^\nu\cdot(\delta(x))^*, \quad x \in D(\delta). \tag{53}$$

Considering a usual $*$-algebra as trivially graded, a symmetric derivation is a graded-symmetric with even degree. We can prove Proposition 5.

Proposition 5: $\underline{Q}_\alpha$ is a graded-symmetric derivation on $\underline{TAS}$ with odd degree.

The last two properties are by-products of the complexity of the definition of $\underline{Q}_\alpha$, but this very complexity is necessary for the validity of the main proposition of this work which is Proposition 6.

Proposition 6: The derivations $\underline{Q}_\alpha$ satisfy the supersymmetric algebra, i.e.,

$$\{\underline{Q}_\alpha,\underline{Q}_\beta\} = 2(\gamma^\mu\cdot\mathbb{C})_{\alpha\beta}\cdot iD_\mu, \tag{54}$$

where

$$D_\mu\underline{\Phi} = \{\{\cdots,\{\cdots,D_\mu\underline{\Phi}^k_n,\cdots\},\cdots\}\}, \tag{55}$$

and

$$(D_\mu{}^n f)(x_1,\ldots,x_n) = \left(\sum_{i=1}^n \frac{\partial}{\partial x_i^\mu}{}^n f\right)(x_1,\ldots,x_n).$$

B. Supersymmetric conditions: Physical interpretation

What we have established up to this point is that there exists a $*$-algebra, $\underline{TAS}$, on which supersymmetry can be represented as an algebra of graded derivations, and such that the spinorial charges are symmetric in some generalized sense. Also we have a map $\mathscr{T}$, from the algebra $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ of test functions for a given number of fields to $\underline{TAS}$, with the property that the images in $\underline{TAS}$ of the elements of $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ are the most likely generalizations of the classical superfields. That is, we have implemented the steps (1) and (2) of the proposed scheme in the Introduction.

From this point, there are two possible routes that one could follow. The first is to study the representaions of $\underline{TAS}$ on indefinite metric spaces, find the Hilbert subspaces and try to interpret the physical meaning of the action of the representations of supersymmetry on these subspcaes.

The second route is to implement the step (3) of the scheme. Here we follow this alternative. We make the following definitions.

Definition 5: $\underline{F} \in (\underline{TAS})'$ is called a *relative state* if $\underline{W}$, defined by

$$\underline{W}(\underline{f}) = \underline{F}(\underline{\mathscr{T}}\,\underline{f}) \tag{56}$$

is a state in $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$.

Definition 6: We say that an element $\underline{F} \in (\underline{TAS})'$ satisfies *Q-conditions* if

$$\underline{F}(\underline{Q}_\alpha\underline{\Phi}) = 0, \quad \alpha = 1,2,3,4, \quad \forall\underline{\Phi} \in \underline{TAS}. \tag{57}$$

$(\underline{TAS})'_Q$ denotes the set of all such elements. $\underline{F} \in (\underline{TAS})'$ satisfies *nth order Q-conditions if*

$$\underline{F}_n(\underline{Q}_\alpha\underline{\Phi}) = 0, \quad \forall\underline{\Phi} \in \underline{TAS}, \quad \alpha = 1,2,3,4. \tag{58}$$

Definition 7: We call a state $\underline{W}$ of $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$ an *inherited state* if it comes from at least one relative state of $\underline{TAS}$, i.e., if there exists a relative state $\underline{F}$ such that:

$$\underline{W}(\underline{f}) = \underline{F}(\underline{\mathscr{T}}\,\underline{f}), \quad \forall\,\underline{f} \in \underline{\mathscr{E}}\,\underline{\mathscr{S}} \tag{59}$$

We call a state $\underline{W} \in (\underline{\mathscr{E}}\,\underline{\mathscr{S}})'$ *Q-invariant* if it is an inherited state and at least on of its relative states is in $(\underline{TAS})'_Q$.

Now we can restate the conjecture made in the step (3) of the scheme

Conjecture: A Q-invariant Wightman state gives a Wightman theory and a representation of supersymmetry in this theory in such a way that the field operators transform covariantly under supersymmery.

If this conjecture is true then the Eq. (2) gives correlations of the structure functions appearing in the form of $\underline{F}$, while Eq. (1) transfers these correlations to the structure functions appearing in the form of $\underline{W}$. These last functions are the Wightman functions of the theory generated by $\underline{W}$.

Thus we need the structure of $\underline{F}$'s which are relative states and satisfy Q-conditions. For that purpose one must investigate the topological structure of $(\underline{TAS})'$ in general and try to establish representations of its elements in the form of generalized Källen–Lehmann representation. Here we will not touch this general problem but we shall restrict ourselves to the 2-point functions and try to show that at this level the proposed framework does indeed point to the right direction (i.e., the correlations are of the type of the perturbative results). As we will see below, for the free fields we can have general results.

IV. APPLICATION AND CONSISTENCY OF THE FRAMEWORK

A. Two-point functions for a general scalar supermultiplet

For the 2-point functions for a general field theory one can derive the Källen–Lehmann representation (KL),[16] which for our scalar supermultiplet is, for $x^0 > y^0$

$$W^{XX}(x-y) = i\int_0^\infty dm^2\,\sigma^X(m^2)\Delta^{(+)}(x-y;m^2),$$

$$X = A,F,G,D, \tag{60}$$

$$W^{YY}_{\alpha\dot\beta}(x-y) = (\Omega_W,\pi^Y_\alpha(x)\bar\pi^Y_\beta(y)\Omega_W)$$
$$= i\int_0^\infty dm^2\,\{w_1^Y(m^2) + w_2^Y(m^2)i\gamma^\mu\partial_\mu\}_{\alpha\beta}$$
$$\times\Delta^{(+)}(x-y;m^2).$$

$$Y=\psi,\chi. \tag{61}$$

Ω_W is the vacuum for the functional W and $\bar{\pi}$ means adjoint spinor

$$W^{A_\mu A_\nu}_{\mu\nu}(x-y)=i\int_0^\infty dm^2\,\{\sigma_1^{A_\nu}(m^2)g_{\mu\nu}+\sigma_2^{A_\nu}(m^2)\partial_\mu\partial_\nu\}\times\Delta^{(+)}(x-y;m^2). \tag{62}$$

We have

$$\sigma^X(m^2)\geqslant 0,\quad X=A,F,G,D; \tag{63a}$$

$$w_2^Y(m^2)\geqslant 0,$$

$$2mw_2^Y(m^2)\geqslant[w_1^Y(m^2)-mw_2^Y(m^2)]\geqslant 0. \tag{63b}$$

Motivated by the KL representation we make the following ansatz:

The functional $F\in(TAS)'$ which is a relative state and satisfies Q-conditions is such that its F_2^k component has the form

$$F_2^k(x,y)=\sum_{\tau_1,\dots\tau_k}\sum_{T_1,\dots T_k}Z^{\tau_1}_{T_1}\otimes\dots\otimes Z^{\tau_k}_{T_k}\otimes\underline{\xi}^{(2,2,k)}_{\begin{bmatrix}1&1\\1&2\end{bmatrix}(n_1,n_2)}\otimes\int_0^\infty dm^2\,i\rho^{T_1,\dots T_k}_{\tau_1,\dots\tau_k}\left(m^2;\frac{\partial}{\partial x}\right)\Delta^{(+)}(x-y;m^2), \tag{64}$$

where $\{\ \}$ denotes the appropriate summations over partitions as is explained in the Appendix, and $\{Z^\tau_T\}$ is the dual basis to $\{M^T_\tau\}$.

From this ansatz we can derive correlations between the structure functions $\rho^{T_1,\dots T_k}_{\tau_1,\dots\tau_k}$ using the Q-conditions [Eq. (2)]. Then, through Eq. (1), we can transfer these correlations to the weight functions appearing in the KL representation of the 2-point functions. We consider these correlations as the predictions of the theory. If we had at our disposal integral representations for n-point functions and an analogous ansatz, we could make predictions for such functions. (See comments below.)

We have

$$\underline{W}(\underline{f})=\sum_{\substack{T_1T_2\\ \tau_1\tau_2}}{}^2W^{T_1T_2}_{\tau_1\tau_2}({}^2f^{\bar{\tau}_1\bar{\tau}_2}_{T_1T_2}),$$

where

$${}^2W^{T_1T_2}_{\tau_1\tau_2}(x-y)=(\Omega_W,\pi^{T_1}_{\tau_1}(x)\pi^{T_2}_{\tau_2}(y)\Omega_W).$$

Now, from the Eq. (1) for $n=2$, we obtain a set of relations of the form

$$\sigma^A=\tfrac{1}{2}\{\rho^{AA}+\rho^A\},$$

$$\sigma^F=\tfrac{1}{32}\{\rho^D+\tfrac{1}{3}\rho^{DA}+\tfrac{1}{3}\rho^{AD}+\tfrac{1}{3}\rho^{FF}\},$$

etc.,

and constrains like

$$\rho^G+\tfrac{1}{3}\rho^{GA}+\tfrac{1}{3}\rho^{AG}-\tfrac{4}{3}\rho_2{}^{\psi\psi}=0,$$

etc.,

where

$$\rho^{YY}_{\alpha\beta}=\rho_1^{YY}\delta_{\alpha\beta}+\rho_2^{YY}(\gamma_5)_{\alpha\beta}+\rho_3^{YY}(i\gamma^\nu)_{\alpha\beta}\partial_\nu+\rho_4^{YY}(i\gamma^\nu\gamma_5)_{\alpha\beta}\partial_\nu.$$

$$Y=\psi,\chi.$$

We also obtain expressions for the mixed functions $W^{T_1T_2}$ in terms of the ρ's.

Now we exploit Eq. (2) for $n=2$, and we obtain a set of equations involving the functions ρ. Using these equations we can eliminate most of the structure functions from the expressions above for the weight functions and the W's. In this way we get the following relations among the weight functions of the supermultiplet:[11] For $m\neq 0$

$$\sigma^A=m^2w_2^\psi,\quad \sigma^F=\sigma^G=\tfrac{1}{32}\rho^D+\tfrac{1}{2}w_2^\psi,$$

$$\sigma_1^{A_\nu}=\tfrac{1}{32}\rho^D-\tfrac{1}{2}w_2^\psi,\quad \sigma_2^{A_\nu}=-(1/m^2)w_2^\psi,$$

$$w_1^\chi=0,\quad w_2^\chi=(1/m^2)w_2^\psi,$$

$$\sigma^D=(1/16m^2)w_2^\psi,$$

$$W^{AF}=W^{AG}=W^{FG}=W^{FD}=0,$$

$$\partial_\mu W_\nu^{AA_\nu}=\partial_\mu W_\nu^{FA_\nu}=\partial_\mu W_\nu^{GA_\nu}=\partial_\mu W_\nu^{DA_\nu}=0,$$

$$W^{AD}=i\int_0^\infty dm^2\left(\frac{1}{2\times 32}\right)\rho^D\Delta^{(+)}(x-y;m^2),$$

$$W^{\psi\chi}_{\alpha\beta}=i\int_0^\infty dm^2\left(\frac{1}{32}\right)\rho^D\delta_{\alpha\beta}\Delta^{(+)}(x-y;m^2). \tag{65}$$

Since $w_1^\chi=0$, (63b) does not hold for χ and we must conclude that χ does not represent physical degrees of freedom. Then the transitions to the field ψ through $\psi^\chi_{\alpha\beta}$ must vanish, i.e.,

$$W^{\psi\chi}_{\alpha\beta}=0,$$

and hence $W^{AD}=0$, and also, in all other 2-point functions, ρ^D will not contribution. Thus we have the independent fields A,F,G,D,A, and ψ.

For free fields and mass $M\neq 0$ for the fermion field ψ we must have

$$w_1^\psi(m^2)=M\delta(m^2-M^2),\quad w_2^\psi(m^2)=\delta(m^2-M^2),$$

$$\rho^D=0. \tag{66}$$

Then we get

$$W^{A_\mu A_\nu}_{\mu\nu}(x-y)=\frac{i}{2}\left[-g_{\mu\nu}-\frac{2}{M^2}\partial_\mu\partial_\nu\right]\times\Delta^{(+)}(x-y;M^2), \tag{67}$$

which cannot give the propagator of a free vector field. Thus A_ν does not provide a physical degree either.

Let us define the following fields

$$\phi_1=MA,\quad \phi_2=\sqrt{2}F,\quad \phi_3=\sqrt{2}G,\quad \phi_4=4mD. \tag{68}$$

Then

$$\sigma^{\phi_1}=\sigma^{\phi_2}=\sigma^{\phi_3}=\sigma^{\phi_4}=w_2^\psi=\delta(m_2-M_2). \tag{69}$$

Thus in the case of free fields we have three scalar fields ϕ_1,ϕ_2,ϕ_4, one pseudoscalar ϕ_3, and one spinor field ψ, all with equal mass M. We may state the above results as follows.

Proposition 7: If the Wightman state for free fields is an inherited state then there exists one relative state which satisfies the second-order Q-conditions.

In Sec. V we indicate not only that the free fields with equal mass come necessarily from an inherited state, but that this state is Q-invariant (i.e., satisfies, Q-conditions of all orders).

B. Free fields and Q-conditions

This part of the work is completely technical and we omit all details. We simply indicate the main points.

Definition 8:Let $\{T_1,...,T_n\}$ be an ordered n-tuple of representations where $T_i = A,\psi...,D$ and let it contain $N(S_i)$ representations of S_i type and suppose that there exist only m types where $m \leqslant \min(n,7)$ and the spinor representations are counted together with their adjoints. We say that the configuration $\{T_1,...,T_n\}$ is *properly paired* if all $N(S_i)$, $i = 1,...,n$ are even numbers. Then n is even.

A Wightman state $\underline{\underline{W}}^0 \in (\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}})'$ that gives a field theory for free field has the following structure:

$$\underline{\underline{W}}^0 = \{1,0,\underline{W}^0_2,...,\underline{W}^0_{2n},...\}, \tag{70}$$

$$\underline{W}^0_{2n} = \sum_{\substack{T_1,...T_{2n} \\ \{T_1,...,T_{2n}\}_{pp}}} {}^0\underline{W}^{T_1,...T_{2n}}_{2n}, \tag{71}$$

where $\{T_1,...,T_{2n}\}_{pp}$ means that ${}^0\underline{W}^{\{T\}}_{2n}$ contains only properly paired configurations.

Since $\underline{\underline{W}}^0$ describes free fields ${}^0\underline{W}^{\{T\}}_{2n}$ can be written as a product of 2-point functions. Then by interchanging the factors and taking into account the various minus signs comming from the fermions we can bring ${}^0\underline{W}^{\{T\}}_{2n}$ into a form which motivates the construction of an appropriate relative state for free fields. Having at hand this appropriate state it is straightforward to prove.

Proposition 8: $\underline{\underline{W}}^0$ is an inherited state.

Proposition 9: $\underline{\underline{W}}^0$ is Q-invariant.

V. DISCUSSION

We proposed a framework in which a Wightman theory can be embedded in such a way that one is able to give criteria for a supersymmetric theory, without having either to mix the Minkowski coordinates and the fields with unphysical Grassmann parameters or to answer the question of the integrability of the supersymmetric algebra.

Also the possibility of the appearance of indefinite metric Hilbert spaces, in this framework, is elevated to the question of the existence of continuous functionals satisfying certain properties (positivity, locality, etc.).

Nevertheless the whole approach relies on a conjecture, the justification of which is still open. Of course the conjecture was motivated by the state of things in the case of C^*-algebras and of usual symmetries and is meant to give a natural extention for the case of nuclear *-algebras and graded-Lie algebras.

The justification of the procedure could be made only at the level of 2-point Wightman functions and under the assumption that the conventional field theoretic results point to the right direction. If we had a generalization of the Källen–Lehmann representation for n-point functions we could study the implications of the framework for this level. Thus we have not proved that our characterization of a supersymmetric Wightman theory is either complete or totally equivalent to the conventional conception of supersymmetry. But the scheme does include the case of free fields.

There are many open problems which have to be solved before the proposed framework is fully axiomatic. One category of problems contains purely mathematical questions and concerns the generalizations to the case of nuclear *-algebras and graded Lie-algebras of the results proved already for C^*-algebras and usual Lie algebras.

Another category of problems concerns the proof of a universality property for $\underline{TAS}$; namely that $\underline{TAS}$ is either unique or minimal, and that the map: $\underline{\mathscr{E}}\,\underline{\mathscr{S}} \to \underline{TAS}$ is a natural extension of the superfield concept.

Finally one could study the possibility of generalized KL representations using the topological properties of the dual of $\underline{\mathscr{E}}\,\underline{\mathscr{S}}$, and then use these representations to study the n-point Wightman functions of a supersymmetric theory (in our sense).

Concluding this work we should point out that, although our framework is not fully developed, it could be taken as a prototype for the study of any graded-Lie algebras which are thought to generated a symmetry of a field theory.

ACKNOWLEDGMENTS

I would like to thank Professor R.F. Streater for very helpful discussions, and the State Scholarships Foundation of Greece for a scholarship during the time of this work.

APPENDIX

Conventions and identities

$$g^{\mu\nu} = (1,-1,-1,-1), \quad \{\gamma^\mu,\gamma^\nu\} = 2g^{\mu\nu},$$

$$\sigma_{\mu\nu} = (i/2)[\gamma_\mu,\gamma_\nu],$$

$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3, \quad \gamma^\dagger_\mu = \beta\gamma_\mu\beta^{-1},$$

$$-\gamma^{\sim}_\mu = C^{-1}\gamma_\mu C, \quad C^\dagger = C^{-1}, \quad C^{\sim} = -C, \quad C\gamma_5 = \gamma^{\sim}_5 C,$$

$$C\gamma^{\sim}_5 = \gamma_5 C, \quad \bar{\psi} = \psi^\dagger\beta.$$

For a Majorana representation we have

$$\gamma^{\mu *} = -\gamma^\mu, \quad \sigma^{\mu\nu *} = -\sigma^{\mu\nu}, \quad \gamma^{\mu\dagger} = g^{\mu\mu}\gamma^\mu,$$

$$\sigma^{\mu\nu\dagger} = g^{\mu\mu}g^{\nu\nu}\sigma^{\mu\nu}, \quad \gamma^\dagger_5 = \gamma^{\sim}_5 = -\gamma_5,$$

$$(\gamma_5\gamma^\mu)^\dagger = g^{\mu\mu}\gamma_5\gamma^\mu, \quad (\gamma_5\sigma^{\mu\nu})^\dagger = -g^{\mu\mu}g^{\nu\nu}\gamma_5\sigma^{\mu\nu},$$

$$\beta = b\gamma_0, \quad C = c\gamma_0,$$

$$b = 1, \quad c = -1.$$

More properties of this representation may be found in L. Gorwin *et al.* (Ref. 2). We collect some useful identities used in this work:

$$\frac{\partial}{\partial\theta_\beta}\theta_\gamma = \delta_{\beta\gamma},$$

$$\theta_\beta\theta_\gamma = \tfrac{1}{4}\{-(\gamma_0)_{\beta\gamma}\bar{\theta}\theta + (\gamma_5)_{\beta\gamma}\bar{\theta}_{\gamma_5}\theta + (i\gamma^\nu\gamma_5\gamma_0)_{\beta\gamma}\bar{\theta}i_{\gamma_\nu\gamma_5}\theta\},$$

$$\frac{\partial}{\partial\theta_\beta}\bar{\theta}\theta = 2\theta_\delta(\gamma_0)_{\beta\delta}, \quad \theta_\beta\bar{\theta}\theta = \bar{\theta}\theta\theta_\beta$$

$$\frac{\partial}{\partial\theta_\beta}\bar{\theta}\theta\theta_\gamma = \tfrac{1}{2}\{\delta_{\gamma\beta}\bar{\theta}\theta + (\gamma_5)_{\gamma\beta}\bar{\theta}_{\gamma_5}\theta + (i\gamma^\nu\gamma_5)_{\gamma\beta}\bar{\theta}i_{\gamma_\nu\gamma_5}\theta\},$$

$$\theta_\beta\bar{\theta}\theta\theta_\gamma = \tfrac{1}{4}(\bar{\theta}\theta)^2(\gamma_0)_{\gamma\beta},$$

$$\frac{\partial}{\partial\theta_\beta}\bar{\theta}_{\gamma_5}\theta = 2\theta_\gamma(\gamma_0\gamma_5)_{\beta\gamma},$$

$$\theta_\beta\bar{\theta}_{\gamma_5}\theta = \bar{\theta}\theta\theta_\gamma(\gamma_5)_{\gamma\beta},$$

$$\frac{\partial}{\partial\theta_\beta}\bar\theta i_{\gamma_\nu\gamma_5}\theta = 2(\gamma_0 i\gamma_\nu\gamma_5)_{\beta\gamma}\theta_\gamma,$$

$$\theta_\beta\bar\theta i_{\gamma_\nu\gamma_5}\theta = \bar\theta\theta_\gamma(\gamma_0\gamma_5 i\gamma_\nu\gamma_0)_{\gamma\beta},$$

$$\frac{\partial}{\partial\theta_\beta}(\bar\theta\theta)^2 = 4\bar\theta\theta_\gamma(\gamma_0)_{\beta\gamma},\quad \theta_\beta(\bar\theta\theta)^2 = 0,$$

$$\bar\theta\theta\bar\theta_{\gamma_5}\theta = \bar\theta_{\gamma_5}\theta\bar\theta\theta = 0,\quad \bar\theta\theta\bar\theta i_{\gamma_\nu\gamma_5}\theta = \bar\theta i_{\gamma_\nu\gamma_5}\theta\bar\theta\theta = 0,$$

$$\bar\theta_{\gamma_5}\theta\bar\theta i_{\gamma_\nu\gamma_5}\theta = \bar\theta i_{\gamma_\nu\gamma_5}\theta\bar\theta_{\gamma_5}\theta = 0,$$

$$\bar\theta_{\gamma_5}\theta\bar\theta_{\gamma_5}\theta = (\bar\theta\theta)^2,\quad \bar\theta i_{\gamma_\mu\gamma_5}\theta\bar\theta i_{\gamma_\nu\gamma_5}\theta = (\bar\theta\theta)^2 g_{\mu\nu}.$$

The Borchers' algebra $\underline{\underline{\mathscr{E}}}\,\underline{\underline{\mathscr{S}}}$

Sum and multiplication by scalar:

$$\underline{\underline{f}} + \lambda\underline{\underline{g}} = \{f_0 + \lambda g_0, \underline{f}_1 + \lambda\underline{g}_1,\ldots,\underline{f}_n + \lambda\underline{g}_n,\cdots\}. \tag{A1}$$

Product:

$$(\underline{\underline{f}}\times\underline{\underline{g}})_n = \sum_{k=1}^{n}\underline{f}_k\times\underline{g}_{n-k},$$

$$(\underline{f}_k\times g_{n-k})(x_1,\ldots,x_n),\quad x_i\in\mathbb{R}^4,$$

$$= \sum_{|\bar T,\tau|}\sum_{|S,\tau|} e^{T_1}_{\tau_1}\otimes\cdots\otimes e^{T_k}_{\tau_k}\otimes e^{S_1}_{\sigma_1}\otimes\cdots\otimes e^{S_{n-k}}_{\sigma_{n-k}}$$

$$\otimes\,{}^k f^{\bar\tau_1\cdots\bar\tau_k}_{\bar T_1\cdots\bar T_k}(x_1,\ldots,x_k)\times{}^{n-k}g^{\bar\sigma_1\cdots\bar\sigma_{n-k}}_{\bar S_1\cdots\bar S_{n-k}}(x_{k+1},\ldots,x_n). \tag{A2}$$

Star operation:

$$\underline{\underline{f}}^* = \{f_0^*,\underline{f}_1^*,\ldots\underline{f}_n^*,\cdots\},\quad (\underline{\underline{f}}^*)_n = \underline{f}_n^*,\quad f_0^* = \bar f_0,$$

$$\underline{f}_n^*(x_1,\ldots,x_n) = \sum_{|\bar T,\tau|}\left[e^{T_1}_{\tau_1}\otimes\cdots\otimes e^{T_n}_{\tau_n}\otimes{}^n f^{\bar\tau_1\cdots\bar\tau_n}_{\bar T_1\cdots\bar T_n}\right]^*(x_1,\ldots,x_n),$$

$$= \sum_{|\bar T,\tau|} e^{T_n}_{\tau_n}\otimes\cdots\otimes e^{T_1}_{\tau_1}\otimes\overline{{}^n f^{\bar\tau_1\cdots\bar\tau_n}_{\bar T_1\cdots\bar T_n}(x_n,\ldots,x_1)},\quad x_i\in\mathbb{R}^4. \tag{A3}$$

The indexing algebras $\underline{\mathscr{A}}^k$

In the text some motivations were presented for the need and the complexity of the indexing algebra. We repeat here that the elements of this algebra guide the action of the operator Q_α on the elements of $\underline{\underline{TAS}}$. This action must involve the operation of taking the partial derivatives with respect to subsets of the set of arguments $(x_1,\ldots,x_n)$. This set of arguments can appear either as a part of a bigger set, the latter formed after multiplication of two elements of $\underline{\underline{TAS}}$, or as composite set formed by the multiplication of two or more elements. Various inversions of order of the parts of subsets or of the subsets themselves due to the star operation, must be taken into account, in order to guarantee the desired properties of $\underline{\underline{Q}}_\alpha$ under conjugation.

The elements of the indexing can be thought of as symbols which contain the information associated to the way a given set of arguments is formed from its subsets (permutations of arguments, inversion of order, permutation of subsets, combination of subsets, etc.). The elements of the indexing algebra of order k, $\underline{\mathscr{A}}^k$, are k-tuples of symbols. We indicate how a general k-tuple is constructed.

Let the following partitions for $k\leqslant l\leqslant n$:

$$r_1+\ldots+r_l = n,\quad 0\leqslant r_i\leqslant n,\quad i=1,\ldots,l.$$

$$n_1+\ldots+n_k = l,\quad 0\leqslant n_j\leqslant l,\quad j=1,\ldots,k.$$

$$R_t = r_1+\ldots+r_{t-1},\quad R_1 = 0,$$

$$\bar R_s = n_1+\ldots+n_{s-1},\quad \bar R_1 = 0.$$

Note the possibility of vanishing r_i, n_j. We introduce the following indexed ordered sets of integers:

$${}_t\alpha^{(n,l)}_{r_t} = \{n,l\,|R_t+1,\ldots,R_t+r_t\},\quad r_t\neq 0,$$
$$= \{n,l\,|0\},\quad r_t = 0,$$

$${}_t\tilde\alpha^{(n,l)}_{r_t}$$
$$= \{n,l\,|n-(R_t+r_t)+1,\ldots,n-(R_t+1)+1\},\quad r_t\neq 0,$$
$$= \{n,l\,|0\},\quad r=0.$$

We define the composition

$${}_t\alpha^{(n,l)}_{r_t}\circ_{t'}\alpha^{(n,l)}_{r_{t'}} = \{n,l\,|R_t+1,\ldots,R_t+r_t,R_{t'}+1,$$
$$\ldots,R_{t'}+r_{t'}\}.$$

Let the permutation $(1,\ldots,l)\to(i_1,\ldots,i_l)$. We introduce the composite symbol

$${}^s\alpha^{(n,l,k)}_{\begin{bmatrix}r_1,\ldots r_l\\ i_1,\ldots i_l\end{bmatrix}(n_s)} = {}_{i_{\bar R_s+1}}\alpha^{(n,l)}_{r_{i_{\bar R_s+1}}}\circ\ldots\circ_{i_{\bar R_s+n_s}}\alpha^{(n,l)}_{r_{i_{\bar R_s+n_s}}},\quad n_s\neq 0,$$
$$= \{n,l\,|0\},\quad n_s = 0.$$

These symbols have the structure: $\{n,l\,|m_1,\ldots,m_p\}$.

We define a composition rule:

$${}^s\alpha^{(n,l,k)}_{[\cdots](n_s)}\circ{}^s\alpha^{(\bar n,\bar l,k)}_{[\cdots](\bar n_s)}$$
$$= \{n+\bar n, l+\bar l\,|m_1,\ldots,m_p,n+\bar m_1,\ldots,n+\bar m_q\}.$$

This rule guarantees the property of $\underline{\underline{Q}}_\alpha$ as a derivation on $\underline{\underline{TAS}}$. We introduce the k-tuples of symbols

$$\underline{\alpha}^{(n,l,k)}_{[\cdots](n_1,\ldots n_k)} = \{{}^1\alpha^{(n,l,k)}_{[\cdots](n_1)},\ldots,{}^k\alpha^{(n,l,k)}_{[\cdots](n_k)}\},$$

and a composition rule

$$\underline{\alpha}^{(n,l,k)}_{[r_1\cdots](n_1,\cdots)}\times\underline{\alpha}^{(\bar n,\bar l,k)}_{[\bar r_1\cdots](\bar n_1,\cdots)}$$
$$= \{\ldots,{}^s\alpha^{(n,l,k)}_{[r_1\cdots](n_s)}\times{}^s\alpha^{(\bar n,\bar l,k)}_{[\bar r_1\cdots](\bar n_s)},\ldots\}. \tag{A4}$$

Let for given k and N pairs of integers n^i, l^j, $j=1,\ldots,N$, such that, $k\leqslant l^j\leqslant n_j$, $\underline{\mathscr{A}}^k_{n^1\ldots n^N}$ be the set of k-tuples of symbols that can be constructed by composing the k-tuples $\underline{\alpha}^{(n^j,l^j,k)}_{[\cdots](\cdots)}$. The general element of $\underline{\mathscr{A}}^k_{n^1\ldots n^N}$ has the form $\underline{\alpha}^{(\cdots)\cdots(\cdots)}_{[\cdots](\cdots)\cdots[\cdots](\cdots)}$. Let $[\underline{\mathscr{A}}^k_{n^1\ldots n^N}]$ be the vector space over $\mathbb{C}$ generated by the set $\underline{\mathscr{A}}^k_{n^1\ldots n^N}$. Let

$$\underline{\mathscr{A}}^k = \sum_{\substack{N,n^1\ldots n^N\\ k\leqslant n^j}}\oplus\left[\underline{\mathscr{A}}^k_{n^1\ldots n^N}\right] \tag{A5}$$

We define a product $\underline{\mathscr{A}}^k$ in such a way that on every $[\underline{\mathscr{A}}^k_{n^1\ldots n^N}]$ coincides with the defined above composition of k-tuples. A star operation can easily be introduced in $\underline{\mathscr{A}}^k$ using the elements ${}_t\tilde\alpha^{(n,l)}_{r_t}$ and $\underline{\mathscr{A}}^k$ becomes an associative *-algebra.[11]

The algebra TAS

We only explain here the notation in the formulas appearing in the text. We simplified the expressions with the conventions:

$$M^{T_i}_{\tau_i} = M_i, M^{T_i}_{\tau_i} = M_i,\ \underline{\alpha}^{(\cdots)\cdots(\cdots)}_{[\cdots](\cdots)\cdots[\cdots](\cdots)} = \alpha\{n\}, f^{(\cdots)\cdots(\cdots),\tau_1\cdots\tau_k}_{[\cdots]\cdots[\cdots],T_1\cdots T_k}$$
$$= f\{{}^{n,\tau}_{T}\}.$$

$$\sum_{s=1}^{N}\sum_{\substack{k\leqslant l^s\leqslant n^s\\ r^s_1+\ldots+r^s_{l^s}=n^s\\ 0<r^s_t\leqslant n^s, t=1,\ldots,l^s}}\ \sum_{\substack{0\leqslant n^s_q\leqslant l^s\\ q=1,\ldots,k\\ n^s_1+\ldots+n^s_k=l^s}} = \sum_{|\ |}.$$

$i^s_1\ldots i^s_{l^s}$ over all perm. of $(1,\ldots,l^s)$

Also

$$\lambda = \nu(N_1)[\nu(M_2) + \ldots + \nu(M_k)] + \nu(N_2)[\nu(M_3) + \ldots + \nu(M_k)] \cdots + \nu(N_{k-1})\nu(M_k).$$

[1]A. Neveu and J. Schwartz, Nucl. Phys. B **31**, 86 (1971); J. Gervais and B. Sakita, Nucl. Phys. B **34**, 633 (1971); P. Ramond, Phys. Rev. D **3**, 2415 (1971); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974); A. Salam, and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); L. O'Raifeartaigh, Phys. Rev. B **139**, 1053 (1965); and Dublin Lecture Notes, 1975.

[2]A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); Phys. Rev. D **11**, 6 (1975); P. Goddard, Nucl. Phys. B **88**, 429 (1975); W. Ruhl and B. Yunn: Universitat Trier-Kaiserlauten Preprints **1**, 2 (1974); J. Kalney, Lett. Nuovo Cimento **11**, 577 (1974); R. Kotecky, Lett. Nuovo Cimento **13**, 269 (1975); C. Ktorides, J. Math. Phys. **16**, 2123 (1975); L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).

[3]F. Berezin and G. Kac, Mat. Sb. **11**, 311 (1970); F. Berezin, *The Method of Second Quantization* (Academic, New York, 1966); J. Dieudonne, *Introduction to the Theory of Formal Groups* (Marcel-Dekker, New York (1973).

[4]J. Bognar, *Indefinite Inner Product Spaces*, (Springer, Berlin, 1974).

[5]R. Haag and D. Kastler: J. Math. Phys. **5**, 848 (1964); G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory*, Wiley-Interscience, New York, 1972); S. Doblicher, R. Haag and J. Roberts, Comm. Math. Phys. **13**, 1 (1969); **25**, 173 (1969); **23**, 199 (1971); **49**, 35 (1974).

[6]R.F. Streater and A. Wightman: *PCT Spin and Statistics and all That* Benjamin, New York, 1964); R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, RI 1966).

[7]H.J. Borchers, Nuovo Cimento **24**, 214 (1962); H.J. Borchers, *Algebraic Aspects of Wightman Field Theory, in Statistical Mechanics and Field Theory*, edited by R.N. Sen and C. Weil, Lectures, Haifa, 1971; W. Wyss, *On Wightmans's Theory Of Quantized Fields*, edited by K. Mahanthappa and W. Britten, in Vol. XI-D of Lectures in Mathematical Physics, Gordon and Breach, New York, 1969).

[8]R. Haag, J. Lopuszanski, and M. Sohnius, Nucl. Phys. B **80**, 257 (1975); L. O'Raifeartaigh, Ref. 1; S. Coleman and J. Mandula: Phys. Rev. B **159**, 1251 (1967); D. Grosser, Universitat Tubingen Preprint, 1975; L. Corwin *et al.*, Ref. 2.

[9]S. Sakai, Recent developments in the Theory of Unbounded derivations in C^*-algebras, Preprint (1977).

[10]S.M. Moore, Lett. Nuovo Cimento **22**, 41 (1978).

[11]D. Ghikas, Thesis, University of London, 1978.

[12]C. Chevalley, *The Construction and the Study of Certain Important Algebras* (Tokyo Math. Society of Japan, Tokyo, 1955).

[13]G. Birkoff and S. MacLane, *Algebra*, (MacMillan, New York, 1967).

[14]F. Treves: *Topological Vector Spaces, Distributions and Kernels*, (Academic, New York, 1967).

[15]A. Robertson and W. Robertson, *Topological Vector Spaces* (Cambridge, U.P., Cambridge, England, 1964).

[16]S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, Row, Peterson, Evanston, IL 1961); J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1964).

Upper bound on the spin-flip cross section from unitarity, total cross section and the forward slope

I. A. Sakmar[a),b)]
Université de Nice, Laboratoire de Physique Théorique,[c)] Parc Valrose, 06034 Nice Cedex, France

(Received 22 July 1980; accepted for publication 28 August 1980)

For the spin 0–spin $\frac{1}{2}$ particle scattering we give a method to construct an upper bound on the spin-flip cross section with the variational calculus using as constraints the total cross section, the forward slope and the full unitarity.

PACS numbers: 11.80.Cr, 11.80.Et

I. INTRODUCTION

Since the 1964 paper of Mac Dowell and Martin[1] much work[2–13] has been done in the application of the variational techniques for finding bounds in particle physics problems, so that the topic has become almost a field in itself. Among the many advances made one could cite the systematic treatment of the inequality constraints,[3,4] such as the unitarity, and the generalization of the method to the cases with spin.[7,8] However, the field has its limitations. There are not many observables which lend themselves to this kind of approach. A number of papers are devoted to the derivative or second derivative of the imaginary part of the amplitude in the forward direction. The method requires a certain consistency in the forms of the quantity to be maximized and the constraints. Otherwise mathematics becomes too complicated. Even so some of the cases investigated require numerical calculations. But the work has been fruitful so that today there exists several rigorous bounds which are based solely on such solid theoretical grounds as the unitarity and such well-known experimental quantities as σ^T and σ^{El}.

To our knowledge the inequality constraints in their full form have not been used before in problems involving spin. In this paper we are applying the variational technique to a spin 0–spin $\frac{1}{2}$ scattering problem to find a bound on the spin-flip cross section when the total cross section and the forward slope are known and the full content of the unitarity is used. For this we follow the elegant treatment of inequality constraints by Einhorn and Blankenbecler.[3] Since we are dealing with the spin case there are two sets of partial wave amplitudes. The variational equations mix those amplitudes and the unitarity imposed in the form of inequality constraints assigns to the amplitudes characteristics such that they can be divided into different classes. The analysis of these classes with the purpose of choosing the Lagrange multipliers for maximizing the spin-slip cross section leads to the solution of the problem.

In Sec. II we set up the equations, write the auxiliary function of Lagrange and differentiating it with respect to the variables we obtain four basic equations. We define the four classes determined by the unitarity equations. Taking second derivatives of the function of Lagrange we find next the maximum conditions.

In Sec. III we analyze the forms of the partial wave amplitudes in the four classes defined before. By imposing the unitarity as well as the maximum condition, the forms of the partial waves are determined. Also the conditions to be satisfied by Lagrange multipliers are found. Subject to these conditions fitting of the total cross section and the forward slope with those parameters and the l values determined by the parameters gives us the values of the multipliers. Spin-flip cross section is then found in terms of σ^T and the foreward slope.

In the Discussion and Conclusion we summarize and discuss our results.

II. FORMALISM

To simplify the formulas we define G, A_0 and S in terms of $\sigma_{\rm SF}$, σ^T and $dA/dt|_{t=0}$ as follows:

$$G = \frac{k^2}{2\pi}\sigma_{\rm SF} = \sum \frac{2l(l+1)}{2l+1}\left[(a_{l+} - a_{l-})^2 + (r_{l+} - r_{l-})^2\right], \tag{1}$$

$$A_0 = \frac{k^2}{4\pi}\sigma^T = \sum \left[(l+1)a_{l+} + la_{l-}\right], \tag{2}$$

$$S = 4k^2\frac{k}{\sqrt{s}}\frac{dA}{dt}\bigg|_{t=0} = \sum l(l+1)\left[(l+1)a_{l+} + la_{l-}\right]. \tag{3}$$

Here $\sigma_{\rm SF}$ is the spin-flip cross section, σ^T the total cross section, A the imaginary part of the scattering amplitude $dA/dt|_{t=0}$ the forward slope, k the c.m. momentum, a_{l+}, a_{l-}, r_{l+}, r_{l-} the imaginary and real parts of the partial waves.

In addition to the equality constraints (2) and (3) we also have the inequality constraints of unitarity:

$$u_l = a_{l+} - a_{l+}^2 - r_{l+}^2 \geqslant 0, \tag{4}$$

$$v_l = a_{l-} - a_{l-}^2 - r_{l-}^2 \geqslant 0. \tag{5}$$

We want to maximize G at a fixed energy subject to the constraints (2)–(5).

The auxiliary function of Lagrange is written in the form

$$L = G + \alpha A_0 + \beta S + \sum (l+1)\lambda_l u_l + \sum l\mu_l v_l. \tag{6}$$

Here $\lambda_l \geqslant 0$, $\mu_l \geqslant 0$ from the theory of inequality constraints. The factors $(l+1)$ and l in front of λ_l and μ_l are arbitrarily chosen by changing the definition of these multipliers to

[a)] On leave from the Applied Mathematics Department, University of Western Ontario, London, Ontario, Canada.
[b)] Research supported in part by the National Research Council of Canada.
[c)] Equipe de Recherche Associée au C.N.R.S.

 022-2488/81/030600-05$1.00

make the resulting equations simpler. To further simplify the formulas we shall show the following, frequently appearing combinations of l by B and D:

$$B \equiv 2l/(2l+1), \quad D \equiv 2(l+1)/(2l+1). \tag{7}$$

Next we differentiate L with respect to the four types of variables:

$$\partial L/\partial a_{l+} = 0 \text{ gives } (B-\lambda_l)a_{l+} - Ba_{l-} + \tfrac{1}{2}[\alpha + l(l+1)\beta + \lambda_l] = 0, \tag{8}$$

$$\partial L/\partial a_{l-} = 0 \text{ gives } Da_{l+} - (D-\mu_l)a_{l-} - \tfrac{1}{2}[\alpha + l(l+1)\beta + \mu_l] = 0, \tag{9}$$

$$\partial L/\partial r_{l+} = 0 \text{ gives } (B-\lambda_l)r_{l+} - Br_{l-} = 0, \tag{10}$$

$$\partial L/\partial r_{l-} = 0 \text{ gives } Dr_{l+} - (D-\mu_l)r_{l-} = 0. \tag{11}$$

We now define the following four classes:

$$I^+I^- = \{l \,|\, u_l > 0, v_l > 0\}\ \lambda_l = 0, \quad \mu_l = 0, \tag{12}$$

$$I^+B^- = \{l \,|\, u_l > 0, v_l = 0\}\ \lambda_l = 0, \quad \mu_l \geqslant 0, \tag{13}$$

$$I^-B^+ = \{l \,|\, u_l = 0, v_l > 0\}\ \lambda_l \geqslant 0, \quad \mu_l = 0, \tag{14}$$

$$B^+B^- = \{l \,|\, u_l = 0, v_l = 0\}\ \lambda_l \geqslant 0, \quad \mu_l \geqslant 0. \tag{15}$$

It is the definition of these classes which generalizes the formalism to the spin case and also makes it possible to study even higher spin cases. For the case of spin 0–spin $\frac{1}{2}$ scattering there are only two types of partial wave amplitudes and a pair f_{l+}, f_{l-} defined by a fixed value of l belongs to one and only one of these four classes.

To find the maximum conditions we take the second derivatives of L,

$$\partial^2 L/\partial a_{l+}\,\partial a_{l+} = 2(l+1)(B-\lambda_l), \tag{16}$$

$$\partial^2 L/\partial a_{l+}\,\partial a_{l-} = -BD(2l+1), \tag{17}$$

$$\partial^2 L/\partial r_{l+}\,\partial r_{l+} = 2(l+1)(B-\lambda_l), \tag{18}$$

$$\partial^2 L/\partial r_{l+}\,\partial r_{l-} = -BD(2l+1), \tag{19}$$

$$\partial^2 L/\partial a_{l-}\,\partial a_{l-} = 2l(D-\mu_l), \tag{20}$$

$$\partial^2 L/\partial r_{l-}\,\partial r_{l-} = 2l(D-\mu_l). \tag{21}$$

All other derivatives vanish. Negative definiteness of the second variation for the maximum gives the conditions

$$\lambda_l \geqslant B, \tag{22}$$

$$\mu_l \geqslant D. \tag{23}$$

III. FOUR CLASSES OF THE PARTIAL WAVES

Class I^+I^-: Since in this class $\lambda_l = 0, \mu_l = 0$ the basic Eqs. (8) and (9) become:

$$B(a_{l+} - a_{l-}) + \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0, \tag{24}$$

$$D(a_{l+} - a_{l-}) - \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0. \tag{25}$$

These two equations are incompatible except when $\alpha + l(l+1)\beta = 0$. But α and β are l-independent and can satisfy this only when they are zero. This would mean no constraints and therefore are not acceptable. Thus the class I^+I^- must be empty.

Class I^+B^-: In this class $\lambda_l = 0, \mu_l \geqslant 0$. The four Eqs. (8)–(11) become:

$$Ba_{l+} - Ba_{l-} + \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0, \tag{26}$$

$$Da_{l+} - D - \mu_l a_{l-} - \tfrac{1}{2}[\alpha + l(l+1)\beta + \mu_l] = 0, \tag{27}$$

$$Br_{l+} - Br_{l-} = 0, \tag{28}$$

$$Dr_{l+} - [D-\mu_l]\,r_{l-} = 0. \tag{29}$$

Equation (28) gives $r_{l+} = r_{l-}$. Equation (29) gives $\mu_l r_{l-} = 0$, for $r_{l-} \neq 0$, $\mu_l = 0$. In this case Eq. (27) becomes

$$D(a_{l+} - a_{l-}) - \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0.$$

But as we saw before this is not compatible with Eq. (26). Hence we must have

$$r_{l-} = 0.$$

In this case also

$$r_{l+} = 0.$$

Because in this class $\lambda_l = 0$, we have $v_l = a_{l-} - a_{l-}^2 = 0$. Hence

$$a_{l-} = \begin{cases} 1 \\ 0 \end{cases}.$$

1) If $a_{l-} = 1$, Eqs. (26) and (27) give

$$a_{l+} = 1 - [\alpha + l(l+1)\beta]/2B \tag{30}$$

and

$$\mu_l = [\alpha + l(l+1)\beta]\,(2l+1)/l. \tag{31}$$

We now impose the unitarity condition $0 \leqslant a_{l+} \leqslant 1$ on Eq. (30) and the maximum condition $\mu_l \geqslant D$ on Eq. (31) and obtain

$$0 \leqslant \alpha + l(l+1)\beta \leqslant 4l/(2l+1), \tag{32}$$

$$\alpha + l(l+1)\beta \geqslant 2l(l+1)/(2l+1)^2. \tag{33}$$

The contributions of $a_{l+} = 1$ and a_{l-} as given by Eq. (46) to A_0 and S are obtained as before by summing the series. addition formulas for powers of integers except one term in A_0. The results are

$$A_0 = (L_2+1)^2 - L_1^2 - \tfrac{1}{4}\alpha[(L_2+2)^2 - (L_1+1)^2] - \tfrac{1}{4}\alpha\sum_{L_1}^{L_2}\frac{1}{l} - \tfrac{1}{24}\beta[(L_2+1)(L_2+2) \times (3L_2^2 + 7L_2 + 3) - L_1(L_1+1)(3L_1^2 + L_1 - 1)], \tag{34}$$

$$S = \tfrac{1}{2}[L_2(L_2+1)^2(L_2+2) - (L_1-1)L_1^2(L_1+1)] - \tfrac{1}{24}\alpha[(L_2+1)(L_2+2)(3L_2^2 + 7L_2 + 3) - L_1(L_1+1)(3L_1^2 + L_1 - 1)] - \tfrac{1}{240}\beta[L_2(L_2+1) \times (L_2+2)(2L_2+1)(2L_2+3)(5L_2+1) - (L_1-1)L_1(L_1+1)(2L_1-1)(2L_1+1)(5L_1+6)]. \tag{35}$$

Hence L_1 and L_2 give the lower and upper ends of the range of l values which satisfy the inequalities (32) and(33) for given values of α and β.

2) If $a_{l-} = 0$, Eqs. (26) and (27) give

$$a_{l+} = -[\alpha + l(l+1)\beta]/B \tag{36}$$

and

$$\mu_l = -[\alpha + l(l+1)\beta](2l+1)/l. \tag{37}$$

In this case both unitarity and $\mu_l \geqslant 0$ show that $[\alpha + l(l+1)\beta]$ must be negative. Again imposing the unitar-

ity condition $0 \leqslant a_{l+} \leqslant 1$ on Eq. (36) and the maximum condition $\mu_l \geqslant D$ on Eq. (37) we find

$$0 \leqslant -[\alpha + l(l+1)\beta] \leqslant \frac{4l}{2l+1} \tag{38}$$

$$-[\alpha + l(l+1)\beta] \geqslant \frac{2l(l+1)}{(2l+1)^2}. \tag{39}$$

Contributions of $a_{l-} = 0$ and a_{l+} as given by Eq. (36) to A_0 and S are:

$$A_0 = -\tfrac{1}{4}\alpha[(L_2+2)^2 - (L_1+1)^2] - \tfrac{1}{4}\alpha \sum_{L_1}^{L_2} \frac{1}{l}$$
$$-\tfrac{1}{24}\beta[(L_2+1)(L_2+2)(3L_2^2 + 7L_2 + 3)$$
$$-L_1(L_1+1)(3L_1^2 + L_1 - 1)], \tag{40}$$

$$S = -\tfrac{1}{24}\alpha[(L_2+1)(L_2+2)(3L_2^2 + 7L_2 + 3)$$
$$-L_1(L_1+1)(3L_1^2 + L_1 - 1)]$$
$$-\frac{\beta}{240}[L_2(L_2+1)(L_2+2)(2L_2+1)(2L_2+3)(5L_2+1)$$
$$-(L_1-1)L_1(L_1+1)(2L_1-1)(2L_1+1)(5L_1+6)]. \tag{41}$$

Class I^-B^+: In this class $\mu_l = 0, \lambda_l \leqslant 0$. The four Eqs. (8)–(11) become:

$$(B - \lambda_l)a_{l+} - Ba_{l-} + \tfrac{1}{2}[\alpha + l(l+1)\beta + \lambda_l] = 0, \tag{42}$$

$$Da_{l+} - Da_{l-} - \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0, \tag{43}$$

$$(B - \lambda_l)r_{l+} - Br_{l-} = 0, \tag{44}$$

$$Dr_{l+} - Dr_{l-} = 0. \tag{45}$$

Equation (45) gives $r_{l+} = r_{l-}$. Equation (44) gives $\lambda_l r_{l+} = 0$. For $r_{l+} \neq 0 \lambda_l = 0$. In this case Eq. (42) becomes

$$B(a_{l+} - a_{l-}) + \tfrac{1}{2}[\alpha + l(l+1)\beta] = 0.$$

But we saw before that this is not compatible with the Eq. (43). Hence we must have

$$r_{l+} = 0.$$

In this case also

$$r_{l-} = 0.$$

Because in this class $\mu_l = 0$ we have $u_l = a_{l+} - a_{l+}^2 = 0$. Hence

$$a_{l+} = \begin{matrix} \nearrow 1 \\ \searrow 0 \end{matrix}.$$

1) If $a_{l+} = 1$, Eqs. (42) and (43) give

$$a_{l-} = 1 - [\alpha + l(l+1)\beta]/2D \tag{46}$$

and

$$\lambda_l = [\alpha + l(l+1)\beta](2l+1)/(l+1). \tag{47}$$

Imposing the unitarity condition $0 \leqslant a_{l-} \leqslant 1$ on Eq. (46) and the maximum condition $\lambda_l \geqslant B$ on Eq. (47), we find

$$0 \leqslant a + l(l+1)\beta \leqslant 4(l+1)/(2l+1) \tag{48}$$

and

$$\alpha + l(l+1)\beta \geqslant 2l(l+1)/(2l+1)^2. \tag{49}$$

The contributions of $a_{l-} = 1$ and a_{l+} as given by Eq. (30) to A_0 and S can be explicitly calculated by using the well-known addition formulas for powers of integers except one term in A_0. The results are

$$A_0 = (L_2+1)^2 - L_1^2 - \tfrac{1}{4}\alpha[L_2^2 - (L_1-1)^2] - \tfrac{1}{4}\alpha \sum_{L_1}^{L_2} \frac{1}{l+1}$$
$$-\tfrac{1}{24}\beta[L_2(L_2+1)(3L_2^2 + 5L_2 + 1)$$
$$-(L_1-1)L_1(3L_1^2 - L_1 - 1)], \tag{50}$$

$$S = \tfrac{1}{2}[L_2(L_2+1)^2(L_2+2) - (L_1-1)L_1^2(L_1+1)]$$
$$-\tfrac{1}{24}\alpha[L_2(L_2+1)(3L_2^2 + 5L_2 + 1)$$
$$-(L_1-1)L_1(3L_1^2 - L_1 - 1)]$$
$$-\frac{\beta}{240}[L_2(L_2+1)(L_2+2)(2L_2+3)(2L_2+1)$$
$$\times(5L_2-1) - (L_1-1)L_1(L_1+1)(2L_1+1)$$
$$\times(2L_1-1)(5L_1-6)]. \tag{51}$$

Again L_1 and L_2 are determined for given α and β by inequalities (48) and (49).

2) If $a_{l+} = 0$ Eqs. (42) and (43) give

$$a_{l-} = -[\alpha + l(l+1)\beta]/2D \tag{52}$$

and

$$\lambda_l = -[\alpha + l(l+1)\beta](2l+1)/(l+1). \tag{53}$$

Again both unitarity and the positiveness of the inequality multipliers require that $[\alpha + l(l+1)\beta]$ be negative. As before we impose the unitarity condition $0 \leqslant a_{l-} \leqslant 1$ on Eq. (52) and the maximum condition $\lambda_l \geqslant B$ on Eq. (53) and find:

$$0 \leqslant -[\alpha + l(l+1)\beta] \leqslant 4(l+1)/(2l+1) \tag{54}$$

and

$$-[\alpha + l(l+1)\beta] \geqslant 2l(l+1)/(2l+1)^2. \tag{55}$$

Contributions of $a_{l+} = 0$ and a_{l-} as given by Eq. (52) to A_0 and S are:

$$A_0 = -\tfrac{1}{4}\alpha[L_2^2 + (L_1-1)^2] - \tfrac{1}{4}\alpha \sum_{L_1}^{L_2} \frac{1}{l+1}$$
$$-\tfrac{1}{24}\beta[L_2(L_2+1)(3L_2^2 + 5L_2 + 1)$$
$$-(L_1-1)L_1(3L_1^2 - L_1 - 1)], \tag{56}$$

$$S = \tfrac{1}{24}\alpha[L_2(L_2+1)(3L_2^2 + 5L_2 + 1)$$
$$-(L_1-1)L_1(3L_1^2 - L_1 - 1)]$$
$$-\frac{\beta}{240}[L_2(L_2+1)(L_2+2)(2L_2+3)(2L_2+1)$$
$$\times(5L_2-1)] - (L_1-1)L_1(L_1+1)$$
$$\times(2L_1+1)(2L_1-1)(5L_1-6)]. \tag{57}$$

Class B^+B^-: In this class neither λ_l nor μ_l is zero. We thus have to solve the four Eqs. (8)–(11) for a_{l+}, a_{l-}, r_{l+} and r_{l-}. We note that the determinant of the Eqs. (8) and (9) is equal to determinant of the homogeneous equations (10) and (11). Thus when the determinant does not vanish Eqs. (10) and (11) have trivial solutions

$$r_{l+} = 0 \text{ and } r_{l-} = 0.$$

This leads to

$$a_{l+} = \begin{matrix} \nearrow 1 \\ \searrow 0 \end{matrix} \text{ and } a_{l-} = \begin{matrix} \nearrow 1 \\ \searrow 0 \end{matrix}.$$

We now consider these four possibilities:

1) When $a_{l+} = a_{l-} = 0$,

$$\alpha + l(l+1)\beta + \lambda_l = 0, \quad \alpha + l(l+1)\beta + \mu_l = 0.$$

Hence

$$\lambda_l = -[\alpha + l(l+1)\beta] \geqslant 0,$$
$$\mu_l = -[\alpha + l(l+1)\beta] \geqslant 0. \tag{58}$$

This case does not contribute to G, A_0 or S.

2) When $a_{l+}=0$, $a_{l-}=1$, Eqs. (8) and (9) give

$$\lambda_l = 2B - [\alpha + l(l+1)\beta] \geqslant 0 \tag{59}$$

and

$$\mu_l = 2D + [\alpha + l(l+1)\beta] \geqslant 0. \tag{60}$$

We shall look later at the inequalities which determine the range of l for given α and β. Here we want to mention that $y=\alpha+l(l+1)\beta$ represents a parabola in the variable l with its extremum at $l=-\frac{1}{2}$. Whether this extremum is a maximum (downward looking parabola) or minimum (upward looking parabola) depends on the sign of β. In the first case (maximum) β is negative, in the second case (minimum) β is positive.

One can see now from the inequality (59) that only if $\beta>0$ can the range of l be limited. Together with (60) this will determine a range for l. On the other hand inequality (60) shows that the range of l given by it can only be finite if $\beta<0$ and together with (59) this will give the range of l. Since both inequalities must be satisfied simultaneously it is sufficient that the range of l is restricted by one of them only.

Inequalities of the type (59) and (60) are best studied by plotting say here $2B=4l/(2l+1)$, $2D=4(l+1)/(2l+1)$ and $y=\alpha+l(l+1)\beta$ against the variable l.

3) When $a_{l+}=1$, $a_{l-}=0$, Eqs. (8) and (9) give

$$\lambda_l = 2B + [\alpha + l(l+1)\beta] \geqslant 0, \tag{61}$$

$$\mu_l = 2D - [\alpha + l(l+1)\beta] \geqslant 0. \tag{62}$$

4) Finally when $a_{l+}=1$, $a_{l-}=1$, Eqs. (8) and (9) give

$$\lambda_l = \mu_l = \alpha + l(l+1)\beta \geqslant 0. \tag{63}$$

It is to be noted that this case does not contribute to G even though it contributes to A_0 and S. Thus one would not want to use this set to maximize G which in turn indicates that $\alpha+l(l+1)\beta$ should be taken negative. When the determinant of the Eqs. (10) and (11) vanishes we have another possible solution. This is obtained when the other determinants of the inhomogeneous set (8) and (9) also vanish. In this case we have

$$\frac{B-\lambda_l}{D} = \frac{B}{D-\mu_l} = -\frac{\alpha+l(l+1)\beta+\lambda_l}{\alpha+l(l+1)\beta+\mu_l}. \tag{64}$$

One solution is $\alpha+l(l+1)\beta=0$ which leads to $\alpha=\beta=0$, that is, no constraints at all.

The other solution is

$$\lambda_l = \mu_l = 2. \tag{65}$$

In this case Eqs. (8) and (9) become identical. Also Eqs. (10) and (11) become equal. We thus have, together with the unitarity equations the following set to solve:

$$Da_{l+} + Ba_{l-} - \tfrac{1}{2}[\alpha + l(l+1)\beta + 2] = 0, \tag{66}$$

$$Dr_{l+} + Br_{l-} = 0, \tag{67}$$

$$a_{l+} - a_{l+}^2 - r_{l+}^2 = 0, \tag{68}$$

$$a_{l-} - a_{l-}^2 - r_{l-}^2 = 0. \tag{69}$$

The solutions of these equations are:

$$a_{l+} = \frac{1}{4}\frac{\alpha+l(l+1)\beta+2}{[\alpha+l(l+1)\beta]D}[\alpha+l(l+1)\beta+D-B], \tag{70}$$

$$a_{l-} = \frac{1}{4}\frac{\alpha+l(l+1)\beta+2}{[\alpha+l(l+1)\beta]B}[\alpha+l(l+1)\beta+B-D], \tag{71}$$

$$r_{l+}^2 = \frac{1}{16}\frac{\alpha+l(l+1)\beta+2}{[\alpha+l(l+1)\beta]^2 D^2}[\alpha+l(l+1)\beta+D-B] \times [B^2-(\alpha+l(l+1)\beta-D)^2], \tag{72}$$

$$r_{l-}^2 = \frac{1}{16}\frac{\alpha+l(l+1)\beta+2}{[\alpha+l(l+1)\beta]^2 B^2}[\alpha+l(l+1)\beta+B-D] \times [D^2-(\alpha+l(l+1)\beta-B)^2], \tag{73}$$

The contributions these amplitudes to A_0 and S will be

$$A_0 = \frac{1}{4}\sum_{L_1}^{L_2}[\alpha+l(l+1)\beta+2](2l+1) = \frac{\alpha+2}{4}[(L_2+1)^2-L_1^2] + \frac{\beta}{8}[L_2(L_2+1)^2(L_2+2)-(L_1-1)L_1^2(L_1+1)], \tag{74}$$

$$S = \frac{1}{4}\sum_{L_1}^{L_2}[\alpha+l(l+1)\beta+2]\,l(l+1)(2l+1) = \frac{\alpha+2}{8}[L_2(L_2+1)^2(L_2+2)-(L_1-1)L_1(L_1+1)] + \frac{\beta}{12}[L_2^2(L_2+1)^2(L_2+2)^2-(L_1-1)^2L_1^2(L_1+1)^2]. \tag{75}$$

In their contribution to G there is an ambiguity in the sign of the square roots of the real parts. But in order to maximize G one has to chose the signs of r_{l+} and r_{l-} opposite. With this, the contribution to G becomes:

$$G = \frac{1}{8}\sum_{L_1}^{L_2}\{4-[\alpha+l(l+1)]^2\}(2l+1) = \frac{4-\alpha^2}{8}[(L_2+1)^2-L_1^2] - \frac{\alpha\beta}{8} \times [L_2(L_1+1)^2(L_2+2)-(L_1-1)L_1(L_1+1)] - \frac{\beta^2}{24}[L_2^2(L_2+1)^2(L_2+2)^2-(L_1-1)^2L_1^2(L_1+1)^2]. \tag{76}$$

For this case the maximum conditions are automatically satisfied since $\lambda_l=\mu_l=2$. Unitarity imposed on Eqs. (70) and (71) in the form

$$0 \leqslant a_{l+}, a_{l-} \leqslant 1$$

gives two possible domains which are common to both amplitudes,

$$2/(2l+1) \leqslant \alpha + l(l+1)\beta \leqslant 2 \tag{77}$$

or

$$-2 \leqslant \alpha + l(l+1)\beta \leqslant -2/(2l+1). \tag{78}$$

This completes the analysis of the four classes. We obtained the forms of the partial waves in all classes. Also the unitarity and the maximum conditions gave relations in form of inequalities between α, β and l.

Now the bound on G is found in principle as follows: α and β determine the lower and upper limits L_1, L_2 of l. Thus L_1 and L_2 are functions of α, β. When we express A_0 and S in terms of all contributing amplitudes, they become functions

of α and β only. Hence α and β can be found in terms of A_0 and S. Once α and β are found, the limits L_1, L_2 for all contributing classes can be found and G calculated.

In order not to clutter expressions we used the same symbols L_1, L_2 for each class. But is in understood that they will in general be different for each class. Also A_0 and S were used for each classe's contribution. To find the total A_0 and S those have to be added.

The signs of α and β are related to the rate of change of G with A_0 and S. Thus (see for example Ref. 3)

$$\partial G/\partial A_0 = -\alpha$$

and

$$\partial G/\partial S = -\beta.$$

Thus if physically one expects G to grow with A_0, α should be negative. Similarly if G increases with incrasing dA/dt, β should be negative.

When α and β are negative (77) is not valid. Thus (78) gives the upper and lower limits of l. They are obtained by equating $\alpha + l(l+1)\beta$ to -2 and $-2/(2l+1)$. In the first case the equation is quadratic and one choses the positive root. In the second case the equation is cubic with only one real root which is positive. All l's between these two values will contribute to A_0, S and G in the forms (74), (75), and (76).

Going backwards we next proceed to case 4) ot B^+B^-. It is obvious that with α and β negative (63) will not be satisfied and this case will not contribute.

In case 3) solving the equality (61) will give the lower and upper values of l for this set. Then (62) is automatically satisfied.

In case 2) solving the equality (60) we find the range of l for this set. Then (59) is automatically satisified.

In class I^-B^+ for the amplitudes of type (52) solving (54) and (55) determines L_1 and L_2 of this case.

When the amplitude has the form (46), inequality (49) can not be satisfied with α and β negative. Thus this set does not contribute.

In class I^+B^- for the amplitudes of type (36) solving (38) and (39) with equalities determines L_1 and L_2 of this case.

When the amplitude has the form (30), inequality (33) can not be satisfied with α and β negative.

Thus we have covered all cases which will contribute to A_0, S and G when α and β are negative. Other cases can be studies in a similar way.

IV. DISCUSSION AND CONCLUSION

We extended the method of variational calculus with inequality constraints to a spin case. Even though the paper deals only with spin 0–spin $\frac{1}{2}$ particle scattering the technique is general and can be applied to arbitrary spins. The basic idea is the definition of classes such that a set of partial wave amplitudes with the same l value belongs to one and only one of these classes. The values of the inequality multipliers in those classes together with unitarity and the maximum condition obtained from second variation determine the form of the partial waves. They also impose conditions on Lagrange multipliers.

Since α and β are global multipliers, not depending on l, one has to find compatible solutions of these conditions, which appear in form of inequalities for given α and β. These solutions determine the range of values which contribute to A_0, S, and G for the type of partial waves in a particular class or subset of a class. A_0 and S are expressed in terms of α, β and $L_1(\alpha,\beta)$ and $L_2(\alpha,\beta)$ where these last are the lower and upper limits of the contributing l values for each set. Solving for α and β in terms of A_0 and S gives the values of the Lagrange multipliers which in turn determine L_1 and L_2 for each type of amplitude. Finally, α, β and L_1's and L_2's determine G.

The results of this paper can be applied to cases like ΠN or KN scattering. However the determination of the multipliers has to be made numerically, since some of the equations which determine the limits of l are cubic and quartic [e.g. (55)].

ACKNOWLEDGMENTS

I would like to thank Université de Nice, Laboratoire de Physique Theéorique and its director Professor François Rocca for their hospitality and support during my stay at the University of Nice where this work was done.

In particular I am grateful to Professor Michel Le Bellac for making it possible for me to spend my sabbatical leave at his institute.

Finally I am thankful to all the members of the Laboratory for their hospitality and to M. Le Bellac, J. L. Femenias, J. P. Provost and G. Vallée for many discussions.

[1] S. W. Mac Dowell and A. Martin, Phys. Rev. B **135** 960 (1964).
[2] S. M. Roy and V. Singh, Phys. Rev. Lett. **24**, 28 (1970); and Phys. Rev. D **1**, 2638 (1970).
[3] M. B. Einhorn and R. Blankenbecler, Ann. Phys. **67**, 480 (1971).
[4] R. Savit, R. Blankenbecler and M. B. Einhorn, J. Math. Phys. **12**, 2092 (1971).
[5] K. H. Mütter, Nucl. Phys. B **31**, 589 (1971).
[6] I. Goldhirsch and M. Kugler, Nucl. Phys. B **104**, 499 (1976).
[7] G. Mahoux, Phys. Lett. **65** B, 139 (1976).
[8] G. Auberson, A. Martin and C. Menessier, Phys. Lett. **67** B, 75 (1977).
[9] G. Auberson and S. M. Roy, Lett. Nuovo Cimento **19**, 97 (1977).
[10] A. D. Gangal and A. S. Vengurlekar, Phys. Rev. D **18**, 682 (1978).
[11] A. D. Gangal and A. S. Vengurlekar, Phys. Rev. D **19**, 1465 (1979).
[12] G. Menessier, S. M. Roy and V. Singh, Nuovo Cimento A **50**, 443 (1974).
[13] S. M. Roy, Phys. Rev. Lett. **43**, 19 (1979).

Confrontation of macroscopic and microscopic nuclear collective models

E. Chacón and M. Moshinsky[a)]
Instituto de Física, UNAM, Apdo. Postal 20-364, México 20, D. F., México

V. Vanagas
Institute of Physics, Lithuanian Academy of Sciences, Vilnius, Lithuania, USSR

(Received 26 June 1980; accepted for publication 5 September 1980)

Since its start nuclear theory has lived with the dichotomy of viewing the nucleus microscopically, as a system of nucleons, or describing it macroscopically in terms of collective coordinates. In the last decade though, a point transformation has been introduced in which single particle coordinates can be expressed in terms of collective ones plus others, opening the possibility of deriving a microscopic collective model. In the present paper we confront the macroscopic and microscopic collective models, first in a space of two dimensions, in which we find explicitly the unitary representation in quantum mechanics of the canonical transformation that relates them. We then show how to extend every step of the analysis to the three-dimensional problem, though there some of the states required are not yet available in analytic form. One of the fundamental problems in collective models of the nucleus is that of shape. We indicate what are the operators whose expectation values give a reasonable description of the shape in the macroscopic and microscopic collective models and confront them critically.

PACS numbers: 21.60.Ev

I. INTRODUCTION

Almost from the initial steps of nuclear theory, after the discovery of the neutron in 1932, there has been a dichotomy in its attitude toward its subject of study. On the one hand, the nucleus has been viewed microscopically as a system of nucleons, first through the shell model[1,2,3] and later through a more realistic Hamiltonian in which shell model states provided the initial approximations.[4] On the other, a collective view prevailed, first in the liquid drop model of Niels Bohr[5] and later in its extension in the work of Bohr and Mottelson,[6] continuing up to the present time in such work as that on transitional nuclei by the Frankfurt group[7] or—in a conceptually different, but mathematical similar, formulation[8]–the interacting boson approximation.[9]

One of the authors of the present paper (M. M.) has recently confronted[8] different collective models, such as those mentioned at the end of the previous paragraph, to show the similarity of their group theoretical background. Another (V. V.) has discussed[10] the group theory underlying the Hamiltonian of A particles, in an appropriate system of coordinates,[11] that brings out their collective behavior. Thus it seemed to the authors that the time is ripe for confronting the collective models derived from macroscopic and microscopic views of the nucleus.

At the start it became clear that a full understanding of the conceptual structure of the problems was required, before getting involved in their mathematical complexities. Thus the main part of the present paper is concerned with the situation in an hypothetical two-rather than the real three-dimensional space. We shall discuss in Sec. II this two-dimensional case of the Bohr–Mottelson collective model and its different generalizations, with particular emphasis on the problem of shape of the nuclei. Then in Sec. III we turn our attention to the microscopic picture of A particles interacting through harmonic oscillator forces, to derive from it the Hamiltonian for the collective part. Once this is available we discuss its solutions, symmetries and again the problem of shape. In Sec. IV, we then confront the macroscopic and microscopic collective Hamiltonian of the previous sections and derive the canonical transformation that relates them.

In the final Sec. V, we then turn our attention to the three-dimensional case and show how step by step we can implement the program developed fully for two dimensions. However, there are analytic expressions for states in two dimensions, whose counterpart we have not yet determined explicitly in three dimensions, but we plan to do this in later publications. The outline in Sec. V leads from a system of A particles interacting through harmonic oscillator potentials, to the collective part of this Hamiltonian using appropriate coordinate transformations, then through a canonical transformation to the oscillator Hamiltonian of the s-d interacting boson model, of which the Bohr–Mottelson oscillator collective Hamiltonian is a particular case. Through all of this analysis, we take particular interest in the shape of the states, which is a very central problem in the collective model.

II. THE MACROSCOPIC COLLECTIVE MODELS

In this section we start from a two-dimensional Bohr–Mottelson model[12] and then generalize it to situations in which we can deal both with the vibrational and rotational limits, as well as with states in the transitional region between the two.

A. The Bohr–Mottelson (BM) model

In two dimensions in which the polar coordinates will be denoted by (ρ, χ), the "liquid drop" instead of being bound by a surface[13] will be bound by a line[12]

[a)]Member of the Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

 022-2488/81/030605-18$1.00

$$\rho = \rho_0[1 + \alpha_x \cos 2\chi + \alpha_y \sin 2\chi\,] = \rho_0[1 + (1/\sqrt{2})(\alpha^+ e^{i2\chi} - \alpha^- e^{-i2\chi})], \tag{2.1}$$

where we limit ourselves to quadrupole vibrations characterized by the two coordinates α_x, α_y or equivalently

$$\alpha_{\pm} = (\mp)(1/\sqrt{2})(\alpha_x \pm i\alpha_y); \alpha^{\pm} = -\alpha_{\mp}, \tag{2.2}$$

corresponding to the five[6,13] α_m of the three-dimensional problem. For small vibrations we get for the Hamiltonian the usual two-dimensional oscillator in α_x, α_y.[12]

Passing to the coordinate system fixed in the body we can write

$$\alpha_{\pm} = \mp(1/\sqrt{2})\beta \exp(\pm i2\vartheta), \tag{2.3}$$

where ϑ, β are, respectively, the sole Euler angle and deformation parameter. In units[12,13] in which $\hbar = B_2 = C_2 = 1$, the Bohr–Mottelson vibrational Hamiltonian becomes

$$H_0 = \tfrac{1}{2}\left(-\frac{1}{\beta}\frac{\partial}{\partial\beta}\beta\frac{\partial}{\partial\beta} - \frac{1}{4\beta^2}\frac{\partial^2}{\partial\vartheta^2} + \beta^2 \right), \tag{2.4}$$

and its eigenstates are given by[14]

$$\psi_{nm}(\beta, \vartheta) = f_n^{|m|}(\beta)(2\pi)^{-\frac{1}{2}} \exp(i2m\vartheta), \tag{2.5a}$$

with[14]

$$f_n^{|m|}(\beta) = [2(n!)/\Gamma(n + |m| + 1)]^{\frac{1}{2}} \times e^{-\frac{\beta^2}{2}} \beta^{|m|} L_n^{|m|}(\beta^2), \tag{2.5b}$$

and $L_n^{|m|}$ being an associated Laguerre polynomial.[15] The corresponding eigenvalue E_{nm} is

$$E_{nm} = 2n + |m| + 1. \tag{2.6}$$

Note that $\exp(i2m\vartheta)$ appears in (2.5a) with m integer as the eigenstates must be invariant under the transfromation[12] $\vartheta \to \vartheta + \pi$ which leaves unchanged the defining equation (2.3).

In the frame of reference fixed in the body the line delimiting the boundary of the "liquid drop" becomes

$$\rho = \rho_0[1 + \beta\cos 2\chi'], \quad \chi' = \chi - \vartheta. \tag{2.7}$$

The two principal axis whose length we denote by ρ_1', ρ_2' correspond then to $\chi' = 0, \pi/2$ i.e.,

$$\rho_1' = \rho_0(1 + \beta), \quad \rho_2' = \rho_0(1 - \beta), \tag{2.8}$$

and a good measure of the deformation is given by

$$\left(\frac{\rho_1' - \rho_2'}{\rho_1' + \rho_2'}\right)^2 = \frac{(\rho_1'^2 - \rho_2'^2)^2}{16\rho_0^4} = \beta^2. \tag{2.9}$$

The prime notation is used for ρ_1', ρ_2' to distinguish them from ρ_1, ρ_2 that appear later for the microscopic problem. Thus the deformation of the vibrational states that are eigenfunctions of the Hamiltonian (2.4) can be estimated from the expectation values[14]

$$\int_0^{\infty}\int_0^{2\pi} \psi_{nm}^*(\beta, \vartheta)\beta^2\psi_{nm}(\beta, \vartheta)\beta\, d\beta\, d\vartheta = 2n + |m| + 1. \tag{2.10}$$

Note that in the units we are using the expectation value of β^2 must be multiplied by some dimensionless function of $\hbar$, B_2, C_2 and thus (2.10) allows us only to compare the deformations of different states rather than give us an absolute measure of them. This implies that we could equally well have used as a measure of the deformation $16\rho_0^4\beta^2 = (\rho_1'^2 - \rho_2'^2)^2$, where ρ_0 is the fixed radius of the nondeformed nucleus. As we shall see later $(\rho_1'^2 - \rho_2'^2)^2$ allows a more direct confrontation with results of the microscopic collective model.

It is not possible to compare (2.10) with the deformation in rotational states as they can only be introduced in a Bohr–Mottelson model in an *ad hoc* fashion.[13] However, we shall see in the next subsections that the BM Hamiltonian can be generalized in several ways so as to include the latter. Thus, we will be able to compare the deformation of vibrational, transitional, and rotational states.

B. Higher-degree terms in the collective Hamiltonian

The BM Hamiltonian (2.4) comes from considering small vibrations of the liquid drop. Had we considered higher-order terms, we would have Hamiltonians $H(\alpha_{\pm}, \pi_{\pm})$, (where $\pi_{\pm} = -i\partial/\partial\alpha^{\pm}$) subject only to the restriction that H is invariant under rotations and reflections, the latter in space and time. If the higher-order terms are static,[7] i.e. dependent only on $\alpha_{\pm}$, the invariance under rotation implies that they give rise to a potential dependent only on $\beta^2 = -2\alpha_+\alpha_-$. Thus the most general higher-order static Hamiltonian can be written as

$$H = H_0 + V(\beta^2), \tag{2.11}$$

where $V(\beta^2)$ is some function of the argument. In the three-dimensional case,[7] where V is function of both β^2 and $\beta^3\cos 3\gamma$, its form is determined from energy levels and transition probabilities of the nucleus being studied.[7] In the hypothetical two-dimensional case we shall select

$$V(\beta^2) = \mathscr{I}^2/\beta^2, \tag{2.12}$$

with $\mathscr{I}$ being an arbitrary constant and show that by varying $\mathscr{I}$ we can reach the vibrational and rotational limits as well as the transitional situation in between. For H of (2.11) and (2.12) the eigenstates can still be written in the form (2.5), if we retain m in the angular part but replace it in the radial part by[14]

$$(\mathscr{I}^2 + m^2)^{\frac{1}{2}}. \tag{2.13}$$

The corresponding energy levels are then given[14] by

$$E_{nm} = 2n + (\mathscr{I}^2 + m^2)^{\frac{1}{2}} + 1 \simeq 2n + \mathscr{I} + 1 + (m^2/2\mathscr{I}), \tag{2.14}$$

where the right-hand side holds only if $\mathscr{I} \gg m$. In this case for each value of n we have a rotational band as in the two-dimensional case m is the quantum number associated with the angular momentum. Thus the rotational limit is achieved when $\mathscr{I} \gg m$ while the vibrational one is reached when $\mathscr{I} \to 0$ and we also have the transitional situation in between.

The deformation of the eigenstates of H is again measured by the expectation value of β^2 with respect to them. But, as shown in Ref. 14, this expectation value remains equal to that of the Hamiltonian and therefore given by (2.14). Thus we see that for the lowest state $n = m = 0$ in the vibrational limit, when $\mathscr{I} = 0$, the expectation value of β^2 is 1, while in the rotational limit the expectation value of β^2 for the corresponding state is $\mathscr{I} + 1 \gg 1$. Thus the deformation

of the lowest states increases considerably when we pass from a vibrational to a rotational limit.

We proceed now to discuss the group theory behind the two-dimensional BM model and show that it can be generalized in a way that allows for a different procedure for including both vibrational and rotational Hamiltonians. This generalization, which is the two-dimensional counterpart of the interacting boson approximation, will be the one that eventually we connect with the microscopic collective model.

C. Group theory and the σ-δ boson collective Hamiltonian

Associated with the coordinates $\alpha_{\pm}$ in the BM model, we can introduce creation and annihilation operators

$$\eta_{\pm} = (1/\sqrt{2})(\alpha_{\pm} - \partial/\partial\alpha^{\pm}),$$
$$\xi^{\pm} = (1/\sqrt{2})(\alpha^{\pm} + \partial/\partial\alpha_{\pm}), \tag{2.15}$$

which we could call δ bosons as they correspond to the values ± 2 of the angular momentum in this two-dimensional space. From these bosons we can construct the generators of an SU(2) algebra

$$T_1 = (-1/\sqrt{2})\eta_+\xi^-, \quad T_0 = \tfrac{1}{2}(\eta_+\xi^+ - \eta_-\xi^-),$$
$$T_{-1} = (1/\sqrt{2})\eta_-\xi^+, \tag{2.16}$$

with the properties

$$[T_0, T_{\pm 1}] = \pm T_{\pm 1}, \quad [T_1, T_{-1}] = -T_0. \tag{2.17}$$

Together with the number operator for δ-bosons,

$$n_\delta = \eta_+\xi^+ + \eta_-\xi^- = H_0 - 1, \tag{2.18}$$

the T_τ, $\tau = 1, 0, -1$ constitute the generators of a U(2) group which is obviously the symmetry group of the two-dimensional BM Hamiltonian as $[T_\tau, n_\delta] = 0$.

We can now attempt to include rotational states, not through special dependence on β^2 as in the previous subsections, but by adding to the δ-bosons a σ boson associated with a scalar coordinate we denote by $\bar{\alpha}$. This would be a procedure similar to the one followed[16] in extending the three-dimensional BM oscillator model to the *s-d* interacting boson approximation. The creation and annihilation operators associated with the σ boson then have the form

$$\bar{\eta} = (1/\sqrt{2})(\bar{\alpha} - \partial/\partial\bar{\alpha}), \quad \bar{\xi} = (1/\sqrt{2})(\bar{\alpha} + \partial/\partial\bar{\alpha}), \tag{2.19}$$

and with their help we can extend the U(2) group (2.16)–(2.18) of the BM model to the U(3) group whose nine generators are

$$T_{\pm 1}, \quad T_0, \quad n_\delta, \quad \eta_{\pm}\bar{\xi}, \ \bar{\eta}\xi^{\pm}, \ \bar{\eta}\bar{\xi}. \tag{2.20}$$

This group will be the equivalent of U(6) in the interacting boson model.

The U(3) of (2.20) has not only as a subgroup SU(2) of (2.16), which characterizes the BM vibrational limit, but also another subgroup SU'(2)$\simeq$O(3) whose generators

$$Q_1 = -(\eta_+\bar{\xi} + \bar{\eta}\xi^-), \quad Q_{-1} = (\eta_-\bar{\xi} + \bar{\eta}\xi^+),$$
$$\mathscr{L}_0 = (\eta_+\xi^+ - \eta_-\xi^-) = 2T_0, \tag{2.21}$$

satisfy the commutation relations

$$[\mathscr{L}_0, Q_{\pm 1}] = \pm Q_{\pm 1}, \quad [Q_1, Q_{-1}] = -\mathscr{L}_0. \tag{2.22}$$

The eigenstates of the number operator

$$\hat{N} = \eta_+\xi^+ + \eta_-\xi^- + \bar{\eta}\bar{\xi} \tag{2.23}$$

of the U(3) group can now be characterized by different chains of subgroups. If we consider U(3)$\supset$SU(2)$\supset O(2)$ of (2.16), the states will be labeled by the eigenvalues of the operators

$$\hat{N}, \quad T^2 = -(T_1T_{-1} + T_{-1}T_1) + T_0^2$$
$$= (n_\delta/2)[(n_\delta/2) + 1] \text{ and } 2T_0, \tag{2.24}$$

where the last one is the angular momentum in two-dimensional space. Denoting the eigenvalues, respectively, by N (total number of bosons), ν (number of δ bosons) and m (angular momentum), the eigenket clearly corresponds to a state of the three-dimensional oscillator in cylindrical coordinates $(\beta, 2\vartheta, \bar{\alpha})$

$$|N\nu m\rangle = \psi_{nm}(\beta, \vartheta)\phi_{N-\nu}(\bar{\alpha}), \quad 2n + |m| = \nu, \tag{2.25}$$

where ψ_{nm} is given by (2.5) and $\phi_{N-\nu}(\bar{\alpha})$ is a one-dimensional oscillator function of $N - \nu$ quanta. For these states the expectation value of β^2 continues to be that of (2.10) i.e., $\nu + 1$, which is also the energy of the state if the Hamiltonian is taken as n_δ of the vibrational limit.

If on the other hand we consider the chain of groups U(3)$\supset$SU'(2)$\simeq$O(3)$\supset$O(2) of (2.21), the states will be determined by the eigenvalues of the operators

$$\hat{N}, \quad \mathscr{L}^2 = -(Q_1Q_{-1} + Q_{-1}Q_1) + \mathscr{L}_0^2, \quad \mathscr{L}_0 = 2T_0, \tag{2.26}$$

which we could denote, respectively, by $N, \lambda(\lambda + 1), m$ with $\lambda = N, N - 2, \ldots 1$ or 0 and $m = \lambda, \lambda - 1, \ldots, -\lambda$. If we introduce the coordinates

$$R^2 = \bar{\alpha}^2 + \beta^2, \quad \tan\Theta = (\beta/\bar{\alpha}), \quad \Phi = 2\vartheta, \tag{2.27}$$

we immediately see that $\mathscr{L}^2$ appearing in (2.26) is the square of the vector $\mathscr{L} = -i(\mathbf{R}\times\nabla)$ and thus the eigenstates will be those of the three-dimensional oscillator in spherical coordinates i.e.,

$$|N\lambda m\rangle = F_N^\lambda(R)Y_{\lambda m}(\Theta, \Phi), \tag{2.28a}$$

where $Y_{\lambda m}$ is a spherical harmonic and[14]

$$F_N^\lambda(R) = \{2[(N-\lambda)/2]!/\Gamma[(N+\lambda+3)/2]\}^{\frac{1}{2}}$$
$$\times e^{-R^2/2}R^\lambda L^{\lambda+\frac{1}{2}}_{(N-\lambda)/2}(R^2), \tag{2.28b}$$

with L being a Laguerre polynomial.

The kets (2.28) are eigenstates of the operator

$$Q^2 = Q_1Q_{-1} + Q_{-1}Q_1 = \mathscr{L}_0^2 - \mathscr{L}^2, \tag{2.29}$$

with eigenvalues

$$m^2 - \lambda(\lambda + 1), \tag{2.30}$$

and thus for each value of λ we have a rotational band whose energy levels are associated with the permissible values of the square of the angular momentum m^2. The operator β^2 characterizing the deformation is

$$\beta^2 = R^2\sin^2\Theta, \tag{2.31}$$

and thus its expectation value with respect to the states (2.28) is given by

$$\langle N\lambda m|\beta^2|N\lambda m\rangle$$

$$= (N+\tfrac{3}{2})\left\{\tfrac{2}{3}\left[1+\frac{3m^2-\lambda(\lambda+1)}{(2\lambda-1)(2\lambda+3)}\right]\right\}. \tag{2.32}$$

From (2.30) we see that the lowest state corresponds to $m=0, \lambda=N$ and thus, if $N \gg 1$, the expectation value of the deformation for it is $\frac{1}{2}(N+\frac{3}{2})$, much larger than the value 1 we get for the lowest state $\nu=0$ in the vibrational limit. In fact this strong deformation holds for all states (2.28) as the curly bracket in (2.32) is always of the order of 1. We note furthermore that for large N, λ the curly bracket in (2.32) goes from $\frac{1}{2}$ to 1 when $|m|$ goes from 0 to λ, which indicates that there is a stretching effect on the shape as the angular momentum increases.

So far we have shown how to get the vibrational and rotational limits in the σ-δ boson picture. A transitional situation will obviously appear if we take as Hamiltonian

$$H=(1-x)n_\delta + xQ^2, \quad 0 \leqslant x \leqslant 1. \tag{2.33}$$

To solve it we consider the complete set of states $|N\lambda m\rangle$ which are eigenstates of $\hat{N}$, $\mathscr{L}_0$ that commute with the Hamiltonian H. The Q^2 is diagonal in this basis as indicated in (2.29) and thus we need only the matrix elements

$$\begin{aligned}\langle N\lambda' m|n_\delta|N\lambda m\rangle = &-\left[\frac{(N-\lambda)(N+\lambda+3)(\lambda-m+2)(\lambda-m+1)(\lambda+m+2)(\lambda+m+1)}{(2\lambda+1)(2\lambda+3)^2(2\lambda+5)}\right]^{\frac{1}{2}}\delta_{\lambda',\lambda+2}\\ &+\left\{\tfrac{2}{3}N+(\tfrac{2}{3}N+1)\frac{[3m^2-\lambda(\lambda+1)]}{(2\lambda-1)(2\lambda+3)}\right\}\delta_{\lambda'\lambda}\\ &-\left[\frac{(N-\lambda+2)(N+\lambda+1)(\lambda-m)(\lambda-m-1)(\lambda+m)(\lambda+m-1)}{(2\lambda-3)(2\lambda-1)^2(2\lambda+1)}\right]^{\frac{1}{2}}\delta_{\lambda',\lambda-2},\end{aligned} \tag{2.34}$$

which follows immediately from the facts that $n_\delta = \hat{N} - \frac{1}{2}(p_z^2+z^2)+\frac{1}{2}$, and p_z^2 has the same matrix elements[8] with respect to the states (2.28) as z^2 and the latter can be written as $z^2 = R^2\cos^2\Theta = \frac{1}{3}R^2[(16\pi/5)^{1/2}Y_{20}(\Theta,\Phi)+1]$.

The eigenstates i of H corresponding to definite N, m can be expressed as a linear combination of $|N\lambda m\rangle$

$$|iNm\rangle = \sum_\lambda a_{i\lambda}(Nm)|N\lambda m\rangle, \tag{2.35}$$

where the coefficient $a_{i\lambda}$ are obtained by the diagonalization of the finite matrix $\|\langle N\lambda' m|H|N\lambda m\rangle\|$. The deformation of the states $|iNm\rangle$ is then given by the expectation value

$$\begin{aligned}\langle iNm|\beta^2|iNm\rangle = \sum_{\lambda,\lambda'}\{&a^*_{i\lambda'}(Nm)a_{i\lambda}(Nm)\\ &\times[-\langle N\lambda' m|R^2\cos^2\Theta|N\lambda m\rangle]\\ &+(N+\tfrac{3}{2})\delta_{\lambda'\lambda}\},\end{aligned} \tag{2.36}$$

where, from the above discussion, we have that

$$\begin{aligned}\langle N\lambda' m|R^2\cos^2\Theta|N\lambda m\rangle &= (N+\tfrac{1}{2})\delta_{\lambda'\lambda}\\ &-\langle N\lambda' m|n_\delta|N\lambda m\rangle,\end{aligned} \tag{2.37}$$

with the last right-hand term given by (2.34).

We have extended the Bohr–Mottelson macroscopic collective model in two different ways, one through a potential energy depending on β^2 in a more complex way than an oscillator and the other through a σ-δ interacting boson model. In both cases we can describe the transitional situation, as well as the vibrational and rotational limits, and we have also been able to discuss the shape of nuclei in all of these cases. In the next section we turn to the microscopic problem of A nucleons interacting through harmonic oscillator forces, to derive from it a collective Hamiltonian which we later compare with the macroscopic one.

III. THE MICROSCOPIC COLLECTIVE MODEL

The macroscopic BM collective model discussed in the previous section starts from the oscillator Hamiltonian (2.4) associated with small vibrations of a liquid drop. In the microscopic case we shall also start with an oscillator Hamiltonian, but now of A particles interacting through harmonic oscillator forces.

A. Many-body oscillator Hamiltonian

We shall consider a system of A particles of mass M in a two-dimensional space, interacting through oscillator forces of frequency $\omega A^{-\frac{1}{2}}$. We designate by $\mathscr{H}_0$ the Hamiltonian just defined from which the kinetic energy of the center of mass has been subtracted. Thus we have

$$\begin{aligned}\mathscr{H}_0 = (2M)^{-1}&\sum_{s=1}^{A}\sum_{i=1}^{2}(p_i'^s)^2 + (M/2A)\omega^2\\ &\times\sum_{s>t=1}^{A-1}\sum_{i=1}^{2}(x_i'^s - x_i'^t)^2 - \tfrac{1}{2}\hbar\omega\sum_{i=1}^{2}(p_i^A)^2,\end{aligned} \tag{3.1}$$

where we denote by $x_i'^s$, $p_i'^s$, $i=1,2$, the coordinates and momenta of particle $s=1,2,\ldots,A$ in standard units, while p_i^A is the center-of-mass momentum defined by

$$p_i^A = (M\hbar\omega)^{-\frac{1}{2}}A^{-\frac{1}{2}}\sum_{s=1}^{A}p_i'^s. \tag{3.2}$$

Introducing now the dimensionless Jacobi coordinates

$$x_i^s = (M\omega/\hbar)^{\frac{1}{2}}[s(s+1)]^{-\frac{1}{2}}\left[\sum_{t=1}^{s}x_i'^t - sx_i'^{s+1}\right], \tag{3.3}$$

and their corresponding momenta $p_i^s = -i\partial/\partial x_i^s$, we immediately see that the Hamiltonian (3.1) in units of $\hbar\omega$ becomes[17]

$$H_0 = (\mathscr{H}_0/\hbar\omega) = \frac{1}{2}\sum_{s=1}^{A-1}\sum_{i=1}^{2}[(p_i^s)^2+(x_i^s)^2]. \tag{3.4}$$

The H_0 will be the basic Hamiltonian of the following discussion as H_0 of (2.4) was for the macroscopic problem. It will be worthwhile though to consider an H with arbitrary two-body interaction which, in the present two-dimensional case, can be restricted to a central one $\Sigma_{s>t}V(|\mathbf{x}'^s-\mathbf{x}'^t|)$, where $\mathbf{x}'^s$, $\mathbf{x}'^t$ stand for two-dimensional vectors and

$|\mathbf{x}'^s - \mathbf{x}'^t|$ is the magnitude of the difference. Adding and subtracting an oscillator interaction we see that H, in units of $\hbar\omega$, can then be written as

$$H = H_0 + (\hbar\omega)^{-1}\sum_{s>t}[V(|\mathbf{x}'^s - \mathbf{x}'^t|) - [(M\omega^2/2A)(\mathbf{x}'^s - \mathbf{x}'^t)^2]. \tag{3.5}$$

Before proceeding to derive from H_0 (and later from H), its collective part, it is useful to discuss the unitary symmetry group of H_0 and subgroups of it, that will be relevant later. Denoting by

$$\eta_i^s = (1/\sqrt{2})(x_i^s - ip_i^s), \quad \xi^s{}_i = (1/\sqrt{2})(x_i^s + ip_i^s); \quad i = 1, 2; \quad s = 1, 2, ..., A-1, \tag{3.6}$$

the creation and annihilation operators associated with the Jacobi coordinates, we see from

$$[\xi_j{}^t, \eta_i^s] = \delta_{ij}\delta^{st}, \tag{3.7}$$

that the operators[18]

$$\mathbf{C}_{ij}^{st} \equiv \eta_i^s \xi_j{}^t; \quad i,j = 1, 2; \quad s,t = 1, 2, ..., A-1, \tag{3.8}$$

satisfy the commutation relations of the generators of a unitary group of $2A-2$ dimension $\mathbf{U}(2A-2)$. These operators clearly commute[8] with

$$H_0 = \sum_{s=1}^{A-1}\sum_{i=1}^{2}\eta_i^s\xi^s{}_i + (A-1), \tag{3.9}$$

and thus $\mathbf{U}(2A-2)$ is the symmetry group of the Hamiltonian. Furthermore we can contract with respect to particle or component indices to get[18]

$$\mathscr{C}_{ij} = \sum_{s=1}^{A-1}\mathbf{C}_{ij}^{ss}, \quad C^{st} = \sum_{i=1}^{2}\mathbf{C}_{ii}^{st}, \tag{3.10a),(3.10b)}$$

which are the generators of an $\mathscr{U}(2)\times\mathbf{U}(A-1)$ subgroup of $\mathbf{U}(2A-2)$ as $[\mathscr{C}_{ij}, C^{st}] = 0$. Finally, we can consider the chain of subgroups

$$\underset{[h_1h_2]}{\mathscr{U}(2)} \supset \underset{j=\frac{1}{2}(h_1-h_2)}{S\,\mathscr{U}(2)} \supset \underset{m}{\mathscr{O}(2)}, \tag{3.11a}$$

$$\underset{[h_1h_2]}{U(A-1)} \supset \underset{(\mu_1\mu_2)}{O(A-1)} \supset \underset{[f]}{D^{[A-1,1]}(S_A)}, \tag{3.11b}$$

where $\mathscr{O}(2)$, $O(A-1)$ are orthogonal groups of the dimensions indicated and $D^{[A-1,1]}(S_A)$ is the representation associated with the partition $[A-1, 1]$ of the symmetric group S_A. Underneath each group we have given the numbers characterizing its irreducible representations (irreps). For $\mathscr{U}(2)$ we have a two rowed partition $[h_1h_2]$, where $h_1 + h_2 = \mathscr{N}$ the total number of quanta and $\frac{1}{2}(h_1 - h_2) = j$ characterizes the irrep of the unitary unimodular group $S\mathscr{U}(2)$, while the total angular momentum m corresponds to the irrep of $\mathscr{O}(2)$. Turning to the group $U(A-1)$ its irrep[18] is the same as that of $\mathscr{U}(2)$, while that of $O(A-1)$ can have also at most two rows[18] $(\mu_1\mu_2)$ and the symmetric group S_A is characterized[19] by the partition $[f] = [f_1 f_2 \cdots f_A]$ of A, while its row is given by the Yamanouchi[19] symbol $r = (r_A r_{A-1} \cdots r_2 r_1)$. Thus the eigenstates of H_0 can be characterized by

$$|\mathscr{N}jm;\Omega(\mu_1\mu_2)\omega[f](r)\rangle \equiv |\Lambda m[f](r)\rangle, \tag{3.12}$$

where Ω corresponds to the set of indices required to distinguish between repeated irreps $(\mu_1\mu_2)$ of $O(A-1)$ appearing in a given irrep $[h_1h_2]$ of $U(A-1)$, while ω does the same for repeated irreps $[f]$ of S_A appearing in a given irrep $(\mu_1\mu_2)$ of $O(A-1)$. We shall also use for these states the short hand notation at the right-hand side of (3.12), where Λ stands then for $\mathscr{N}j\Omega(\mu_1\mu_2)\omega$.

The states (3.12) which, even in three dimensions,[20] can be derived in a systematic though laborious fashion, are a convenient complete basis as m, $[f]$, (r) will remain good quantum numbers for the general Hamiltonian H of (3.5). Thus the matrix of H will respect to the states (3.12) has elements of the form

$$\langle\Lambda' m[f](r)|H|\Lambda m[f](r)\rangle = (\mathscr{N} + A - 1)\delta_{\Lambda'\Lambda} + [A(A-1)/2]\langle\Lambda' m[f](r)| \times V[(2\hbar/\mu\omega)^{\frac{1}{2}}|\mathbf{x}^1|] - A^{-1}(\mathbf{x}^1)^2|\Lambda m[f](r)\rangle. \tag{3.13}$$

The eigenstates $|\mathrm{Im}[f](r)\rangle$ of H enumerated by the index $I = 1, 2, 3 ...$, are then linear combinations

$$|\mathrm{Im}[f](r)\rangle = \sum_{\Lambda} a_{I\Lambda}(m[f](r))|\Lambda m[f](r)\rangle, \tag{3.14}$$

where the coefficients $a_{I\Lambda}$ come from the diagonalization of the matrix (3.13).

B. The collective coordinates

The derivation of the collective part of the oscillator Hamiltonian H_0 of (3.4) or, more generally, of H of (3.5), can be implemented if we pass from the Jacobi coordinates to those introduced by Dzublik *et al.* and by Zickendraht.[11] In the notation of one of us[10] this transformation can be written as

$$x_i^s = \sum_{k=1}^{2}\rho_k D^1_{ki}(\vartheta)D^1_{A-3+k,s}(\alpha); \quad i,k = 1, 2; \quad s = 1, 2, ..., A-1, \tag{3.15}$$

where,[10] as we shall also show later, ρ_1^2, ρ_2^2 are connected with the principal moments of inertia of the A body system, ϑ is the Euler angle taking us to the frame of reference fixed in the body, and we have $2A-5$ coordinates α that parametrize the $O(A-1)$ group mentioned in the previous subsection. In (3.15)

$$\mathbf{D}^1(\vartheta) = \|D^1_{ki}(\vartheta)\| = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix}, \tag{3.16}$$

is a 2×2 matrix defining the irrep characterized by 1 (which is the reason of the upper index) of the $\mathscr{O}(2)$ group. We have a similar interpretation for $\|D^1_{is}(\alpha)\|$ only that now the group is $O(A-1)$ and, as we do not need the full matrix of the representation but just the rows $t = A-3+k$, we have only[10] $2A$-5 of the α's rather than the full complement of $[(A-1)(A-2)/2]$. Furthermore, in (3.15), we shall frequently substitute ρ_1, ρ_2 for

$$\rho_1 = \rho\cos\gamma, \quad \rho_2 = \rho\sin\gamma, \tag{3.17a}$$

which implies that we pass from the $2A-2$ coordinates x_i^s to ρ, γ, ϑ and the $2A-5$ variables α's.

We now see that in the Hamiltonian H_0 of (3.4) we have

$$\sum_{i=1}^{2}\sum_{s=1}^{A-1} x_i^s = \rho^2, \tag{3.17b}$$

where we take advantage of the orthogonal character of the matrices $\|D^1_{ki}(\vartheta)\|$ and $\|D^1_{ts}(\alpha)\|$. The situation for the momenta-dependent part $\Sigma_{i=1}^{2}\Sigma_{s=1}^{A-1}(p_i^s)^2$ is more complex, but explicit expressions for it have been derived[10,11] and thus H_0 can be written in terms of ρ, γ, ϑ, $-i\partial/\partial\rho$, $-i\partial/\partial\gamma$, $-i\partial/\partial\vartheta$ as well as in the α's and their derivatives.

It is possible to show[10] that ρ, ϑ and essentially γ, are invariant under permutations of the particles, while the α's are strongly affected by them. Thus, because of the interpretation mentioned above for $\rho_1, \rho_2, \vartheta$, and their invariant character under permutation, we can think of them as collective coordinates. On the other hand, the α's are more closely connected with individual particle motion as they are affected by the permutations. Thus if we project H_0 to the subspace of the full Hilbert space of the x_i^s, $i = 1, 2$; $s = 1, \ldots A-1$, in which the momenta associated with the α coordinates are zero, we get the collective part of the oscillator Hamiltonian. This collective Hamiltonian is not essentially sensitive to the number A of particles in H_0 from which we project it, so long as[10,11] this number is larger than the dimension of our space. As the latter is 2 in our case, we shall derive explicitly in Sec. 4D the collective Hamiltonian for $A = 3$, but before doing this we discuss the problem of shape of a many-body system. This will allow us later to compare the operators describing the shape in the macroscopic and microscopic collective models.

C. Shape of many-body systems

If we have a system of A particles of mass M, its inertial tensor with respect to a given frame of reference is

$$\begin{aligned} I_{ij} &= M\left[\sum_{s=1}^{A}\sum_{k=1}^{2}(x_k'^s x_k'^s)\delta_{ij} - \sum_{s=1}^{A} x_i'^s x_j'^s\right] \\ &= (\hbar/\omega)\left\{\left[\sum_{s=1}^{A-1}\sum_{k=1}^{2}(x_k^s x_k^s)\delta_{ij} - \sum_{s=1}^{A-1} x_i^s x_j^s\right]\right. \\ &\quad \left. + \left[\sum_{k=1}^{2}(x_k^A x_k^A)\delta_{ij} - x_i^A x_j^A\right]\right\}, \end{aligned} \tag{3.18}$$

where on the right-hand side we substituted the dimensionless Jacobi vectors of (3.3) and x_k^A which is $A^{1/2}$ times the center-of-mass coordinate. If now the frame of reference is fixed at the center of mass i.e., $x_i^A = 0$ and we make use of the expression (3.17b) for ρ^2 we obtain

$$q_{ij} \equiv \sum_{s=1}^{A-1} x_i^s x_j^s = \rho^2\delta_{ij} - (\omega/\hbar)I_{ij}. \tag{3.19}$$

If we pass to a frame of reference in which the two-dimensional matrix $\mathbf{q} = \|q_{ij}\|$ is diagonal, we see from (3.19) that its eigenvalues are related with the principal moments of inertia. Furthermore we note that x_i^s are given by (3.15) and, from the orthogonal character of the matrix $\|D^1_{ts}(\alpha)\|$, we immediately see that

$$q_{ij} = \sum_{k=1}^{2}\rho_k^2 D^1_{ki}(\vartheta)D^1_{kj}(\vartheta), \tag{3.20}$$

which in matrix form becomes

$$\mathbf{q} = \tilde{\mathbf{D}}^1(\vartheta)\begin{pmatrix}\rho_1^2 & 0\\ 0 & \rho_2^2\end{pmatrix}\mathbf{D}^1(\vartheta), \tag{3.21}$$

where $\mathbf{D}^1(\vartheta)$ is given by (3.16) and $\tilde{\mathbf{D}}^1(\vartheta)$ is its transposed.

Thus, we immediately conclude that ϑ is the angle of rotation that takes $\mathbf{q}$, and therefore $\mathbf{I} = \|I_{ij}\|$, to diagonal form. As $\rho^2 = \rho_1^2 + \rho_2^2$, we conclude from (3.19) that the eigenvalues ρ_1^2, ρ_2^2 of $\mathbf{q}$ are then the principal moments of inertia, and besides ρ_1, ρ_2 give a measure of the deformation of the A body system along the principal axis.

If we are dealing with specific states of H or H_0, such as (3.12) or (3.14), the expectation values of ρ_1, ρ_2 with respect to them provide a measure of the deformation. But as ρ_1, ρ_2 are not polynomial functions of x_i^s it is better to take a function of ρ_1, ρ_2 that has this property. We note from (3.21) that

$$\mathrm{tr}\mathbf{q} = \rho_1^2 + \rho_2^2 = \rho^2, \quad \det\mathbf{q} = \rho_1^2\rho_2^2 = (\rho^4\sin^2 2\gamma)/4, \tag{3.22}$$

and

$$(\mathrm{tr}\mathbf{q})^2 - 4\det\mathbf{q} = (\rho_1^2 - \rho_2^2)^2 = (\rho^2\cos 2\gamma)^2, \tag{3.23}$$

where the right-hand side is a good measure of the deformation and besides is a polynomial function of the q_{ij} and thus of the x_i^s. In the following discussion, we shall use the expectation value of $\rho^4\cos^2 2\gamma$ as a measure of the deformation in the microscopic collective model.

D. The three-body system

As indicated at the end of Sec. 3B, we can derive the collective part of the oscillator Hamiltonian (3.4) even if we restrict ourselves to three particles. In this case $A = 3$, $2A - 5 = 1$, $s = 1, 2$ and thus in the transformation (3.15) there is only one α and the $D^1_{ks}(\alpha)$ has the same form as $D^1_{ki}(\vartheta)$ of (3.16). Introducing in this subsection the short hand notation

$$b = \cos\vartheta,\ r = \sin\vartheta;\ c = \cos\alpha,\ s = \sin\alpha; \tag{3.24}$$

$$d = \cos\gamma,\ t = \sin\gamma,\ \sigma = \sec 2\gamma,$$

we see from (3.15) that we can write

$$x_1^1 \equiv y_1 = \rho(dbc + trs),\ x_1^2 \equiv y_2 = \rho(-dbs + trc), \tag{3.25a, b}$$

$$x_2^1 \equiv y_3 = \rho(-drc + tbs),\ x_2^2 \equiv y_4 = \rho(drs + tbc). \tag{3.25c, d}$$

Furthermore we shall also use the notation

$$\rho \equiv y_1',\ \gamma \equiv y_2',\ \vartheta \equiv y_3',\ \alpha \equiv y_4', \tag{3.26}$$

in which case the matrix $\|\partial y_n/\partial y_m'\|$, $m = n = 1, 2, 3, 4$ becomes

$$\|\partial y_n/\partial y_m'\| = \begin{array}{c|cccc} m\backslash n & 1 & 2 & 3 & 4 \\ \hline 1 & (dbc + trs) & (-dbs + trc) & (-drc + tbs) & (drs + tbc) \\ 2 & \rho(-tbc + drs) & \rho(tbs + drc) & \rho(trc + dbs) & \rho(-trs + dbc) \\ 3 & \rho(-drc + tbs) & \rho(drs + tbc) & \rho(-dbc - trs) & \rho(dbs - trc) \\ 4 & \rho(-dbs + trc) & \rho(-dbc - trs) & \rho(drs + tbc) & \rho(drc - tbs) \end{array} \tag{3.27}$$

Inverting this matrix we obtain

$$\|\partial y'_n/\partial y_m\| = \begin{array}{c|cccc} m \backslash n & 1 & 2 & 3 & 4 \\ \hline 1 & (bdc+rts) & \rho^{-1}(-btc+rds) & \rho^{-1}(-rdc-bts)\sigma & \rho^{-1}(-rtc-bds)\sigma \\ 2 & (-bds+rtc) & \rho^{-1}(bts+rdc) & \rho^{-1}(rds-btc)\sigma & \rho^{-1}(rts-bdc)\sigma \\ 3 & (-rdc+bts) & \rho^{-1}(rtc+bds) & \rho^{-1}(-bdc+rts)\sigma & \rho^{-1}(-btc+rds)\sigma \\ 4 & (rds+btc) & \rho^{-1}(-rts+bdc) & \rho^{-1}(bds+rtc)\sigma & \rho^{-1}(bts+rdc)\sigma \end{array} \tag{3.28}$$

as can be easily checked by direct multiplication, and from it the contravariant metric tensor becomes

$$g^{mn} = \sum_{l=1}^{4} \frac{\partial y'_m}{\partial y_l}\frac{\partial y'_n}{\partial y_l};\ l, m, n = 1, 2, 3, 4. \tag{3.29}$$

The Laplacian in the Hamiltonian H_0 of (3.4), in which $A = 3$, is then

$$\sum_{l=1}^{4} \frac{\partial^2}{\partial y_l^2} = \sum_{m,n} g^{-1/2}\frac{\partial}{\partial y'_m} g^{mn} g^{1/2}\frac{\partial}{\partial y'_n}, \tag{3.30}$$

where $g^{-1} = \det\|g^{mn}\|$ and thus we obtain

$$H_0 = \frac{1}{2}\left\{ -\frac{1}{\rho^3}\frac{\partial}{\partial\rho}\rho^3\frac{\partial}{\partial\rho} - \frac{1}{\rho_2}\left[\frac{1}{\cos2\gamma}\frac{\partial}{\partial\gamma}\cos2\gamma\frac{\partial}{\partial\gamma} + \frac{1}{\cos^2 2\gamma}\left(\frac{\partial^2}{\partial\vartheta^2} + \frac{\partial^2}{\partial\alpha^2}\right) + \frac{2\sin2\gamma}{\cos^2 2\gamma}\frac{\partial^2}{\partial\vartheta\partial\alpha}\right] + \rho^2 \right\}. \tag{3.31}$$

It is not enough to have only H_0 in the coordinates $\rho, \gamma, \vartheta, \alpha$. We need also to consider the generators of the $S\mathcal{U}(2)$ group $\mathcal{C}_{ij}$ of (3.10a) in these coordinates and momenta, as we shall project this group onto the Hilbert space associated with the collective variables, thus getting integrals of motion of the collective Hamiltonian. It will be enough for our purpose to discuss these generators in classical mechanics as the corresponding quantum mechanical operators shall be obtained directly in Sec. E, when we transform the collective part of H_0 into the Hamiltonian of the pseudo-coulomb problem.[14]

We note from (3.10a), (3.6), and (3.7) that we can write

$$\mathcal{C}_{ij} = \frac{1}{2}\sum_{s=1}^{2}(x_i^s x_j^s + p_i^s p_j^s) + \frac{i}{2}\sum_{s=1}^{2}(x_i^s p_j^s - x_j^s p_i^s). \tag{3.32}$$

Thus if we use for x_i^s the notation y_n, $n = 1, 2, 3, 4$ as in (3.25) and denote by p_n the corresponding momenta of the latter,i.e.,

$$p_1 \equiv p_1^1,\ p_2 \equiv p_1^2,\ p_3 \equiv p_2^1,\ p_4 \equiv p_2^2, \tag{3.33}$$

we can define

$$K_1 \equiv \tfrac{1}{2}(\mathcal{C}_{11} - \mathcal{C}_{22}) = \tfrac{1}{4}(y_1^2 + y_2^2 - y_3^2 - y_4^2) + \tfrac{1}{4}(p_1^2 + p_2^2 - p_3^2 - p_4^2), \tag{3.34a}$$

$$K_2 \equiv -\tfrac{1}{2}(\mathcal{C}_{12} + \mathcal{C}_{21}) = -\tfrac{1}{2}(y_1y_3 + y_2y_4 + p_1p_3 + p_2p_4), \tag{3.34b}$$

$$K_3 \equiv \frac{i}{2}(\mathcal{C}_{12} - \mathcal{C}_{21}) = -\tfrac{1}{2}(y_1p_3 - y_3p_1 + y_2p_4 - y_4p_2), \tag{3.34c}$$

where the K's satisfy the Poisson bracket relation $\{K_1, K_2\} = K_3$ and its cyclic permutations.

The coordinates y_n are given in terms of y'_n through (3.25) and (3.26). Furhtermore denoting by p'_n the canonically conjugate variable to y'_n we know that

$$p_m = \sum_{n=1}^{4}\frac{\partial y'_n}{\partial y_m}p'_n,\ m = 1, 2, 3, 4, \tag{3.35}$$

where the matrix $\|\partial y'_n/\partial y_m\|$ is given in (3.28). Thus we can immediately write K_1, K_2, K_3 in terms of y'_n, p'_m and we give them below where, for clarity sake, we replaces y'_n and p'_n, $n = 1, 2, 3, 4$ by $\rho, \gamma, \vartheta, \alpha$ and $p_\rho, p_\gamma, p_\vartheta, p_\alpha$

$$K_1 = \tfrac{1}{4}\cos2\gamma\cos2\vartheta\left\{\rho^2 + p_\rho^2 + \frac{1}{\rho^2}[-p_\gamma^2 + \sec^2 2\gamma(-p_\vartheta^2 + p_\alpha^2)]\right\} - \frac{1}{2\rho}\left[\sin2\gamma\cos2\vartheta p_\rho p_\gamma + \frac{\sin2\vartheta}{\cos2\gamma}(p_\vartheta p_\rho + \sin2\gamma p_\rho p_\alpha)\right] - \frac{1}{2\rho^2}\sin2\vartheta p_\gamma p_\alpha, \tag{3.36a}$$

$$K_2 = \tfrac{1}{4}\cos2\gamma\sin2\vartheta\left\{\rho^2 + p_\rho^2 + \frac{1}{\rho^2}[-p_\gamma^2 + \sec^2 2\gamma(-p_\vartheta^2 + p_\alpha^2)]\right\} - \frac{1}{2\rho}[\sin2\gamma\sin2\vartheta p_\rho p_\gamma - \cos2\vartheta\sec2\gamma(p_\rho p_\vartheta + \sin2\gamma p_\rho p_\alpha)] + \frac{1}{2\rho^2}\cos2\vartheta p_\gamma p_\alpha, \tag{3.36b}$$

$$K_3 = -\tfrac{1}{2}p_\vartheta. \tag{3.36c}$$

We notice the K's do not depend on α though they do on p_α. This we could have predicted from the fact that $\mathscr{C}_{ij}$ of (3.10a) commutes with C^{sl} of (3.10b). As the generator of the O(2) group associated with the variable α is $C^{12}-C^{21}$, it commutes with $\mathscr{C}_{ij}$ and thus the latter is an invariant of O(2) and cannot change its form when $\alpha\to\alpha+\alpha_0$ i.e., it must be independent of α. This property is of interest because then K_l, H_0 (when the latter is considered as a classical observable) will continue to satisfy the Poisson bracket relations[21]

$$\{K_1, K_2\} = K_3$$

and cyclically,

$$\{K_l, H_0\} = 0 \quad l = 1, 2, 3, \tag{3.37}$$

even when $p_\alpha = 0$ as can be seen from the fact that, for example,

$$\{K_l, H_0\} = \sum_{n=1}^{3}\left(\frac{\partial K_l}{\partial y_n'}\frac{\partial H_0}{\partial p_n'} - \frac{\partial K_l}{\partial p_n'}\frac{\partial H_0}{\partial y_n'}\right) + \left(\frac{\partial K_l}{\partial \alpha}\frac{\partial H_0}{\partial p_\alpha} - \frac{\partial K_l}{\partial p_\alpha}\frac{\partial H_0}{\partial \alpha}\right), \tag{3.38}$$

and as the K_l, H_0 are independent of α, the final bracket on the right-hand side will always be zero.

As a last point in our analysis of the three-body problem, we would like to derive explicitly the states (3.12) for this case, in terms of polynomials in the creation operators η_i^s acting in the ground state $|0\rangle$. We first notice the behavior of the Jacobi coordinates under permutation. As any permutation can be constructed from the transposition (1, 2) and the cyclic permutation of all the particles,[19] we need only to apply (1, 2), (1, 2, 3) to $\mathbf{x}^s$, $s = 1, 2$,[17] i.e.,

$$(1,2)\begin{pmatrix}\mathbf{x}^1\\ \mathbf{x}^2\end{pmatrix} = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{x}^1\\ \mathbf{x}^2\end{pmatrix},$$

$$(1,2,3)\begin{pmatrix}\mathbf{x}^1\\ \mathbf{x}^2\end{pmatrix} = \begin{pmatrix}-\frac{1}{2} & \frac{1}{2}\sqrt{3}\\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}\mathbf{x}^1\\ \mathbf{x}^2\end{pmatrix}. \tag{3.39}$$

The same property holds for the creation operator $\boldsymbol{\eta}^s$, $s = 1, 2$, of (3.6), and thus if we define[17]

$$\boldsymbol{\eta}^{\pm} \equiv (1/\sqrt{2})(\mp i\boldsymbol{\eta}^1 + \boldsymbol{\eta}^2), \tag{3.40}$$

$$(1,2)\begin{pmatrix}\boldsymbol{\eta}^+\\ \boldsymbol{\eta}^-\end{pmatrix} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{\eta}^+\\ \boldsymbol{\eta}^-\end{pmatrix},$$

$$(1,2,3)\begin{pmatrix}\boldsymbol{\eta}^+\\ \boldsymbol{\eta}^-\end{pmatrix} = \begin{pmatrix}e^{-i\frac{2\pi}{3}} & 0\\ 0 & e^{i\frac{2\pi}{3}}\end{pmatrix}\begin{pmatrix}\boldsymbol{\eta}^+\\ \boldsymbol{\eta}^-\end{pmatrix}. \tag{3.41}$$

It is furthermore convenient to consider the vector $\boldsymbol{\eta}^{\pm}$ not in Cartesian coordinates $\eta_i^{\pm}$, $i = 1, 2$, but in the form

$$\eta_+^{\pm} = (1/\sqrt{2})(-i\eta_1^{\pm} + \eta_2^{\pm}),$$
$$\eta_-^{\pm} = (1/\sqrt{2})(i\eta_1^{\pm} + \eta_2^{\pm}), \tag{3.42}$$

and similar relations for $\xi_{\pm}^{s}$ in terms of the ξ_i^s. The generators $\mathscr{C}_{ij}$ of $\mathscr{U}(2)$ and C^{sl} of $U(2)$ appearing in (3.10) can be transformed then into

$$\mathscr{C}_{++},\ \mathscr{C}_{--},\ \mathscr{C}_{+-},\ \mathscr{C}_{-+}$$

and

$$C^{++},\ C^{--},\ C^{+-},\ C^{-+}, \tag{3.43}$$

Clearly the generators $\mathscr{O}(2)$ and $O(2)$ in (3.11) instead of $\mathscr{C}_{12}-\mathscr{C}_{21}$ and $C^{12}-C^{21}$ will now be given by $\mathscr{C}_{++}-\mathscr{C}_{--}$ and $C^{++}-C^{--}$.

The normalized state of highest weight[18] associated with the partition $[h_1h_2] = [(\mathscr{N}/2)+j, (\mathscr{N}/2)-j]$, where j is integer or semi-integer according to whether $\mathscr{N}$ is even or odd, is then given by[18]

$$(2j+1)^{\frac{1}{2}}\{[(\mathscr{N}/2)+j+1]![(\mathscr{N}/2)-j]!\}^{-\frac{1}{2}} (\eta_+^+)^{2j}(\eta_+^+\eta_-^- - \eta_-^+\eta_+^-)^{((\mathscr{N}/2)-j)}|0\rangle, \tag{3.44}$$

as the action of $\mathscr{C}_{+-}$, C^{+-} on it gives zero. Applying then the operator [18]

$$\left[\frac{(j+m)!}{(j-m)!(2j)!}\right]^{\frac{1}{2}}(\mathscr{C}_{-+})^{j-m}\left[\frac{(j+\mu)!}{(j-\mu)!(2j)!}\right]^{\frac{1}{2}}(C^{-+})^{j-\mu}, \tag{3.45}$$

to (3.44), we get the normalized states $|\mathscr{N}jm\mu\rangle$ corresponding to $\mathscr{N}$ quanta and characterized by the irreps j, m, μ of $S\mathscr{U}(2)$, $\mathscr{O}(2)$, $O(2)$, respectively, in the form

$$\begin{aligned}|\mathscr{N}jm\mu\rangle &= (2j+1)^{\frac{1}{2}}\{[(\mathscr{N}/2)+j+1]![(\mathscr{N}/2)-j]!\}^{-\frac{1}{2}}\\ &\times[(j+m)!(j-m)!(j+\mu)!(j-\mu)!]^{\frac{1}{2}}\\ &\times(\eta_+^+\eta_-^- - \eta_-^+\eta_+^-)^{((\mathscr{N}/2)-j)}\\ &\times\sum_s\{[(j-m-s)!s!(\mu+m+s)!(j-\mu-s)!]^{-1}\\ &\times(\eta_-^+)^{j-m-s}(\eta_-^-)^s(\eta_+^+)^{\mu+m+s}(\eta_+^-)^{j-\mu-s}\}|0\rangle.\end{aligned} \tag{3.46}$$

To get now the states $|\mathscr{N}jm,\mu\{f\}(r)\rangle$, corresponding to (3.12) for $A = 3$ we need to apply the permutations (1, 2) and (1, 2, 3) to the states (3.46). From (3.41) we then easily obtain that

$$(1,2,3)|\mathscr{N}jm\mu\rangle = [\exp(-i2\pi/3)]^{2\mu}|\mathscr{N}jm\mu\rangle, \tag{3.47a}$$

$$(1,2)|\mathscr{N}jm\mu\rangle = (-1)^{((\mathscr{N}/2)-j)}|\mathscr{N}jm-\mu\rangle. \tag{3.47b}$$

Thus we immediately see that

$$\left|\mathscr{N}jm,\mu\begin{Bmatrix}\{3\} & (111)\\ \{111\} & (321)\end{Bmatrix}\right\rangle = (1/\sqrt{2})[|\mathscr{N}jm\mu\rangle \pm (-1)^{((\mathscr{N}/2)-j)}|\mathscr{N}jm-\mu\rangle], \tag{3.48a}$$

when $2\mu\equiv 0 \bmod 3$ but $\mu\neq 0$, where the symmetric (antisymmetric)states correspond to $+(-)$ sign on the right-hand side. For $\mu = 0$ we have

$$\left|\mathscr{N}jm0\begin{Bmatrix}\{3\} & (111)\\ \{111\} & (321)\end{Bmatrix}\right\rangle = \begin{cases}|\mathscr{N}jm0\rangle \text{ if } ((\mathscr{N}/2)-j \text{ even}\\ |\mathscr{N}jm0\rangle \text{ if } ((\mathscr{N}/2)-j \text{ odd}\end{cases} \tag{3.48b}$$

Finally, when $2\mu\equiv 1, 2 \bmod 3$ we have

$$\left|\mathscr{N}jm,\mu\{21\}\begin{matrix}(211)\\ (121)\end{matrix}\right\rangle = (1/\sqrt{2})[|\mathscr{N}jm\mu\rangle] \pm(-1)^{((\mathscr{N}/2)-j)}|\mathscr{N}jm-\mu\rangle, \tag{3.48c}$$

with Yamanouchi symbols (211) and (121) corresponding, respectively to the $+$, $-$ signs on the right-hand side.

As the last point we wish to find the number Δ of states $|\mathcal{N} j m \mu\rangle$ of total number of quanta $\mathcal{N}$, as well as the number of states d of fixed $\mathcal{N}$ but with $\mu = 0$, such that these states are symmetric under interchange of all particles which, from (3.48b), imply $(\mathcal{N}/2) - j$ is even. For the former, when $\mathcal{N}$ is even, we see that the number is given by[22]

$$\begin{aligned}\Delta &= \sum_{j=0}^{(\mathcal{N}/2)} (2j+1)^2 \\ &= 4\sum_{j=1}^{(\mathcal{N}/2)} j^2 + 4\sum_{j=1}^{(\mathcal{N}/2)} j + (\mathcal{N}/2) + 1 \\ &= (1/6)(\mathcal{N}+1)(\mathcal{N}+2)(\mathcal{N}+3),\end{aligned} \tag{3.49}$$

and the same result holds for $\mathcal{N}$ odd. Clearly Δ represents the degeneracy of the basis for the single rowed irreducible representation $[\mathcal{N}]$ of U(4).

For the number d we note that if $\mu = 0$ then j must be integer and thus $\mathcal{N}$ is even as we remarked before that $(\mathcal{N}/2) - j$ is always integer. We then require

$$d = \sum_{j}{}' {}^{(\mathcal{N}/2)} (2j+1), \tag{3.50}$$

where $'$ implies that j is taken over values such that $(\mathcal{N}/2) - j$ is also even. Assuming that $(\mathcal{N}/4)$ is an integer we then get

$$\begin{aligned}d &= \sum_{k=0}^{(\mathcal{N}/4)} \{2[(\mathcal{N}/2) - 2k] + 1\} \\ &= \sum_{k=0}^{(\mathcal{N}/4)} (\mathcal{N} - 4k + 1) = (1/2)[(\mathcal{N}/2)+1][(\mathcal{N}/2)+2],\end{aligned} \tag{3.51}$$

and the same result holds when $(\mathcal{N}/4)$ is half-integer. Thus the total number d of symmetric states i.e., $\{f\} = \{3\}$ with $\mu = 0$, corresponding to a given even number of quanta $\mathcal{N}$ is the same as the degeneracy of the basis for the single rowed irrep $[\mathcal{N}/2]$ of a three-dimensional unitary group. As we show in the next subsection that the symmetric states with $\mu = 0$ are the solutions of the microscopic collective Hamiltonian, they are then characterized by an irrep of $U(3)$.

E. The microscopic collective Hamiltonian

In this subsection we proceed to project out from the three-body oscillator Hamiltonian (3.31) its collective part and discuss the relevant integrals of motion of the latter. To achieve this purpose we first analyze the behavior of the coordinates ρ, γ, ϑ, α introduced in (3.15), (3.24), and (3.25) under the permutations of the three particle system. We could use (3.25) to express ρ, γ, ϑ, α in terms of the x_i^s and, from the transformation properties (3.39) of the latter under permutation, we obtain

$$\begin{aligned}&(1,2)\rho = \rho, \quad (1,2,3)\rho = \rho; \\ &(1,2)\vartheta = \vartheta, \quad (1,2,3)\vartheta = \vartheta; \\ &(1,2)\gamma = -\gamma, \quad (1,2,3)\gamma = \gamma; \\ &(1,2)\alpha = \pi - \alpha, \quad (1,2,3)\alpha = \alpha - 2\pi/3.\end{aligned} \tag{3.52}$$

Inversely, if we assume (3.52) we easily check that x_i^s transfrom as Jacobi vectors i.e., in the form (3.39).

We note that ρ, ϑ are invariant under arbitrary permutations as these can be constructed from products of (1, 2) and (1, 2, 3). The γ variable is not completely invariant under permutations as the transposition (1, 2) affects its sign. On the other hand, the α variable is strongly affected by the permutation. From these properties, together with the fact discussed in (3.21) that $\rho_1^2 = \rho^2\cos^2\gamma$, $\rho_2^2 = \rho^2\sin^2\gamma$, give the principal moments of inertia of the three-particle system and ϑ provides the rotation that diagonalizes the inertia tensor, we conclude that the collective part of the Hamiltonian (3.31) is the one that is independent of $p_\alpha = -i\partial/\partial\alpha$ [i.e., its eigenstates are scalars of O(2)] and designating it as H_C we have

$$\begin{aligned}H_C = \frac{1}{2}\Bigg\{ &-\frac{1}{\rho^3}\frac{\partial}{\partial\rho}\rho^3\frac{\partial}{\partial\rho} - \frac{1}{\rho^2} \\ &\times\left[\frac{1}{\cos 2\gamma}\frac{\partial}{\partial\gamma}\cos 2\gamma\frac{\partial}{\partial\gamma} + \frac{1}{\cos^2 2\gamma}\frac{\partial^2}{\partial\vartheta^2}\right] + \rho^2\Bigg\}.\end{aligned} \tag{3.53}$$

We note furthermore that we are only interested in eigenstates of this Hamiltonian that would be *invariant* [10] under permutations of the particles. As only γ is affected, and this by the the transportation (1, 2) $\gamma = -\gamma$, we conclude that the only collective solutions of H_C are those that remain invariant under the operation that takes $\gamma \to -\gamma$.

The eigenstates of (3.53) with the above property are trivial to obtain as the Hamiltonian is separable. We shall prefer though to carry the point transformation

$$r = \rho^2/2, \quad \theta = 2\gamma + (\pi/2), \quad \varphi = 2\vartheta, \tag{3.54}$$

as this reduces the Hamiltonian to the well-known pseudo-Coulomb problem[14]

$$\tfrac{1}{2}H_C = \tfrac{1}{2}r(-\nabla^2 + 1), \tag{3.55}$$

where ∇^2 is the Laplacian in the spherical coordinates.

One set of eigenstates of (3.55) is well known

$$|nlm\rangle = \mathcal{R}_{nl}(r)Y_{lm}(\theta,\varphi), \tag{3.56a}$$

where the Y's are spherical harmonics and the normalized $\mathcal{R}_{nl}(r)$ are given by[14]

$$\begin{aligned}\mathcal{R}_{nl}(r) = &(-1)^l 2^{l+1/2}[2(n-l)!/\Gamma(n+l+2)]^{1/2} \\ &\times r^l e^{-r} L_{n-l}^{2l+1}(2r),\end{aligned} \tag{3.56b}$$

where L is an associated Laguerre polynomial. The eigenvalue[14] of H_C corresponding to the eigenstate (3.56) is $2(n+1)$, $n = 0, 1, 2, 3, \ldots$.

Note that the eigenstates (3.56) must be restricted by the fact that they are invariant under the operation $\gamma \to -\gamma$, i.e.

$$\theta \to \pi - \theta. \tag{3.57}$$

From the properties of the associated Legendre polynomials (p. 1015 of Ref. 15) we conclude that the spherical harmonics $Y_{lm}(\theta,\varphi)$ transform under the operation (3.57) as

$$Y_{lm}(\theta,\varphi) \to (-1)^{l-m}Y_{lm}(\theta,\varphi), \tag{3.58}$$

Thus the only allowed collective solutions (3.56) are those in which $l - m$ is even.

There are though other solutions of (3.53) or equivalently (3.55), that are more interesting. As (3.53) is obtained from (3.31) when $p_\alpha = 0$, clearly the solutions of the latter with

$\mu = 0$ and invariant under permutations, i.e., the upper row of (3.48b) are solutions of (3.53). They are characterized by a j quantum number associated with the irrep of the $S\mathscr{U}(2)$ symmetry group (3.10a). For these states the m takes all the values $m = j, j-1, \ldots, -j$ and not as is in (3.56) only those in which $l - m$ is even. Unfortunately the states $|\mathscr{N} jm\mu\rangle$ of (3.46), of which the upper line of (3.48b) is a particular case, are expressed as creation operators acting on the ground state and not as functions of ρ, γ, ϑ or equivalently r, θ, φ. We shall proceed to show that, with the help of the generators of the symmetry group O(4) of the pseudo-Coulomb problem,[23] it is possible to express the solutions $|\mathscr{N} jm0\rangle$ when $\mathscr{N} = 2n$, $n - j$ even in terms of those of the pseudo-Coulomb problem in parabolic coordinates or the ones of (3.56) when $l - m$ is even.

We recall that the pseudo-Coulomb Hamiltonian[23,24] (3.55), besides the integrals of motion that are components of the angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \tag{3.59}$$

has the components of the Runge–Lenz vector[23,24]

$$\mathbf{A} = \tfrac{1}{2}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}) + \tfrac{1}{2}\mathbf{r}(p^2 + 1), \tag{3.60}$$

where $\mathbf{r}$ is the vector characterized in spherical coordinates by (r, θ, φ) and, in the operator case, $\mathbf{p} = -i\nabla$ is the corresponding momentum. The commutation rules are then[23,24]

$$[H_C, \mathbf{L}] = [H_C, \mathbf{A}] = 0, \tag{3.61a}$$

$$\mathbf{L}\times\mathbf{L} = i\mathbf{L}, \quad \mathbf{L}\times\mathbf{A} = i\mathbf{A}, \quad \mathbf{A}\times\mathbf{A} = i\mathbf{L}. \tag{3.61b}$$

The kets $|nlm\rangle$ of (3.56) are then eigenstates of H_C, L^2, L_3. Another set of eigenstates is obtained when we consider the vector operators[24]

$$\mathbf{M} = \tfrac{1}{2}(\mathbf{L} + \mathbf{A}), \quad \mathbf{N} = \tfrac{1}{2}(\mathbf{L} - \mathbf{A}), \tag{3.62}$$

which from (3.61) satisfy the commutation relations

$$\mathbf{M}\times\mathbf{M} = i\mathbf{M}, \quad \mathbf{N}\times\mathbf{N} = i\mathbf{N}, \quad [\mathbf{M}, \mathbf{N}] = 0. \tag{3.63}$$

Furthermore from (3.59) and (3.60) $\mathbf{L}\cdot\mathbf{A} = 0$ and thus $M^2 = N^2$.

From (3.55), (3.60), and (3.62) we also conclude that[24]

$$H_C^2 = 4(4N^2 + 1) = 4(4M^2 + 1). \tag{3.64}$$

Thus the eigenstates of H_C can be characterized by the eigenvalues $s(s+1)$ of $M^2 = N^2$ [where from (3.64) $s = (n/2)$], σ of M_3, and τ of N_3 as these are commuting operators. Denoting the corresponding ket by $|(n/2)\sigma\tau\rangle$ we review in the Appendix its connection with the solution in terms of parabolic coordinates. As from (3.62)

$$\mathbf{L} = \mathbf{M} + \mathbf{N}, \tag{3.65}$$

we immediately conclude that[23,24]

$$|nlm\rangle = \sum_{\sigma\tau} |(n/2)\sigma\tau\rangle \langle (n/2)\sigma, (n/2)\tau | lm\rangle, \tag{3.66}$$

where the bracket is a standard Clebsch–Gordan coefficient. From the orthonormal properties of these coefficients we have furthermore that

$$|(n/2)\sigma\tau\rangle = \sum_{l=0}^{n} \sum_{m=-l}^{l} |nlm\rangle \langle (n/2)\sigma, (n/2)\tau | lm\rangle. \tag{3.67}$$

The invariance we must require of the collective states when $\gamma \to -\gamma$ or, equivalently, $\theta \to \pi - \theta$, reflects itself in the fact that for $|nlm\rangle$, $l - m$ is even. For $|(n/2)\sigma\tau\rangle$ we show in the Appendix that the states that have this invariance property are

$$(1/2)[\,|(n/2)\sigma\tau\rangle + (-1)^{n-\sigma-\tau}|(n/2)\tau\sigma\rangle\,]. \tag{3.68}$$

It is easy to check, from the symmetry properties of Clebsch–Gordan coefficients,[25] that if we multiply (3.68) by $\langle (n/2)\sigma, (n/2)\tau | lm\rangle$ and sum over σ, τ we get the states $|nlm\rangle$ with the selection rule $l - m$ even.

While the eigenkets $|nlm\rangle$, $|(n/2)\sigma\tau\rangle$ of the pseudo-Coulomb problem (3.55) are well known,[23,24] neither provides the type of solution associated with $|\mathscr{N} jm0\rangle$, $(\mathscr{N}/2) - j$ even, discussed in the previous section. The reason is that we do not take into account *ab initio* in the vectors $\mathbf{L}, \mathbf{A}$ of (3.59) and (3.60) the condition of invariance under the operation $\theta \to \pi - \theta$. To find then how to avoid this problem, we start by noticing that the components (x_1, x_2, x_3) of the vector $\mathbf{r}$

$$x_1 = r\sin\theta\cos\varphi, \quad x_2 = r\sin\theta\sin\varphi, \quad x_3 = r\cos\theta, \tag{3.69}$$

transform under the operation $\theta \to \pi - \theta$ as

$$x_1 \to x_1, \quad x_2 \to x_2, \quad x_3 \to -x_3. \tag{3.70}$$

This implies for the momentum vector $\mathbf{p} = -i\nabla$ that the corresponding components (p_1, p_2, p_3) transform as

$$p_1 \to p_1, \quad p_2 \to p_2, \quad p_3 \to -p_3. \tag{3.71}$$

Turning then to $\mathbf{L}, \mathbf{A}$ of (3.59) and (3.60) we have that

$$L_1 \to -L_1, \quad L_2 \to -L_2, \quad L_3 \to L_3; \tag{3.72a}$$

$$A_1 \to A_1, \quad A_2 \to A_2, \quad A_3 \to -A_3. \tag{3.72b}$$

Thus of the six generators $\mathbf{L}, \mathbf{A}$ of the symmetry group O(4) of the pseudo-Coulomb problem, only three A_1, A_2, L_3 are invariant under the transformation $\theta \to \pi - \theta$.

From the commutation relations (3.61b) the invariant operators satisfy the commutation relations

$$[A_1, A_2] = iL_3, \quad [L_3, A_1] = iA_2, \quad [A_2, L_3] = iA_1, \tag{3.73}$$

and therefore they are generators of an $S\mathscr{U}(2)$ group. Thus we could also characterize the eigenstates of the H_C of (3.55) with eigenvalue $2(n+1)$ by the irreps of this $S\mathscr{U}(2)$ group and of its $\mathscr{O}(2)$ subgroup i.e., as eigenkets $|njm\rangle$ of

$$J^2 \equiv A_1^2 + A_2^2 + L_3^2, \quad J_3 = L_3, \tag{3.74}$$

with eigenvalues $j(j+1)$ and m.

We shall now proceed to show that $J_1 = A_1, J_2 = A_2$, $J_3 = L_3$ considered as classical observables are identical to the K_1, K_2, K_3 of (3.36) if we take in the latter $p_\alpha = 0$. For this purpose we notice that classically

$$\mathbf{A} = (\mathbf{L}\times\mathbf{p}) + \tfrac{1}{2}\mathbf{r}(p^2 + 1) = (\mathbf{r}\cdot\mathbf{p})\mathbf{p} - \tfrac{1}{2}\mathbf{r}(p^2 - 1). \tag{3.75}$$

Introducing now the spherical coordinates r, θ, φ, and their corresponding momenta p_r, p_θ, p_φ, we get for the three components (x_1, x_2, x_3) of $\mathbf{r}$ the expression (3.69) while for the corresponding momenta we have

$$p_i = \frac{\partial r}{\partial x_i} p_r + \frac{\partial \theta}{\partial x_i} p_\theta + \frac{\partial \varphi}{\partial x_i} p_\varphi, \quad i = 1, 2, 3, \tag{3.76}$$

from which it follows that

$$(\mathbf{r}\cdot\mathbf{p}) = rp_r, \quad p^2 = p_r^2 + r^{-2}p_\theta^2 + (r\sin\theta)^{-2}p_\varphi^2. \tag{3.77}$$

Thus we obtain

$$J_1 = \tfrac{1}{2}\sin\theta\cos\varphi\left\{r + rp_r^2 - \frac{1}{r}(p_\theta^2 + \frac{1}{\sin^2\theta}p_\varphi^2)\right\} + \cos\theta\cos\varphi p_r p_\theta - (\sin\varphi/\sin\theta)p_r p_\varphi, \quad (3.78a)$$

$$J_2 = \tfrac{1}{2}\sin\theta\sin\varphi\left\{r + rp_r^2 - \frac{1}{r}(p_\theta^2 + \frac{1}{\sin^2\theta}p_\varphi^2)\right\} + \cos\theta\sin\varphi p_r p_\theta + (\cos\varphi/\sin\theta)p_r p_\varphi, \quad (3.78b)$$

$$J_3 = -p_\varphi. \quad (3.78c)$$

We can also express J_1, J_2, J_3 in terms of ρ, γ, ϑ by writing r, θ, φ in terms of the former using (3.54) and p_r, p_θ, p_φ in terms p_ρ, p_r, p_ϑ, again using (3.54),

$$p_r = \rho^{-1}p_\rho, \quad p_\theta = p_\gamma/2, \quad p_\varphi = p_\vartheta/2. \quad (3.79)$$

It is immediate then to see that J_i coincides with K_i, $i = 1, 2, 3$ of (3.36) when $p_\alpha = 0$.

The $S\mathscr{U}(2)$ group of the three or, for that matter, A particle problem, whose generators are K_1, K_2, K_3, can be projected into the subspace $p_\alpha = 0$, still maintaining the Poisson bracket relations $\{K_1, K_2\} = K_3$ and cyclically, as shown in (3.38). These projected generators K_i, $i = 1, 2, 3$ of $S\mathscr{U}(2)$ are then integrals of motion of the collective Hamiltonian and, as we just showed, can be identified with the J_i, $i = 1, 2, 3$.

In the Appendix we show that the eigenstates $|njm\rangle$ of the operators H_C, J^2, J_3 can be written down immediately in terms of the $|(n/2)\sigma\tau\rangle$, of the parabolic coordinate solution through the relation,

$$|njm\rangle = \sum_\sigma (-1)^{(n/2)-\sigma+m}\langle(n/2)\sigma, (n/2)m - \sigma|jm\rangle |(n/2)\sigma m - \sigma\rangle. \quad (3.80)$$

But from the discussion at the end of the previous subsection we see that $|\mathscr{N}jm0\rangle$ of (3.48b), where $\mathscr{N}$ and $(\mathscr{N}/2) - j$ are even, is an eigenstate $K^2 = K_1^2 + K_2^2 + K_3^2$ and K_3 when the eigenvalue μ of p_α is taken as 0. Thus we have

$$|\mathscr{N}jm0\rangle = |njm\rangle, \quad \mathscr{N} = 2n, \quad n - j \text{ even}, \quad (3.81)$$

and the only values of j of interest to our problem are

$$j = n, n-2, \ldots, 1 \text{ or } 0. \quad (3.82)$$

For a fixed n the number of states (3.80) with the restriction (3.82) is then clearly $(1/2)(n+1)(n+2)$ and thus, as the spectrum of the Hamiltonian $(H_C/2)$ is equally spaced, we could map the states (3.80) on those of a three-dimensional oscillator with definite angular momentum j and projection m. This will be done in Sec. IV to obtain the quantum mechanical representation of the canonical transformation that takes us from the macroscopic to the microscopic collective model.

In the final subsection of this section, we shall analyze the general collective Hamiltonian and the problem of shape for its states.

F. The general microscopic collective Hamiltonian and the shape problem

In the previous subsection, we took the three-particle *oscillator* Hamiltonian and projected out the part with $p_\alpha = 0$ to get the corresponding microscopic collective Hamiltonian.

We could have started with an arbitrary particle Hamiltonian H of the type (3.5) and projected out from it[10] the *general* collective microscopic Hamiltonian. Because the eigenstates of H are characterized by the irreps $\{f\}$ of the symmetric group S_A, the part $H - H_0$ of (3.5) can be replaced by

$$[A(A-1)/2]\{(\hbar\omega)^{-1}V[(2\hbar/M\omega)^{1/2}|\mathbf{x}^1|] - A^{-1}(\mathbf{x}^1)^2\}, \quad (3.83a)$$

as discussed in (3.13), where $|\mathbf{x}^1|$ is the magnitude of the first Jacobi vector. We can then project the collective parts[10] of $V[(2\hbar/M\omega)^{1/2}|\mathbf{x}^1|], (\mathbf{x}^1)^2$, by replacing $\mathbf{x}^1$ by its value (3.15) in terms of $\rho, \gamma, \vartheta, \alpha$, and averaging over the 2A-5 internal coordinates α.

We shall implement this program for V of the Gaussian type i.e.,

$$V = V_0\exp[-(\mathbf{x}^1)^2/b^2], \quad (3.83b)$$

where $b^2 = (M\omega/2\hbar)a^2$, with a being the range in normal units of the Gaussian potential. From (3.25) and the above discussion we see that the collective part of the potential V_C is[26]

$$\begin{aligned} V_C &= V_0(2\pi)^{-1} \\ &\times \int_0^{2\pi} \exp[-(\rho^2/b^2)(\cos^2\gamma\cos^2\alpha + \sin^2\gamma\sin^2\alpha)]d\alpha \\ &= V_0(2\pi)^{-1}\exp(-\rho^2/2b^2) \\ &\times \int_0^{2\pi} \exp[-(\rho^2/2b^2)\cos2\gamma\cos2\alpha]d\alpha \\ &= V_0\exp(-\rho^2/2b^2)I_0[(\rho^2/2b^2)\cos2\gamma], \end{aligned} \quad (3.83c)$$

where I_0 is a Bessel function (p. 958 of Ref. 15). Note that this formula is valid only within certain restrictions for its domain as discussed by Katkevicius and Vanagas.[26]

Similarly the collective part of $(\mathbf{x}^1)^2$ is $(\rho^2/2b^2)$ and thus for $A = 3$ and Gaussian potentials the collective Hamiltonian takes the form

$$H_{GC} = H_C + 3\{(V_0/\hbar\omega)\exp(-r/b^2) \times I_0(r\sin\theta/b^2) - (r/Ab^2)\}, \quad (3.84)$$

where H_C is given by (3.55) and we prefer using the coordinates r, θ, φ related to ρ, γ, ϑ by the point transformation (3.54). The projection from an A-partcle problem gives a similar result.[26]

The Hamiltonian H_{GC} still commutes with the angular momentum, which for this two-dimensional problem is $J_3 = L_3 = -i\partial/\partial\varphi$, and thus we can factor $\exp(im\varphi)$ from its eigenstates, but it is no longer separable in r, θ. We can though calculate, with the techniques discussed in the Appendix, the matrix of elements

$$\langle n'j'm|H_{GC}|njm\rangle, \quad (3.85)$$

where $|njm\rangle$ is given by (3.80) and, restricting ourselves to a maximum number of quanta $0 \leqslant n \leqslant N$, from the diagonalization of this finite matrix obtain the eigenstates $|im\rangle$, $i = 1, 2, \ldots$, of H_{GC} as

$$|im\rangle = \sum_{\substack{nj \\ n-j \text{ even}}} a_{inj}(m)|njm\rangle. \quad (3.86)$$

We now turn our attention to the problem of shape of states of this generalized collective Hamiltonian. The discussion following (3.23) indicates that the shape would be given by the expectation value with respect to the states $|im\rangle$ of

$$\rho^4 \cos^2 2\gamma = 4r^2 \sin^2\theta. \tag{3.87}$$

Thus from (3.86), we require the matrix elements of

$$\langle n'j'm|r^2 \sin^2\theta\,|njm\rangle, \tag{3.88}$$

which can be evaluated in parabolic coordinates by the procedure indicated in the Appendix. We also note that from (3.80) and (3.67) we can write

$$|njm\rangle = \sum_{l=0}^{n} A_{jl}(nm)|nlm\rangle, \tag{3.89}$$

where

$$A_{jl}(nm) = \sum_{\sigma}\{(-1)^{(n/2)-\sigma+m}\langle (n/2)\sigma, (n/2)m-\sigma|jm\rangle \times \langle (n/2)\sigma, (n/2)m-\sigma|lm\rangle\}. \tag{3.90}$$

Since the matrix element $\langle n'l'm|r^2\sin^2\theta\,|nlm\rangle$ can be determined directly,[14,25] we can finally write

$$\begin{aligned}\langle n'j'm|r^2\sin^2\theta\,|njm\rangle = {} & \sum_{l'=0}^{n'}\sum_{l=0}^{n}\{A_{j'l'}(n'm)A_{jl}(nm)\tfrac{2}{3}\left[\delta_{l'l} - \left(\frac{2l+1}{2l'+1}\right)^{\frac{1}{2}}\langle lm, 20|l'm\rangle\langle l0, 20|l'0\rangle\right] \\ & \times \tfrac{1}{8}\left[\frac{(n-l)!(n'-l')!}{(n+l+1)!(n'+l'+1)!}\right]^{\frac{1}{2}}(-1)^{n+n'}(l'-l+2)!(l-l'+2)! \\ & \times \sum_{t}(l+l'+3+t)![t!(n-l-t)!(n'-l'-t)!(l'-n+2+t)!(l-n'+2+t)!]^{-1}\},\end{aligned} \tag{3.91}$$

where we made use of the volume element (A2) of the Appendix and of the generating functions[15] of the Laguerre polynomials.

IV. CANONICAL TRANSFORMATIONS RELATING THE MACROSCOPIC AND MICROSCOPIC COLLECTIVE MODELS

In Sec. II, we saw that the two-dimensional Bohr–Mottelson model can be generalized to a σ-δ interacting boson model that contains the Hamiltonians of both the vibrational and rotational limits. This model is associated with the U(3) symmetry of the three-dimensional oscillator and the states correspond to the irrep N of this group. They can be further characterized by the irrep λ of the $O(3)$ subgroup of U(3) which gives the different rotational bands. Finally the irreps m of the subgroup $O(2)$ of $O(3)$, are associated with the angular momentum in our two-dimensional space. Thus, we get the states $\langle \mathbf{R}|N\lambda m\rangle$ of (2.28), where we indicate explicitly the coordinate $\mathbf{R}$ (of spherical components R, Θ, Φ) on which they depend. They are eigenstates of $\hat{N}$, $\mathscr{L}^2$, $\mathscr{L}_0$ in (2.26) with eigenvalues N, $\lambda(\lambda+1)$, m, respectively.

We turn now to the Hamiltonian of a system of particles interacting through harmonic oscillator forces. The states then are characterized by the total number of quanta $\mathscr{N}$, the irreps j and m, respectively, of an $S\mathscr{U}(2)$ group and its subgroup $\mathscr{O}(2)$, as well as on irreps of other subgroups as indicated in (3.12). With the help of the collective coordinates introduced by Dzublik *et al.*[11] and by Zickendraht,[11] we extracted the collective part of these states, which turns out to depend only on even $\mathscr{N} = 2n$, integer j, such that $n-j$ even and $m = j, j-1, \ldots, -j$. These states, given in (3.80) or (3.89), are denoted by $\langle r|njm\rangle$, where we explicitly indicate the coordinate r (of spherical components r, θ, φ). They are eigenstates of the pseudo-Coulomb Hamiltonian $(H_C/2)$ of (3.55), as well as of J^2, J_3, of (3.74), with eigenvalues $n+1, j(j+1)$ and m, respectively.

As both $\langle \mathbf{R}|N\lambda m\rangle$ and $\langle \mathbf{r}|njm\rangle$ are a complete set of states, with the same range of values for the quantum numbers appearing in them, we could define a unitary transformation from one of these sets of states to the other by

$$\langle \mathbf{R}|U|\mathbf{r}\rangle = \sum_{\substack{n \\ n-j \text{ even}}}\sum_{\substack{j \\ j \text{ integral}}}\sum_{m=-j}^{j}\langle \mathbf{R}|njm\rangle\langle njm|\mathbf{r}\rangle, \tag{4.1}$$

where we took $N = n, \lambda = j$ and $\langle njm|\mathbf{r}\rangle$ as the complex conjugate of $\langle \mathbf{r}|njm\rangle$.

This unitary transformation is a representation of a canonical transformation.[27] It guarantees, among others, that the classical observables H_0, $\mathscr{L}^2$, $\mathscr{L}_3$ of (2.26), which are functions of $\mathbf{R}, \mathbf{P}$, transform into $(H_C/2)$, J^2, J_3 of (3.55) and (3.74) which are functions of $\mathbf{r}, \mathbf{p}$, where $\mathbf{P}, \mathbf{p}$ are, respectively, the canonically conjugate variables to $\mathbf{R}, \mathbf{r}$. The explicit form of this canonical transformation i.e., $\mathbf{R}, \mathbf{P}$ as function of $\mathbf{r}, \mathbf{p}$ or vice versa[27] will be discussed elsewhere by Moshinsky and Seligman. For our present purposes all we need is the unitary representation (4.1) as given any *operator* F in $\mathbf{R}$, $\mathbf{P} = -i\nabla_{\mathbf{R}}$ we can transform it into the corresponding *operator* f in $\mathbf{r}$, $\mathbf{p} = -i\nabla_{\mathbf{r}}$ or vice versa. This transformation is most easily achieved if, for example, we calculate the matrix elements of F with respect to the states $|N\lambda m\rangle$ of (2.28). Then from (4.1) the matrix elements of f with respect to the states $|njm\rangle$ of (3.80) or (3.89) must have the same value when $n = N, \lambda = j$. As in quantum mechanics we can either work in the operator or matrix representation, the possibility of mapping the latter on each other in the macroscopic and microscopic collective models, provides then the correspondence between them that we are searching for.

As an example we note that the general Hamiltonian H of (2.33) of the macroscopic collective model, can be translated into the microscopic collective model. We note that $Q^2 = \mathscr{L}_0^2 - \mathscr{L}^2$ appearing in (2.33) actually becomes $J_3^2 - J^2 = A_1^2 + A_2^2$ with the A's given by (3.60). For the n_δ of (2.33) we have its matrix element (2.34) and if we replace $N = n, \lambda = j$ we have the corresponding matrix elements with respect to $|njm\rangle$. Thus we can write in the microscopic collective model the matrix associated with the combination $(1-x)n_\delta + xQ^2$, $0 \leqslant x \leqslant 1$, which includes the vibrational

and rotational limits as well as the transitional region in between.

More interestingly we could start with a realistic two-body interaction in the Hamiltonian of an A-body system and project out its microscopic collective part H_{GC}, as we illustrated in (3.84) for the $A = 3$ case with a Gaussian interaction. Then we can calculate the matrix elements of H_{GC} with respect to the states $|njm\rangle$ as indicated in (3.85). Thus when we replace n by N, j by λ, we have the Hamiltonian matrix in a σ-δ interacting boson approximation model, corresponding to the microscopic two-body interaction that was considered.

Another interesting point concerns the shape problem. In the macroscopic collective model we saw that the deformation of a state can be estimated from the expectation value of $R^2 \sin^2\Theta$ with respect to it. On the other hand, in the microscopic case, we saw from (3.86)–(3.91) that the deformation of the state can be estimated from the expectation value of $r^2 \sin^2\theta$ with respect to it. But the canonical transformation whose unitary representation is (4.1) *does not* map $r^2\sin^2\Theta$ on $r^2 \sin^2\theta$. In fact the matrix elements $\langle N'\lambda' m|R^2 \sin^2\Theta\,|N\lambda m\rangle$ given by (2.36) (suitably generalized to $N' \neq N$ through the analysis in p. 247 of Ref. 13) are different from those of $\langle n'j'm|r^2 \sin^2\theta\,|njm\rangle$ given by (3.91). Thus the shape problem in collective models has different connotations when we approach it from the macroscopic or the microscopic end. With other collaborators, we shall explore this problem further in future publications.

We have given a thorough discussion of the macroscopic and microscopic collective models, and of their interrelations, for the two-dimensional case. The interesting problem though is in three dimensions where the mathematical analysis is much more difficult. Yet conceptually the problem is very similar and in the next section we proceed to outline it, leaving for a later article the detailed implementation of some of the steps.

V. OUTLINE OF THE GENERALIZATION TO THREE DIMENSIONS

We wish now to outline the main steps required to implement for the three-dimensional case the analysis given in this paper for the two-dimensionsal problem.

A. The macroscopic collective model

As in Sec. II we start with the Bohr–Mottelson model, where now the surface of the drop is given by[6,13]

$$R = R_0\Big[\,+\,1\sum_m \alpha_m Y^*_{2m}(\Theta, \Phi)\Big], \tag{5.1}$$

where here R, Θ, Φ are the spherical coordinates in the physical three-dimensional case. The collective motion in this macroscopic model are then characterized by the five α_m, $m = 2, 1, 0, -1, -2$ and their corresponding momenta are $\pi_m = -i\partial/\partial\alpha^m$. For small vibrations the Hamiltonian is then an oscillator in these variables.[6,13,28]

As in Sec. II we pass from the frame of reference fixed in space, in which the coordinates are the α_m, to those fixed in the body i.e., β, γ and the Euler angles θ_i, $i = 1, 2, 3$ which now instead of (2.3) are related by [6,13]

$$\alpha_m = [D^{2*}_{m2}(\theta_i) + D^{2*}_{m-2}(\theta_i)](1/\sqrt{2})\beta \sin\gamma + D^{2*}_{m0}(\theta_i)\beta\cos\gamma. \tag{5.2}$$

The eigenstates of the oscillator Hamiltonian, corresponding to those given by (2.5) in the two-dimensional case, were fully determined recently[28] and can be written as

$$\langle\beta\gamma\theta_i|\nu\Lambda tLM\rangle. \tag{5.3}$$

The quantum numbers are related with the irreps of the chain of groups of the Bohr–Mottelson vibrational Hamiltonian, i.e.,[28]

$$\underset{\nu}{U(5)} \supset \underset{\Lambda}{O(5)} \supset \underset{L}{O(3)} \supset \underset{M}{O(2)}, \tag{5.4}$$

and t distinguishes between repeated irreps L of $O(3)$ in a given irrep Λ of $O(5)$.

The kets (5.3) represent only vibrational states. We can get though rotational or transitional ones by considering higher-order terms in the Bohr–Mottelson Hamiltonian, as in Sec. B of II for the two-dimensional case, getting then what has been called the Frankfurt model.[8] It is more relevant though, for our final objective, to achieve this purpose by extending the Hilbert space, as was done in Sec. C of II for the two-dimensional case. Thus we can add[16] an s coordinate to the five d coordinates α_m getting a six-dimensional oscillator whose symmetry group is U(6). In this s-d boson model one chain of subgroups is (5.4) leading to the state

$$\langle\bar{\alpha}\beta\gamma\theta_i|N\nu\Lambda tLM\rangle, \tag{5.5}$$

in which N is the irrep of U(6). This state can be obtained in a trivial way[16] from (5.3) and it corresponds to the $\langle\bar{\alpha}\beta\vartheta\,|Nvm\rangle$ of (2.25) for the two-dimensional case.

There are though other chains of subgroups of U(6) and in particular,

$$\underset{N}{U(6)} \supset \underset{[2e_1,2e_2,2e_3]}{U(3)} \supset \underset{L}{O(3)} \supset \underset{M}{O(2)} \tag{5.6}$$

will be very relevant for the objective of finding the unitary representation of the canonical transformation that relates the three-dimensional macroscopic and microscopic collective models.

In (5.6) we have indicated the irreps of the different groups in the chain. For a given single-rowed partition $[N]$ that characterizes the irrep of U(6), the irreps of the subgroup U(3) are given by three even numbers $2e_i$, $i = 1, 2, 3$, with the property that[10,16]

$$e_1 + e_2 + e_3 = N \quad e_1 \geqslant e_2 \geqslant e_3 \geqslant 0, \tag{5.7}$$

The irreps L of $O(3)$ contained in a given irrep $[2e_1, 2e_2, 2e_3]$ of U(3) are given by the inequalities discussed in p. 52 of Ref. 18.

The states associated with the chain of groups (5.6) can then be written as

$$\langle\bar{\alpha}\beta\gamma\theta_i|[2e_1, 2e_2, 2e_3]\Omega LM\rangle, \tag{5.8}$$

where Ω distinguishes between the repeated irreps L of $O(3)$ contained in a given irrep of U(3). These states corrspond to $\langle R\Theta\Phi\,|N\lambda m\rangle$ of (2.28) of the two-dimensional case and they are associated with rotational properties.[16] The ket (5.8) is not presently explicitly available, but as we showed in Ref. 16 how to calculate the matrix elements of the Casimir operator of U(3) in the basis (5.5), we can, by diagonalizing a finite

matrix, obtain the coefficients of the expansion of (5.8) in terms of (5.5). Thus, at least in principle, we have the rotational states and we could also get transitional ones by considering, in the basis (5.5), Hamiltonians equivalent to (2.33).

Having then discussed a macroscopic collective model of the *s-d* interacting boson type, we turn our attention to the microscopic collective model in three-dimensional space.

B. The microscopic collective model

As in Sec. A of III we start with an A body oscillator Hamiltonian from which we want to project out the collective part. As indicated at the end of Sec. B of III, this can be done from any A so long as it is larger than the dimension of our space.[10] Thus in the three-dimensional case the smallest A we can take is $A = 4$ for which we have the three Jacobi vactors x_i^s, $s = 1, 2, 3$ of components $i = 1, 2, 3$. The collective coordinates are then introduced through a transformation of the type (3.15) which for three dimensions and $A = 4$ becomes[10]

$$x_i^s = \sum_{k=1}^{3} \rho_k D^1_{ki}(\vartheta_j) D^1_{ks}(\alpha_j); \quad i,j,k = 1,2,3, \tag{5.9}$$

where $\|D^1_{ki}\|$ are the standard rotation matrices [e.g. as in p. 95 of Ref. 13] in terms of the ordinary Euler angles ϑ_j or those of the angles in particle index space α_j. The ρ_k^2's give the principal moments of inertia as discussed in Sec. 3 of III.

All of the analysis of Sec. A of III applies to the three-dimensional case if i,j take values 1, 2, 3 rather than only 1, 2. Thus the states of the four-body problem in three dimensions can be characterized by the irrep $\mathcal{N}$ of the $\mathbf{U}(3A-3) = \mathbf{U}(9)$ group and its subgroups

$$\begin{array}{ccccccc} \mathbf{U}(9) & \supset & \mathscr{U}(3) & \times & U(3) & [h_1h_2h_3] & \\ & & \cup & & \cup & & \\ & & \mathscr{O}(3) & L & O(3) & \bar{L} & , \\ & & \cup & & \cup & & \\ & & \mathscr{O}(2) & M & D^{[3,1]}(S_4) & \{f\} & \end{array} \tag{5.10}$$

where we indicate to the right of the groups the corresponding irreps. We recall that, as in (3.11), $[h_1h_2h_3]$ is the irrep of both $\mathscr{U}(3)$ and U(3) and besides[18]

$$h_1 + h_2 + h_3 = \mathcal{N}. \tag{5.11}$$

The eigenstates of the oscillator Hamiltonian, corresponding to (3.12) in the two-dimensional case, can then be denoted by

$$\langle \rho_i \vartheta_i \alpha_i | [h_1h_2h_3] \Omega LM, \bar{\Omega} \bar{L} K \{f\}(r) \rangle, \tag{5.12}$$

where $\Omega, \bar{\Omega}$ distinguish between repeated irreps $L, \bar{L}$ of $\mathscr{O}(3)$, $O(3)$ in a given irrep $[h_1h_2h_3]$ of $\mathscr{U}(3)$, U(3), while K does the same[20] for the repeated irrep $\{f\}$ of S_4 in a given $\bar{L}$ of $O(3)$.

The states (5.12) can be determined as polynomials in the η_i^s, $s = 1, 2, 3$; $i = 1, 2, 3$ acting on the ground state by procedures[18] similar to those that lead to the $|\mathcal{N} j m \mu \{f\}(r)\rangle$ of (3.48). We do not have though at present an explicit expression of these states as functions $\rho_i \vartheta_i \alpha_i$.

The microscopic collective states can be projected from (5.12) in a procedure similar to the one used in the two-dimensional case to get (3.81) from (3.48). In particular, this implies considering those symmetric states (5.12) that are scalars of $O(3)$ i.e., in which $\{f\} = \{4\}$ and $\bar{L} = 0$, which are obviously independent of the angular coordinates α_i associated with the particle indices $s = 1, 2, 3$. We note that for $\bar{L}$ to vanish the inequalities in p. 52 of Ref. 18 indicate that $h_1 - h_3$, $h_2 - h_3$ must be even and, as in this case $\bar{L} = 0$ appears only once,[18] the $\bar{\Omega}$ index is irrelevant. Furthermore for the state to be symmetric, i.e., $\{f\} = 4$, when $\bar{L} = 0$ we require that[20] $\mathcal{N}$ must be even i.e., $\mathcal{N} = 2N$ and thus from (5.11) we conclude that the states (5.12) with $\bar{L} = 0$, $\{f\} = \{4\}$ must correspond to a partition

$$[h_1h_2h_3] = [2e_1, 2e_2, 2e_3]; \quad \mathcal{N} = 2N, \tag{5.13}$$

where the e_i's satisfy (5.7).

The collective state can then be denoted by

$$\langle \rho_i, \vartheta_i | [2e_1, 2e_2, 2e_3] \Omega LM \rangle, \tag{5.14}$$

and as the quantum numbers have the same range as in the macroscopic collective state (5.8) (when we write $\mathcal{N} = 2N$) we conclude that for a fixed N the number of states (5.14) corresponds to the degeneracy of the representation $[N]$ of U(6). Besides as (5.14) are the particular oscillator states (5.12) for which $\bar{L} = 0$, $\{f\} = \{4\}$, the energy levels to which (5.14) are associated continue to be equally spaced. Thus we see that both the macroscopic and microscopic collective Hamiltonians are related with the same chain of subgroups of U(6) despite the fact that the latter were derived from states characterized by the irreps of the chain of subgroups (5.10) of U(9).

C. Canonical transformations relating macroscopic and microscopic collective models

From the analysis of Sec. IV and the discussion in this section we immediately conclude that the unitary representation of the canonical transformations relating the macroscopic and microscopic collective models can be written as

$$\begin{aligned} &\langle \bar{\alpha}\beta\gamma\theta_i | U | \rho_1\rho_2\rho_3\vartheta_i \rangle \\ &\quad = \sum_{e_1e_2e_3} \sum_{\Omega LM} \langle \bar{\alpha}\beta\gamma\theta_i | [2e_1, 2e_2, 2e_3] \Omega LM \rangle \\ &\qquad \times \langle [2e_1, 2e_2, 2e_3] \Omega LM | \rho_1\rho_2\rho_3\vartheta_i \rangle, \end{aligned} \tag{5.15}$$

where care should be taken with the domain of the variables as discussed in Chap. 11 in the last reference of Ref. 10.

We indicated in the previous subsection that the states (5.8) of the macroscopic collective model that appear in (5.15), can be obtained as linear combinations of the states (5.5) where an explicit expression is available for the latter.[28] The states (5.14) of the microscopic collective model also appearing in (5.15) are more difficult to obtain and we proceed to outline a possible, though not necessarily practical, way of determining them.

To begin with the oscillator Hamiltonian H_0 of (3.4) when $A = 4$, $i = 1, 2, 3$ can be written in terms of the collective coordinates ρ_i, ϑ_i, α_i of (5.9), as was done in the 1969 paper of Zickendraht.[11] If in H_0 we equate to zero the angular momentum vector $\bar{\mathbf{L}}$ associated with the coordinates α_i, $i = 1, 2, 3$ we get a Hamiltonian H_C of which the kets (5.14) are eigenstates. We need though that (5.14) should also be eigenstate of integrals of motion associated with the $\mathscr{U}(3)$ group and its subgroups. Following the analysis given in Sec. A of III, we have for, the generators of $\mathscr{U}(3)$,

$$\mathscr{C}_{ij} = \sum_{s=1}^{3} \eta_i^s \xi_j^s, \quad i,j = 1,2,3, \tag{5.16}$$

with the definition (3.6) for creation and annihilation operators. We then write $\mathscr{C}_{ij}$ in terms of the collective coordinates $\rho_i, \vartheta_i, \alpha_i$ to obtain results equivalent to the generators K_1, K_2, K_3 of $S\mathscr{U}(2)$ for the two-dimensional case that are given in (3.36). Finally, we equate in $\mathscr{C}_{ij}$ the $\overline{L}$ to 0, to obtain $\mathscr{C}^C_{ij}$ the generators of $\mathscr{U}(3)$ dependent only on ρ_i, ϑ_i and their derivatives which are equivalent to the generators J_1, J_2, J_3 of $S\mathscr{U}(2)$ for the two-dimensional case given in (3.78). Thus finally, the kets (5.14) are eigenstates of the operators

$$H_C, \quad G = \sum_{i,j=1}^{3} \mathscr{C}^C_{ij}\mathscr{C}^C_{ji}, \quad L^2, \quad L_3, \tag{5.17}$$

where G is the Casimir operator[18] of the $\mathscr{U}(3)$ group.

We plan to implement in another publication the steps given in the previous paragraph, so as to obtain an explicit expression for the states (5.14) and thus also for the unitary transformation that takes us from the macroscopic to the microscopic collective model.

D. Shape of states in the three-dimensional case

We discussed extensively in the previous sections the deformation of the states for the two-dimensional case. This problem is of course much more interesting in the real three-dimensional case. We proceed to outline the steps that allow us to discuss the problem of shape both in the macroscopic and microscopic collective models.

We consider first the macroscopic collective model of Bohr–Mottelson and replace in Eq. (5.1) for the surface of the liquid drop the α_m by (5.2) and Θ, Φ by Θ', Φ' where the latter are angular coordinates in the frame of reference fixed in the body. The deformations R_1, R_2, R_3 along the three principal axis fixed in the body correspond then to the angles $\Theta', \Phi' = \pi/2, 0; \pi/2, \pi/2; 0, 0$, and we get[13]

$$R_k = R_0\{1 + (5/4\pi)^{\frac{1}{2}}\beta \cos[\gamma - (2\pi k/3)]\}, \quad k = 1, 2, 3. \tag{5.18}$$

Thus clearly a measure of the deformation along the three principal axis can be obtained from the expectation values of $\beta \cos[\gamma - (2\pi k/3)]$ with respect to the states of the macroscopic collective Hamiltonian. These states can be expressed as linear combinations of (5.5) or, equivalently, of (5.8).

However, it is more convenient to consider some functions of $\beta \cos[\gamma - (2\pi k/3)]$, $k = 1, 2, 3$ for the expectation values rather than these expressions themselves. We notice, for example, that we have the following relation:

$$\{x - \beta\cos[\gamma - (2\pi/3)]\}\{x - \beta\cos[\gamma - (4\pi/3)]\} \times \{x - \beta\cos\gamma\} = x^3 - \tfrac{3}{4}x\beta^2 - \tfrac{1}{4}\beta^3\cos 3\gamma, \tag{5.19}$$

and thus the cubic equation resulting from equating the right-hand side to zero has precisely the roots $x_k = \beta\cos[\gamma - (2\pi k/3)]$, $k = 1, 2, 3$.

We note that[28]

$$\beta^2 = 5^{\frac{1}{2}}[\alpha\times\alpha]^0_0,$$

$$\beta^3\cos 3\gamma = -(35/2)^{\frac{1}{2}}[[\alpha\times\alpha]^2\times\alpha]^0_0, \tag{5.20}$$

where the square brackets indicate angular momentum couplings. Then $\beta^2, \beta^3\cos 3\gamma$ are polynomials in the α's and their matrix elements with respect to the states (5.5) can be obtained[16,28] with the help of the isoscalar factors for the $O(5)\supset O(3)$ chain of groups. Thus we can discuss the deformation of states in the three-dimensional macroscopic collective model, corresponding to (2.35) in the two-dimensional case.

Turning our attention now to the microscopic collective model, we note from the discussion in Sec. C of III that ρ_1, ρ_2, ρ_3 themselves measure the deformation along the three principal axis. We can use then, for example, the expectation values of $\rho_1^2, \rho_2^2, \rho_3^2$ with respect to the eigenstates of a microscopic collective Hamiltonian H_{GC} to get an estimation of the shape of these states. Again though it may be more convenient[29] to write $\rho_1^2, \rho_2^2, \rho_3^2$ in terms of ρ, b, δ through the relations

$$\rho_k^2 = (\rho^2/3)\{1 + 2b\cos[\delta - (2\pi k/3)]\}, \tag{5.21}$$

from which it follows that $\rho^2 = \rho_1^2 + \rho_2^2 + \rho_3^2$.

Clearly then, by a similar analysis to that of the macroscopic case, rather than the expectation values of ρ_k^2 we could take those of

$$\rho^2, \quad \rho^4 b^2, \quad \rho^6 b^3\cos 3\delta, \tag{5.22}$$

and, solving a cubic equation, get the deformation parameters. As the eigenstates of H_{GC} can be expanded in terms of the complete set of eigenstates (5.14) of H_C, we require then the matrix elements of the operators (5.22) with respect to the states (5.14). This is also a problem we intend to tackle in a future publication once we obtain the states (5.14) explicitly.

As indicated in Sec. IV, the shape problem in collective models has different connotations when approached from the macroscopic or the microscopic end. This is due to the fact that operators such as $\beta^2, \beta^3\cos 3\gamma$ do not transform into $\rho^4 b^2, \rho^6 b^3\cos 3\delta$ under the unitary operator (5.15). Thus, as mentioned at the end of Sec. IV, we intend, with other collaborators, to confront the derformations obtained for states in the macroscopic and microscopic collective models.

E. Flow diagrams relating the nuclear models

As a last point, we summarize the relations between the nuclear models discussed in this paper by a kind of flow diagram given in Fig. 1. We consider first the different oscillator Hamiltonians and their interconnections and then extend the discussion to general Hamiltonians.

We start with the A particle oscillator Hamiltonian—which is the three-dimensional generalization of (3.1)—and denote it by the circle containing A in Fig. 1. We then project from it the microscopic collective (MC) Hamiltonian by the procedure outlined in this section getting then the three-dimensional generalization of H_C of (3.53). This Hamiltonian is indicated by MC in a circle and, as we obtain it from projection from the particle oscillator Hamiltonian, we connect A and MC by a line going from the first to the second.

We then consider an oscillator boson approximation (OBA) Hamiltonian containing s and d bosons and the discussion in this section indicates that MC and OBA (where the latter also appears in a circle in the figure) are related by a unitary representation of a canonical transformation given

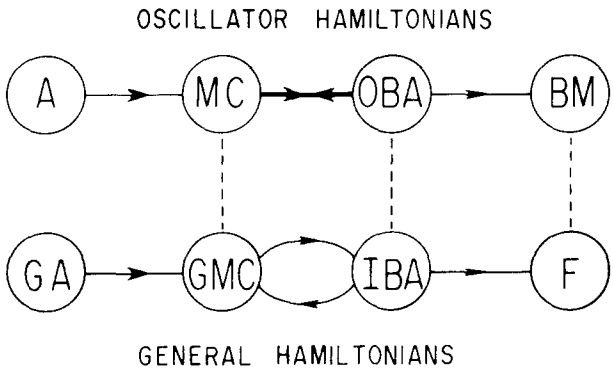


FIG. 1. The flow diagram is explained in Sec. E of V. We indicate here only the meaning of the acronymics of the Hamiltonians appearing in the circles: A-A-nucleon oscillator; MC-Microscopic Collective; OBA-Oscillator (*s-d*) boson approximation; BM-Bohr—Mottelson oscillator; GA-General A nucleon problem; GMC-General microscopic collective; IBA-Interacting (*s-d*) boson approximation; F-Generalized Bohr–Mottelson model as developed by the Frankfurt group.

in (5.15). This fact is shown in the figure by the heavy line connecting MC and OBA with arrows in both directions, as the canonical transformation allows us to go from the MC to the OBA or vice versa. Finally, by projecting out the *d*-boson part of the OBA we get a Hamiltonian which is identical to the oscillator Hamiltonian of Bohr–Mottelson (BM) as indicated by the last circle of the upper line of Fig. 1.

We turn now to the general A particle (GA) Hamiltonian which is the three-dimensional equivalent of (3.5). By a procedure similar to the one discussed in Sec. E of II for the two-dimensional case, we can project the general microscopic collective (GMC) Hamiltonian corresponding to (3.84). This GMC Hamiltonian can be expressed as a matrix with respect to the eigenstates (5.14) of the oscillator microscopic collective (MC) Hamiltonian, which is indicated in Fig. 1 by the dotted connection between the two models.

Turning now to the macroscopic *s-d* interacting boson approximation (IBA) analyzed in Ref. 9, we can express its most general Hamiltonian as a matrix with respect to the eigenstates (5.8) of the oscillator *s-d* boson approximation (OBA) as discussed in Ref. 16. Thus, through the canonical transformation relating MC and OBA, we can pass from GMC to a kind of IBA or vice versa as indicated by the two lines with arrows in the diagram. Finally, as discussed in Ref. 8, there is a procedure by which we can relate IBA Hamiltonians with those of a generalized Bohr–Mottelson type which we designate as Frankfurt (F) Hamiltonians as most work on them[7] was done there. These Hamiltonians can be expressed as matrices with respect to the eigenstates (5.3) of the oscillator Bohr–Mottelson (BM) Hamiltonian, which is indicated by the dotted line connecting BM and F.

Thus we have outlined a procedure by which starting from the Hamiltonian of a microscopic system of A nucleons, we can arrive finally to a macroscopic collective model which is a suitable generalization of the one introduced by Bohr and Mottelson.

APPENDIX: THE PSEUDO-COULOMB HAMILTONIAN AND ITS EIGENSTATES

In this Appendix we review some of the well-known[23] solutions in parabolic coordinates of the Hamiltonian H_C of (3.55)—which we call pseudo-Coulomb[14]—and then derive from them the eigenstates of H_C, J^2, J_3, where the last two operators are given by (3.74).

In terms of the spherical coordinates r, θ, φ defined in (3.54) the parabolic ones u, v, ϕ are given by

$$u = r(1 - \cos\theta), \quad v = r(1 + \cos\theta), \quad \phi = \varphi. \tag{A1}$$

To derive the volume element in these coordinates we note that from the determinant of the matrix (3.27), relating the Jacobi vectors x_i^s, $i = 1, 2$, $s = 1, 2$ with $\rho, \gamma, \vartheta, \alpha$, we have

$$\begin{aligned} dV &= \rho^3 \cos 2\gamma d\rho d\gamma d\vartheta d\alpha \\ &= (r/2)\sin\theta dr d\theta d\varphi d\alpha = \tfrac{1}{4} du dv d\phi d\alpha, \\ &0 \leqslant u, v \leqslant \infty, \quad 0 \leqslant \phi, \alpha \leqslant 2\pi, \end{aligned} \tag{A2}$$

where we made use of the transformations (3.54) and (A1). From the definitions (3.55), (3.59), and (3.60) and the commutation rules (3.61), we have that the operators H_C, A_3, L_3 commute among themselves and the last two are given by

$$\begin{aligned} A_3 &= -\frac{\partial}{\partial x_3} r \frac{\partial}{\partial r} + \tfrac{1}{2} x_3 (\nabla^2 + 1), \\ L_3 &= \frac{1}{i}\left(x_1 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_1}\right). \end{aligned} \tag{A3}$$

From (3.69) and (A1) we see that the common eigenstate ψ of these three operators satisfies, in parabolic coordinates, the equations[23]

$$\begin{aligned} \tfrac{1}{2} H_C \psi &= \left\{\left[-\frac{\partial}{\partial u} u \frac{\partial}{\partial u} - \frac{1}{4u}\frac{\partial^2}{\partial \phi^2} + \frac{u}{4}\right] \right. \\ &\quad \left. + \left[-\frac{\partial}{\partial v} v \frac{\partial}{\partial v} - \frac{1}{4v}\frac{\partial^2}{\partial \phi^2} + \frac{v}{4}\right]\right\}\psi \\ &= (n+1)\psi, \end{aligned} \tag{A4a}$$

$$\begin{aligned} A_3\psi &= \left\{ -\left[-\frac{\partial}{\partial u} u \frac{\partial}{\partial u} - \frac{1}{4u}\frac{\partial^2}{\partial \phi^2} + \frac{u}{4}\right] \right. \\ &\quad \left. + \left[-\frac{\partial}{\partial v} v \frac{\partial}{\partial v} - \frac{1}{4v}\frac{\partial^2}{\partial \phi^2} + \frac{v}{4}\right]\right\}\psi = a\psi, \end{aligned} \tag{A4b}$$

$$L_3\psi = -i\frac{\partial\psi}{\partial\phi} = m\psi, \tag{A4c}$$

where we denote by $n+1, a, m$, respectively, the eigenvalues of $H_C/2, A_3, L_3$. If instead of A_3, L_3, we consider the operators $M_3 = (L_3 + A_3)/2$, $N_3 = (L_3 - A_3)/2$, whose eigenvalues we denote by σ, τ, the ψ is also an eigenstate of them in which $m = \sigma + \tau$, $a = \sigma - \tau$.

The explicit form of the eigenstate ψ of (A4) is then the ket

$$\begin{aligned} |(n/2)\sigma\tau\rangle &= (-1)^{(n/2)-\sigma}(uv)^{-\frac{1}{4}}(2\pi)^{-\frac{1}{2}}\exp[i(\sigma+\tau)\phi] \\ &\times \begin{cases} F^{|\sigma+\tau|/2}_{(n/2)-\sigma}(u/2) F^{|\sigma+\tau|/2}_{(n/2)-\tau}(v/2), & \sigma+\tau \geqslant 0 \\ F^{|\sigma+\tau|/2}_{(n/2)+\tau}(u/2) F^{|\sigma+\tau|/2}_{(n/2)+\sigma}(v/2), & \sigma+\tau \leqslant 0, \end{cases} \end{aligned} \tag{A5}$$

where the factor $(-1)^{(n/2)-\sigma}$ is required[23] for the Condon and Shortley[22] phase convention and the function $F_n^{m/2}(\rho)$, $m \geqslant 0$ of the variable ρ satisfies the equation[14]

$$\rho\left[-\frac{d^2}{d\rho^2}+\frac{[(m/2)^2-(1/4)]}{\rho^2}+1\right]F_n^{m/2}(\rho) = (2n+m+1)F_n^{m/2}(\rho), \tag{A6}$$

so that its normalized form with respect to the volume element (A2) becomes[14]

$$F_n^{m/2}(\rho) = 2^{(m+1)/2}[(n!)/\Gamma(n+m+1)]^{\frac{1}{2}}\rho^{(m+1)/2} e^{-\rho}L_n^m(2\rho),\ m \geqslant 0, \tag{A7}$$

where L_n^m is a Laguerre polynomial.

The kets (A5) are then eigenkets of the operators $M^2 = N^2$, M_3, N_3, with eigenvalues $(n/2)[(n/2)+1]$, σ, τ, respectively. Thus with respect to them, the matrix elements of $L_\pm = L_1 \pm iL_2$, L_3; $A_\pm = A_1 \pm iA_2$, A_3 become

$$\langle (n/2)\sigma'\tau'|L_\pm|(n/2)\sigma\tau\rangle = \{[(n/2)\mp\sigma][(n/2)\pm\sigma+1]\}^{\frac{1}{2}}\delta_{\sigma'\sigma\pm1}\delta_{\tau'\tau} + \{[(n/2)\mp\tau][(n/2)\pm\tau+1]\}^{\frac{1}{2}}\delta_{\sigma'\sigma}\delta_{\tau'\tau\pm1}, \tag{A8a}$$

$$\langle (n/2)\sigma'\tau'|L_3|(n/2)\sigma\tau\rangle = (\sigma+\tau)\delta_{\sigma'\sigma}\delta_{\tau'\tau}, \tag{A8b}$$

$$\langle (n/2)\sigma'\tau'|A_\pm|(n/2)\sigma\tau\rangle = \{[(n/2)\mp\sigma][(n/2)\pm\sigma+1]\}^{\frac{1}{2}}\delta_{\sigma'\sigma\pm1}\delta_{\tau'\tau} - \{[(n/2)\mp\tau][(n/2)\pm\tau+1]\}^{\frac{1}{2}}\delta_{\sigma'\sigma}\delta_{\tau'\tau\pm1}, \tag{A8c}$$

$$\langle (n/2)\sigma'\tau'|A_3|(n/2)\sigma\tau\rangle = (\sigma-\tau)\delta_{\sigma'\sigma}\delta_{\tau'\tau}. \tag{A8d}$$

From the fact that $\mathbf{L} = \mathbf{M} + \mathbf{N}$ it is clear that the eigenstate $|nlm\rangle$ of H_C, L^2, L_3 is given by

$$|nlm\rangle = \sum_\sigma \langle (n/2)\sigma, (n/2)m-\sigma|lm\rangle\,|(n/2)\sigma m-\sigma\rangle, \tag{A9}$$

where $\langle|\rangle$ is a Clebsch–Gordan coefficient. Then applying $L_\pm$ to the left- and right-hand sides of (A9) and making use of (A8a), we see that the Clebsch–Gordan coefficients must satisfy the recursion relations

$$[(l\mp m)(l\pm m+1)]\langle (n/2)\sigma, (n/2)m\pm1-\sigma|lm\pm1\rangle = \{[(n/2)\pm\sigma][(n/2)\mp\sigma+1]\}^{\frac{1}{2}} \times\langle (n/2)\sigma\mp1, (n/2)m-\sigma\pm1|lm\rangle + \{[(n/2)\mp m\pm\sigma][(n/2)\pm m\mp\sigma+1]\}^{\frac{1}{2}} \times\langle (n/2)\sigma, (n/2)m-\sigma|lm\rangle, \tag{A10}$$

as is actually the case.[25]

We now wish to obtain the eigenstates $|njm\rangle$ of H_c, J^2, J_3, where the last two are given by (3.74). These eigenstates will have the property that

$$J_\pm|njm\rangle = A_\pm|njm\rangle = [(j\mp m)(j\pm m+1)]^{\frac{1}{2}}|njm\pm1\rangle. \tag{A11}$$

If we propose now that

$$|njm\rangle = \sum_\sigma (-1)^{(n/2)-\sigma+m}\langle (n/2)\sigma, (n/2)m-\sigma|jm\rangle\,|(n/2)\sigma m-\sigma\rangle, \tag{A12}$$

then applying $A_\pm$ to the left- and right-hand sides of $|njm\rangle$ and using (A8c) rather than (A8a) we arrive at precisely the same recurrence relation (A10) for the Clebsch–Gordan coefficients. Thus the states $|njm\rangle$ are explicitly determined through (A12) and (A5).

As a final point, we determine the matrix elements of

$$r^2\sin^2\theta = uv, \tag{A13}$$

with respect to the states $|njm\rangle$. As indicated in (3.88), this matrix element is relevant to the discussion of the shape problem in the microscopic collective model.

From the equation relating Laguerre polynomials (Ref. 15, p. 1037)

$$\rho L_\nu^m(\rho) = -(\nu+1)L_{\nu+1}^m(\rho) + (2\nu+m+1)L_\nu^m(\rho) - (\nu+m)L_{\nu-1}^m(\rho), \tag{A14}$$

we immediately conclude that

$$\int_0^\infty F_{\nu'}^{m/2}(\rho)F_\nu^{m/2}(\rho)\,d\rho = -[(\nu+m+1)(\nu+1)]^{\frac{1}{2}}\delta_{\nu'\nu+1} + (2\nu+m+1)\delta_{\nu'\nu} - [\nu(\nu+m)]^{\frac{1}{2}}\delta_{\nu'\nu-1}. \tag{A15}$$

Taking into account then the explicit expression of $|njm\rangle$ given by (A12) and (A5) we obtain

$$\begin{aligned}\langle n'j'm'|r^2\sin^2\theta\,|njm\rangle &= \sum_{\sigma,\sigma'}[\langle (n'/2)\sigma', (n'/2)m-\sigma'|j'm\rangle\langle (n/2)\sigma, (n/2)m-\sigma|jm\rangle \\ &\times\tfrac{1}{4}\{-\delta_{(n'/2)-\sigma',(n/2)-\sigma+1}[(n/2)-\sigma+1]^{\frac{1}{2}}[(n/2)-\sigma+m+1]^{\frac{1}{2}} + \delta_{(n'/2)-\sigma',(n/2)-\sigma}[n-2\sigma+m+1] \\ &-\delta_{(n'/2)-\sigma',(n/2)-\sigma-1}[(n/2)-\sigma]^{\frac{1}{2}}[(n/2)-\sigma+m]^{\frac{1}{2}}\} \\ &\times\{-\delta_{(n'/2)+\sigma',(n/2)+\sigma+1}[(n/2)-m+\sigma+1]^{\frac{1}{2}}[(n/2)+\sigma+1]^{\frac{1}{2}} + \delta_{(n'/2)+\sigma',(n/2)+\sigma}[n+2\sigma-m+1] \\ &-\delta_{(n'/2)+\sigma',(n/2)+\sigma-1}[(n/2)-m+\sigma]^{\frac{1}{2}}[(n/2)+\sigma]^{\frac{1}{2}}\}].\end{aligned} \tag{A16}$$

[1] W. Elsasser, J. Phys. Radium **5**, 625 (1934).
[2] M. G. Mayer, Phys. Rev. **75**, 1969 (1949); **78**, 16 (1956).
[3] M. G. Mayer and J. H. D. Jensen, *Elementary Theory of Nuclear Shell Structure* (Wiley, New York, 1955).
[4] A. de Shalit and H. Feshbach, *Theoretical Nuclear Physics*, Vol. I Chapter VII (Wiley, New York, 1974).

[5]N. Bohr, Nature **137**, 344 (1936); N. Bohr and F. Kalckar, Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd. **14**, 10 (1937).
[6]A. Bohr and B. Mottelson, Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd. **27**, 16 (153).
[7]G. Gneuss and W. Greiner, Nuc. Phys. A **171**, 449 (1971); R. Sedlmayer *et al.*, Nucl. Phys. A **232**, 465 (1971).
[8]M. Moshinsky, Nucl. Phys. A **338**, 156 (1980).
[9]A. Arima and F. Iachello, Ann. Phys. (N. Y.) **99**, 253 (1976); **111**, 201 (1978); A. Arima, F. Iachello and O. Scholten, Ann. Phys. (N. Y.) **115**, 325 (1978); R. V. Dzholos, F. Dönau, and D. Janssen, Yad. Fiz. **22**, 965 (1975) [Sov. J. Nuclear Phys. **22**, 503 (1976)].
[10]V. Vanagas, E. Nadjakov, P. Raychev, Preprint IC/75/140, Trieste 1976 and Bulg. J. Phys. **11**, 6 (1975); V. Vanagas, "The microscopic nuclear theory within the framework of the restricted dynamics", *Lecture Notes in Physics* (University of Toronto, Toronto, 1977).
[11]W. Zickendraht, J. Math. Phys. **10**, 30 (1969); **12**, 1663 (1971); A. Ya. Dzublik V. I. Ovcharenko, A. I. Steshenko and G. F. Filippov, Yad. Fiz. **15**, 869 (1972); [Sov. J. Nuc. Phys. **15**, 487 (1972)].
[12]M. Bouten and P. van Leuven, Am. J. Phys. **45**, 455 (1977).
[13]J. M. Eisenberg and W. Greiner, *Nuclear Models* (North–Holland, Amsterdam, 1970).
[14]M. Moshinsky, T. H. Seligman and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[15]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[16]O. Castaños, E. Chacón, A. Frank and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[17]M. Moshinsky, *The Harmonic Oscillator in Modern Physics; From Atoms to Quarks* (Gordon & Breach, New York, 1969).
[18]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon & Breach, New York, 1967).
[19]M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, Massachussets, 1962).
[20]P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).
[21]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, Massachussetts, 1959).
[22]E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U. P., (1953), p. 72.
[23]M. J. Englefield *Group Theory and the Coulomb Problem* (Wiley, New York, 1972).
[24]M. Moshinsky, Phys. Rev. **126**, 1880 (1962).
[25]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[26]O. Katkevicius and V. Vanagas, Liet. fiz. rinkinyis **21**, (1980) (to be published).
[27]P. A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1975); M. Moshinsky and T. H. Seligman, Ann. Phys. **114**, 243 (1978); **120**, 402 (1979).
[28]E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 1668 (1976); E. Chacón and M. Moshinsky, *ibid.* **18**, 870 (1977).
[29]Yu. F. Smirnov and G. F. Filippov, Yad. Fiz. **27**, 73 (1978); [Sov. J. Nucl. Phys. **27**, 39 (1978)].

Bianchi type IX cosmological models with homogeneous spinor fields

Octavio Obregón
Department of Physics, Universidad Autónoma Metropolitana–Iztapalapa, México 13, D. F., México

Michael P. Ryan, Jr.[a)]
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, C. U., México 20, D. F., México

(Received 16 July 1980; accepted for publication 5 September 1980)

The diagonal and symmetric Bianchi type IX models are coupled to a homogeneous spinor field. An action for the combined fields is constructed, where the orthonormal basis used is given explicitly in terms of the metric. This allows one to vary the action with respect to the metric and the spinor fields only. Next, a Hamiltonian formulation is given, and a qualitative solution for the problem is presented. We also show that the $k = +1$ FRW (Friedmann–Robertson–Walker) model is not compatible with a homogeneous spinor field, while the more complicated models are.

PACS numbers: 98.80.Dr

I. INTRODUCTION

Recently Isham and Nelson[1] have studied Friedmann–Robertson–Walker (FRW) models with matter given by a homogeneous spinor field in both the classical and quantum regimes. One of the most discouraging results of their work was that the $k = +1$ models do not admit such a homogeneous spinor field. Since the $k = +1$ models are special cases of the Bianchi type IX models, more general type IX models could be compatible with homogeneous spinor fields.

The reason for the failure of the $k = +1$ models is that G_{0i} is automatically zero for these models, while T_{0i} is not. The constraints $T_{0i} = 0$ are so restrictive that the Hamiltonian for the spinor field vanishes, leaving only the vacuum part of the gravitational Hamiltonian. It is well known that there are no vacuum solutions for $k = +1$ FRW models, so the solution breaks down. However, the more complicated models all admit vacuum solutions, and we will show that the diagonal and symmetric[2] (or nontumbling[3]) type IX models with a homogeneous spinor field exist. In fact, the symmetric type IX models have $G_{0i} \neq 0$, which allows an even richer solution structure.

This paper has three aims: (i) to demonstrate the compatibility of diagonal and symmetric type IX models with a homogeneous spinor field; (ii) to give a Hamiltonian formulation for spinor fields in homogeneous models; and (iii) to present a qualitative solution[2] of the combined field equations in the diagonal and symmetric cases.

We begin by writing the Einstein–Dirac action

$$I = \int [\sqrt{-g}R + 16\pi \mathcal{L}_M]\, d^4x, \tag{1.1}$$

where R is the scalar curvature and $\mathcal{L}_M$ is the Dirac Lagrangian density in a curved background.[4] In the usual formulation, $\mathcal{L}_M$ is a function of the spinor field ψ and $\hat{h}^{(\rho)}{}_{\alpha}$, an orthonormal basis which satisfies

$$\hat{h}_{(\rho)\mu}\hat{h}^{(\rho)}{}_{\nu} = g_{\mu\nu}, \quad \hat{h}_{(\alpha)\mu}\hat{h}^{(\beta)\mu} = \delta^{\alpha}{}_{\beta}, \tag{1.2}$$

and their derivatives. In order to connect this density to R, which is a function of $g_{\mu\nu}$ and its derivatives, we can do one of two things; we can write R as a function of $\hat{h}^{(\rho)}{}_{\alpha}$ and vary with respect to $\hat{h}^{(\rho)}{}_{\alpha}$ to obtain the gravitational equations, or we can explicitly give $\hat{h}^{(\rho)}{}_{\alpha}$ as a function of $g_{\mu\nu}$ and vary with respect to $g_{\mu\nu}$. We may construct a Hamiltonian formulation of the problem for either one of these two approaches. For an example of the first approach see Nelson and Teitelboim.[5] We will concentrate on the second method.

One problem with the second approach is that it is usual to prescribe a certain variation of $\hat{h}^{(\rho)}{}_{\alpha}$ with respect to $g_{\mu\nu}$ in order to achieve a canonical form for the energy–momentum tensor. However, if we choose a basis for other reasons, the energy–momentum tensor may not take the usual form. In fact, the bases usually chosen for Bianchi type models, when varied do not always satisfy the relations that lead to the canonical energy–momentum tensor. This problem is discussed in the Appendix.

In Sec. II we develop a Hamiltonian formulation of the Einstein–Dirac system, then apply it to Bianchi type IX cosmological models. In Sec. III, we discuss the equations of motion, demonstrating the incompatibility of $k = +1$ FRW model with a homogeneous spinor field, and the fact that other type IX models do admit such fields. Section IV is devoted to qualitative solutions of the field equations for the cases where such a solution exists. The last section includes conclusions and suggestions for further work.

II. AN ACTION FOR HOMOGENEOUS SPINOR FIELDS IN BIANCHI-TYPE COSMOLOGICAL MODELS

We begin by considering the Einstein–Dirac action ($\hbar = 1, c = 1$) in the form

$$I = \frac{1}{16\pi}\int [\pi^{ij}\dot{g}_{ij} - N\mathcal{H}_{\perp} - N_i\mathcal{H}^i + 16\pi\mathcal{L}_M]\, d^4x, \tag{2.1}$$

with

$$\mathcal{L}_M = \sqrt{-g}\{\tfrac{1}{2}i(\bar{\psi}\gamma^{\mu}\psi_{;\mu} - \bar{\psi}_{;\mu}\gamma_{\mu}\psi) - m\bar{\psi}\psi\}, \tag{2.2}$$

where Greek indices run from 0 to 3 and latin indices run from 1 to 3. The dot means derivative with respect to $x^0 = t$,

[a)]Also, Asesor, Department of Physics, Universidad Autónoma Metropolitana-Iztapalapa, México 13, D. F., México.

 022-2488/81/030623-08$1.00

the ordinary partial derivative will be designated by $(,\mu)$ and the covariant vector or spinor derivative by $(;\mu)$.

The first three terms in the action (2.1) correspond to the Einstein Lagrangian $I = \int R\,(\sqrt{\ } - g)\,d^4x$ reparametrized according to the Arnowitt–Deser-Misner (ADM) formulation,[6] by introducing quantities

$$N = (-{}^4g^{00})^{-1/2}, \quad N_i = {}^4g_{0i},$$
$$g_{ij} = {}^4g_{ij}, \quad \pi^{ij}, \quad \mathscr{H}_\perp, \quad \text{and} \quad \mathscr{H}^i. \tag{2.3}$$

Here the superscript 4 denotes a four-dimensional geometrical object, and the superscript 3 will be used for objects on selected three-dimensional hypersurfaces. The π^{ij} are defined in terms of the ${}^4g_{\mu\nu}$ and the ${}^4\Gamma^\alpha_{\mu\nu}$, while $\mathscr{H}_\perp$ and $\mathscr{H}_i$ are algebraic combinations for the π^{ij}, the g_{ij}, and their derivatives.

In the matter Lagrangian $\mathscr{L}_M$, $g = \det{}^4g_{\mu\nu}$, $\bar\psi = \psi^\dagger\gamma^0$ and γ^μ are the generalized Dirac matrices defined by

$$\{\gamma^\mu,\gamma^\nu\} = -2g^{\mu\nu}, \tag{2.4}$$

with signature $(-,+,+,+)$ for the metric tensor $g_{\mu\nu}$.

The covariant spinorial derivative of the Dirac 4-spinor is given by

$$\psi_{;\mu} = \psi_{,\mu} - \Gamma_\mu\psi, \tag{2.5}$$

with the connection coefficients

$$\Gamma_\mu = -\tfrac{1}{4}\hat h_{(\rho)\alpha;\mu}\hat h^{(\rho)}{}_\beta\gamma^\beta\gamma_\alpha. \tag{2.6}$$

The vierbein basis $\hat h_{(\rho)\mu}$ satisfies

$$\hat h_{(\rho)\mu}\hat h^{(\rho)}{}_\nu = {}^4g_{\mu\nu}, \quad \hat h_{(\alpha)\mu}\hat h^{(\beta)\mu} = \delta^\beta{}_\alpha,$$
$$\hat h_{(\rho)\mu} = \eta_{(\rho\alpha)}\hat h^{(\alpha)}{}_\mu, \tag{2.7}$$

so that the generalized Dirac matrices are related to the standard matrices of special relativity $\gamma^{(\alpha)}$ by

$$\gamma^\mu = \hat h_\mu{}^{(\alpha)}\gamma_{(\alpha)}. \tag{2.8}$$

For the standard γ matrices the following representation is used:

$$\gamma^{(0)} = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad \gamma^{(k)} = \begin{bmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{bmatrix}, \tag{2.9}$$

with

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},$$
$$\sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$
$$\gamma^{(5)} = i\gamma^{(0)}\gamma^{(1)}\gamma^{(2)}\gamma^{(3)} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}. \tag{2.10}$$

By definition we also have

$$[\gamma^\mu,\gamma^\nu] = -2i\sigma^{\mu\nu}. \tag{2.11}$$

Our first step is to break up[6] the matter Lagrangian in (2.1) into terms such as $p\,dq$, $N\mathscr{L}^0{}_M$ and $N_i\mathscr{L}^i{}_M$. We will obtain such a Lagrangian density for the Class A Bianchi-type cosmological models.

Since the matter Lagrangian $\mathscr{L}_M$ is a function of $g_{\mu\nu}$ mainly through the basis $\hat h_{(\alpha)\mu}$, we must choose a suitable basis. It has been usual[7] to write the metric of the Bianchi-type models as

$$ds^2 = -(N^2 - N_iN^i)\,dt^2 + 2N_i\,dt\,\omega^i + e^{-2\Omega(t)}e^{2\beta(t)}{}_{ij}\omega^i\omega^j, \tag{2.12}$$

where $\Omega(t)$ is a scalar, β_{ij} is a 3×3 matrix, $N = N(t)$, and $N_i = N_i(t)$. The forms ω^i are invariant one-forms appropriate to the model under consideration and obey

$$d\omega^i = -\tfrac{1}{2}C^i{}_{jk}\omega^j\wedge\omega^k, \tag{2.13}$$

where the $C^i{}_{jk}$ are group structure constants of the particular Bianchi type under consideration. For type IX models, $C^i{}_{jk} = -\epsilon_{ijk}$.

A convenient orthonormal basis[8] is

$$\boldsymbol\sigma^{(0)} = N\,dt, \quad \boldsymbol\sigma^{(i)} = N^i\,dt + \hat h^{(i)}{}_l\,dx^l, \tag{2.14}$$

with $\hat h^{(i)}{}_k$ an orthonormal triad $[\hat h^{(i)}{}_l\hat h^{(j)}{}_k\delta_{ij} = g_{lk}$, $\hat h_{(i)}{}^l\hat h_{(j)l} = \delta_{ij}]$. Here g_{lk} is the metric on $t = $ const surfaces. In order to define our orthonormal triad we take a basis of one forms $\boldsymbol\sigma^i = e^{-\Omega}e^\beta{}_{ij}\omega^j$.

To calculate the Lagrangian $\mathscr{L}_M$ we need to evaluate expressions of the type $\Gamma_{(\mu)}\gamma^{(\mu)}$, $\Gamma_{(\mu)} \equiv \hat h_{(\mu)}{}^\alpha\Gamma_\alpha$. These terms turn out to be functions of the $\gamma^{(\mu)}$ themselves and $C^{(\mu)}{}_{(\alpha\beta)}$, the structure coefficients in the equation

$$d\boldsymbol\sigma^{(\mu)} = -\tfrac{1}{2}C^{(\mu)}{}_{(\alpha\beta)}\boldsymbol\sigma^{(\alpha)}\wedge\boldsymbol\sigma^{(\beta)}. \tag{2.15}$$

The $C^{(\mu)}{}_{(\alpha\beta)}$ for the basis (2.14), if $N = N(t)$ and $N_i = N_i(t)$, are

$$C^{(0)}{}_{(j0)} = 0,$$
$$C^{(l)}{}_{(j0)} = \frac{1}{N}(-\dot{\hat h}_{(j)}{}^k\hat h_{(l)k} + N^r\hat h_{(j)}{}^k{}_{,r}\hat h^{(l)}{}_k). \tag{2.16}$$
$$C^{(l)}{}_{(jk)} = e^\Omega e^\beta{}_{lm}e^{-\beta}{}_{jp}e^{-\beta}{}_{ki}C^m{}_{pi},$$

the $C^m{}_{pi}$ are those that appear in Eq. (2.13). If we now calculate $\Gamma_{(\mu)}$ using Eq. (2.6) and insert the results in $\mathscr{L}_M$, we obtain for the Class A Bianchi-type cosmological models

$$\begin{aligned}\mathscr{L}_M = {}& \tfrac{1}{2}ie^{-3\Omega}\bar\psi\gamma^{(0)}\dot\psi - \tfrac{1}{2}ie^{-3\Omega}\dot{\bar\psi}\gamma^{(0)}\psi \\ & + \tfrac{1}{4}ie^{-3\Omega}\bar\psi(e^\beta{}_{ik})\dot{e}^{-\beta}{}_{kj}\sigma^{(ij)}\gamma^{(0)}\psi \\ & + Ne^{-3\Omega}[\tfrac{1}{8}e^{2\Omega}\bar\psi e^\beta{}_{mr}e^{-\beta}{}_{is} \\ & \times e^{-\beta}{}_{jt}C^r{}_{st}\gamma^{(m)}\sigma^{(ij)}\psi - m\bar\psi\psi] \\ & + N^i[\tfrac{1}{4}e^{-3\Omega}\bar\psi e^{-\beta}{}_{jr}e^\beta{}_{lt}C^t{}_{ir}\gamma^{(l)}\sigma^{(j0)}\psi].\end{aligned} \tag{2.17}$$

Variation of this Lagrangian density with respect to the field variables ψ and $\bar\psi$ gives the generally covariant Dirac equations. Variation with respect to the metric tensor $g^{\mu\nu}$ must be performed using the tetrad field we have chosen. This variation is in general not unique and our vierbein basis does not fulfill the generally assumed relation

$$\frac{\delta\hat h_{(\rho)}{}^\alpha}{\delta g_{\mu\nu}} = -\tfrac{1}{2}\hat h_{(\rho)}{}^\mu g^{\nu)\alpha}, \tag{2.18}$$

by means of which the "standard" symmetric energy–momentum tensor of the Dirac field is normally obtained. In view of this fact our energy–momentum tensor differs from the "standard" one (see Appendix).

In this paper we will study Bianchi type IX models which can be divided into three cases: the FRW $k = +1$ case ($\beta_{ij} = 0$), the diagonal case $\beta_{ij} = \beta_{dij}$ $= \operatorname{diag}(\beta_+ + (\sqrt{3})\beta_-, \beta_+ - (\sqrt{3})\beta_-, -2\beta_+)$, and the symmetric case where $\beta_{ij} = e^{-\phi\kappa_i}\beta_d e^{\phi\kappa_j}$,

$$\kappa_3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We will construct an action for the symmetric case, which we can always specialize later to the other two, putting $\phi = 0$ and $\beta_{dij} = 0$ respectively.

We can now insert the metric (2.12) with $\beta_{ij} = e^{-\phi\kappa_3}\beta_d e^{\phi\kappa_3}$ in the action (2.1), and substitute $\mathscr{L}_M$ given by Eq. (2.17). We define

$$p_{ij} = (e^{\beta}{}_{is}\pi^s{}_t e^{-\beta}{}_{tj} - \tfrac{1}{3}\delta_{ij}\pi^l{}_l), \tag{2.19}$$

with

$$6p_{ij} = e^{-\phi\kappa_3}\{\alpha_1 p_+ + \alpha_2 p_- + \alpha_3 \frac{3p_\phi}{\sinh[2(\sqrt{3})\beta_-]}\} e^{\phi\kappa_3}, \tag{2.20}$$

$\alpha_1 = \mathrm{diag}(1,1,-2), \alpha_2 = \mathrm{diag}\ (\sqrt{3}, -\sqrt{3}, 0)$,

$$\alpha_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Using

$$p'_\phi = p_\phi + 8\pi^{-3\Omega}\bar{\psi}\sigma^{(12)}\gamma^{(0)}\psi\sinh^2[(\sqrt{3})\beta_-], \tag{2.21}$$

we find

$$I = \frac{1}{8\pi}\int [p_+\dot{\beta}_+ + p_-\dot{\beta}_- + p'_\phi\dot{\phi} - \dot{\Omega}\pi^k{}_k + 4\pi i e^{-3\Omega}\bar{\psi}\gamma^{(0)}\dot{\psi} - 4\pi i e^{-3\Omega}\dot{\bar{\psi}}\gamma^{(0)}\psi - \tfrac{1}{2}N\mathscr{H}_\perp - \tfrac{1}{2}N^i\mathscr{H}_i]\, dt\, d\omega. \tag{2.22}$$

Here $d\omega \equiv \omega^1\wedge\omega^2\wedge\omega^3$, and

$$\mathscr{H}_\perp = \tfrac{1}{2}e^{2\Omega}(1-V) + e^{6\Omega}\left(\tfrac{1}{6}(\pi^k{}_k)^2 - \tfrac{1}{6}p_+^2 - \tfrac{1}{6}p_-^2 - \frac{\frac{1}{2}p'^2_\phi}{\sinh^2(2\sqrt{3}\beta_-)} + 2\pi\frac{e^{-3\Omega}\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi p'_\phi}{\cosh^2(\sqrt{3}\beta_-)} - 8\pi^2 e^{-6\Omega}\tanh^2(\sqrt{3}\beta_-)\times(\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi)^2\right) - 8\pi e^{\Omega}[e^{2\beta_+}\cosh(2\sqrt{3}\beta_-) + \tfrac{1}{2}e^{-4\beta_+}]\times\bar{\psi}\gamma^{(0)}\gamma^{(5)}\psi - 16\pi m\bar{\psi}\psi,$$

$$\mathscr{H}_1 = 4\pi\cosh(3\beta_+)\sin2\phi\sinh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(2)}\gamma^{(5)}\psi - 4\pi\sinh(3\beta_+)\cosh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(1)}\gamma^{(5)}\psi + 4\pi\cosh(3\beta_+)\cos2\phi\sinh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(1)}\gamma^{(5)}\psi,$$

$$\mathscr{H}_2 = 4\pi\cosh(3\beta_+)\cosh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(2)}\gamma^{(5)}\psi + 4\pi\sinh(3\beta_+)\cos2\phi\sinh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(2)}\gamma^{(5)}\psi - 4\pi\sinh(3\beta_+)\sin2\phi\sinh(\sqrt{3}\beta_-)\bar{\psi}\gamma^{(1)}\gamma^{(5)}\psi,$$

$$\mathscr{H}_3 = -p'_\phi + 4\pi e^{-3\Omega}\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi,$$

where

$$V = 1 + \tfrac{2}{3}e^{4\beta_+}[\cosh(4\sqrt{3}\beta_-) - 1] + \tfrac{1}{3}e^{-8\beta_+} - \tfrac{4}{3}e^{-2\beta_+}\cosh(2\sqrt{3}\beta_-).$$

III. EQUATIONS OF MOTION

Given the action (2.22), we only have to vary it with respect to $\beta_\pm, p_\pm, p'_\phi, \phi, N_i, N, \psi$, and $\bar{\psi}$ to get all the equations of motion. To obtain the final system, we have to choose N and N_i and also solve the constraint equations $\mathscr{H}_\mu = 0$.

For convenience, we choose $N_i = 0$ and $N = 6e^{-3\Omega}/\pi^k{}_k$ [this second choice is equivalent to taking $\Omega(t) = t$]. We will solve the constraint $\mathscr{H}_\perp = 0$ for $\pi^k{}_k$ and define $H = \pi^k{}_k$. It is not difficult to show that we may solve the constraint $\mathscr{H}_3 = 0$ for p'_ϕ in terms of $\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi$, thus eliminating ϕ from the problem (except as a quadrature). Substituting this value for p'_ϕ into $\mathscr{H}_\perp$ and solving for H as a function of $\beta_\pm, p_\pm, \Omega, \bar{\psi}$, and ψ, the action for $\beta_\pm$ becomes

$$I = \int [p_+\, d\beta_+ + p_-\, d\beta_- - H\, d\Omega\,], \tag{3.1a}$$

where

$$H^2 = p_+^2 + p_-^2 + 48\pi^2 e^{-6\Omega}(\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi)^2\times\coth^2(2\sqrt{3}\beta_-) + 48\pi e^{-5\Omega}V_S(\beta_\pm)\bar{\psi}\gamma^{(0)}\gamma^{(5)}\psi + 96\pi m e^{-6\Omega}\bar{\psi}\psi + 9e^{-4\Omega}(V(\beta_\pm) - 1). \tag{3.1b}$$

The spinor potential $V_S(\beta_\pm)$ is

$$V_S(\beta_\pm) = e^{2\beta_+}\cosh(2\sqrt{3}\beta_-) + \tfrac{1}{2}e^{-4\beta_+}. \tag{3.2}$$

Note that in this Hamiltonian form we have dropped the velocity terms in ψ and $\bar{\psi}$. This is because the action (3.1) with these terms added, when varied with respect to ψ or $\bar{\psi}$ does not give the Dirac equations. To see the reason for this, we have to return to the action (2.22) and vary with respect to ψ and $\bar{\psi}$, and then impose the conditions $N_i = 0$ and the constraint $\mathscr{H}_3 = 0$. Unfortunately, the $N_i = 0$ coordinate choice is incompatible with the action (3.1) with the velocity terms in ψ and $\bar{\psi}$ added. In fact, it is easy to see that with $N = 6e^{-3\Omega}/H$, the equations for ψ and $\bar{\psi}$ become ($\cdot \equiv d/d\Omega$)

$$\dot{\psi} = \tfrac{1}{2}\psi - \frac{3\pi i}{H}e^{-3\Omega}(\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi)\times\gamma^{(0)}\gamma^{(3)}\gamma^{(5)}\psi[1 + \tanh^2(\sqrt{3}\beta_-)] - \frac{3i}{H}V_S(\beta_\pm)e^{-2\Omega}\gamma^{(5)}\psi - \frac{6im}{H}e^{-3\Omega}\gamma^{(0)}\psi, \tag{3.3a}$$

$$\dot{\bar{\psi}} = \tfrac{1}{2}\bar{\psi} + \frac{3\pi i}{H}e^{-3\Omega}(\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi)\times\bar{\psi}\gamma^{(3)}\gamma^{(5)}\gamma^{(0)}[1 + \tanh^2(\sqrt{3}\beta_-)] - \frac{3i}{H}V_S(\beta_\pm)e^{-2\Omega}\bar{\psi}\gamma^{(5)} - \frac{6im}{H}e^{-3\Omega}\bar{\psi}\gamma^{(0)}. \tag{3.3b}$$

Actually the difference between these equations and those obtained by varying (3.1) with the appropriate spinor terms consists only in replacing $1 + \tanh^2(\sqrt{3}\beta_-)$ by $2\coth^2(2\sqrt{3}\beta_-)$ in (3.3).

Now, varying (2.22) with respect to N_1 and N_2 gives us the following constraints on the spinor variables (excluding impossible conditions on the metric variables):

$$\bar{\psi}\gamma^{(2)}\gamma^{(5)}\psi = \bar{\psi}\gamma^{(1)}\gamma^{(5)}\psi = 0. \tag{3.4}$$

In order to study the structure of Eqs. (3.3) and (3.4) we take a spinor basis in which

$$\psi = \begin{pmatrix} a_1 e^{i\theta_1} \\ a_2 e^{i\theta_2} \\ b_1 e^{i\phi_1} \\ b_2 e^{i\phi_2} \end{pmatrix}. \tag{3.5}$$

The constraints (3.4) tell us that $b_2 = -a_1a_2/b_1$ and that

$\phi_2 = \theta_2 - \theta_1 + \phi_1$. The content of Eqs. (3.3) now reduces to

$$\dot{a}_1 - \tfrac{1}{2}a_1 - \frac{3V_S(\beta_\pm)}{H} e^{-2\Omega} b_1 \sin(\phi_1 - \theta_1) = 0, \tag{3.6a}$$

$$\dot{\theta}_1 - \frac{3\pi}{H} e^{-3\Omega} \frac{(a_1^2 + b_1^2)(a_2^2 - b_1^2)}{b_1^2} [1 + \tanh^2(\sqrt{3}\beta_-)] + \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} \cos(\phi_1 - \theta_1) \frac{b_1}{a_1} + \frac{6m}{H} e^{-3\Omega} = 0, \tag{3.6b}$$

$$\dot{a}_2 - \tfrac{1}{2}a_2 + \frac{3V_S(\beta_\pm)}{H} e^{-2\Omega} \frac{a_1 a_2}{b_1} \sin(\phi_1 - \theta_1) = 0, \tag{3.6c}$$

$$\dot{\theta}_2 + \frac{3\pi}{H} e^{3\Omega} \frac{(a_1^2 + b_1^2)(a_2^2 - b_1^2)}{b_1^2} [1 + \tanh^2(\sqrt{3}\beta_-)] - \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} \cos(\phi_1 - \theta_1) \frac{a_1}{b_1} + \frac{6m}{H} e^{-3\Omega} = 0, \tag{3.6d}$$

$$\dot{b}_1 - \tfrac{1}{2}b_1 + \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} a_1 \sin(\phi_1 - \theta_1) = 0, \tag{3.6e}$$

$$\dot{\phi}_1 - \frac{3\pi}{H} e^{-3\Omega} \frac{(a_1^2 + b_1^2)(a_2^2 - b_1^2)}{b_1^2} [1 + \tanh^2(\sqrt{3}\beta_-)] + \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} \frac{a_1}{b_1} \cos(\phi_1 - \theta_1) - \frac{6m}{H} e^{-3\Omega} = 0, \tag{3.6f}$$

$$\dot{\theta}_2 + \dot{\phi}_1 - \dot{\theta}_1 + \frac{3\pi}{H} e^{-3\Omega} \frac{(a_1^2 + b_1^2)(a_2^2 - b_1^2)}{b_1^2} \times [1 + \tanh^2(\sqrt{3}\beta_-)] - \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} \frac{b_1}{a_2} \cos(\phi_1 - \theta_1) - \frac{6m}{H} e^{-3\Omega} = 0, \tag{3.6g}$$

$$-\left(\frac{a_1 a_2}{b_1}\right)^{\cdot} + \tfrac{3}{2}\left(\frac{a_1 a_2}{b_1}\right) + \frac{3}{H} V_S(\beta_\pm) e^{-2\Omega} a_2 \sin(\phi_1 - \theta_1) = 0. \tag{3.6h}$$

It is not difficult to show that Eqs. (3.6g) and (3.6h) are consequences of the first six of Eqs. (3.6), which means that the constraints (3.4) are compatible with the Dirac equation. On the other hand, the addition of (3.6d) and (3.6f) gives

$$\phi_1 = -\theta_2 + \text{const}, \tag{3.7}$$

so in Eqs. (3.6), Eq. (3.6f) can be substituted by (3.7). Moreover, from (3.6f) and (3.6b) we can calculate $\dot{\phi}_1 - \dot{\theta}_1$. With $\gamma \equiv \phi_1 - \theta_1$, we get

$$\dot{\gamma} + \frac{3V_S(\beta_\pm)}{H} \cos\gamma \left[\frac{a_1}{b_1} - \frac{b_1}{a_1}\right] e^{-2\Omega} - \frac{12m}{H} e^{-3\Omega} = 0. \tag{3.8}$$

Now, Eqs. (3.6a), (3.6e), and (3.8) constitute a reduced set of equations whose integration, if possible, would allow the integration of all of Eqs. (3.6) because, knowing a_1, b_1, and γ as functions of Ω one could integrate (3.6c) and obtain a_2. With these four functions (3.6b) would be integrable and we would know θ_1, and then ϕ_1 could be calculated from γ and θ_1 and θ_2 from (3.7).

For the diagonal type IX model and the FRW case $\phi = p_\phi' = 0$ and we have the additional constraint

$$\bar{\psi}\gamma^{(3)}\gamma^{(5)}\psi = 0. \tag{3.9}$$

With ψ given by (3.5) this equation tell us that $a_2^2 = b_1^2$. The addition of this constraint to the system of Eqs. (3.6) makes Eqs. (3.6c) and (3.6e) equal and some terms zero. However, this system of equations is still consistent and could also be solved using the reduced set (3.6a), (3.6e), and (3.8).

For our three cases the symmetric, the diagonal and the FRW, the problem reduces to the action (2.22) which gives us $\beta_\pm$, the constraints (3.4) for the symmetric case and also the constraint (3.9) for the diagonal and the FRW cases, and the set of Eqs. (3.6) for ψ. The only equation we lack is one for H. Varying (2.22) with respect to Ω, and using $N = 6e^{-3\Omega}/H$, and Eqs. (3.3) to substitute for $\dot{\psi}$ and $\dot{\bar{\psi}}$, we find that

$$\dot{H} = -\frac{48\pi}{H} V_S(\beta_\pm) \bar{\psi}\gamma^{(0)}\gamma^{(5)}\psi e^{-5\Omega} - \frac{144\pi m}{H} e^{-6\Omega} \bar{\psi}\psi - \frac{18}{H} e^{-4\Omega} [V(\beta_\pm) - 1]. \tag{3.10}$$

Calculating $\bar{\psi}\gamma^{(0)}\gamma^{(5)}\psi$ and $\bar{\psi}\psi$, we see that they are proportional to $a_2^2 - b_1^2$. Thus, in the diagonal and FRW cases where we must apply the constraint (3.9), the spinor part of the Hamiltonian vanishes, and H as given by Eq. (3.1) reduces to the vacuum Hamiltonian. In the FRW case H becomes imaginary, so a homogeneous spinor field is incompatible with $k = +1$ FRW models, as was pointed out by Isham and Nelson. However, in the diagonal case vacuum solutions exist and are well known, so we will be able to give qualitative solutions for this case very simply. The vanishing of the spinor part of the Hamiltonian may imply that spinor fields satisfying the constraints (3.4) and (3.9) correspond to "ghost" spinor fields.[9]

In the next section we will show that for $\Omega \to \infty$ it is possible to find qualitative solutions in the sense of Ref. 2 for the diagonal and symmetric cases.

IV. QUALITATIVE SOLUTIONS

Now that we have the equations governing the behavior of $\beta_\pm$, H and ψ as functions of Ω, we can in principle proceed to solve them. Because of the highly nonlinear character of the equations, the possibility of finding exact solutions seems remote. There are several methods for obtaining approximate solutions for type IX models in the pure gravitational case[2,10-12] and these can be extended to models containing a spinor field. We will use the graphical technique of Ref. 2.

We begin by noting that the equations of motion are similar to those of a particle, the universe point, moving in an eight-dimensional space with coordinates $\beta_\pm$, a_1, a_2, b_1, θ_1, θ_2, and ϕ_1 for the symmetric case and a seven-dimensional space with $\beta_\pm$, a_1, b_1, θ_1, θ_2, and ϕ_1 as coordinates for the diagonal case. Here Ω plays the role of time. The analogy is not exact because the equations for the spinor variables are only first order. In the pure gravity case, it is possible to display the evolution of $\beta_\pm$ as a trajectory in the $\beta_+\beta_-$ plane, where the potential $V(\beta_\pm)$ governs the motion. Be-

cause of the first order nature of the spinor equations, such a potential description in the full eight- or (seven-) dimensional configuration space seems impossible. However, the behavior of the spinor variables is relatively simple, and it turns out to be possible to construct, at least near the singularity, a diagrammatic solution for $\beta_\pm$, where the spinor variables are given in terms of quadratures. We can begin by defining new spinor variables A_1, A_2, and B_1: $a_1 = A_1 e^{3\Omega/2}$, $a_2 = A_2 e^{3\Omega/2}$, $b_1 = B_1 e^{3\Omega/2}$; and δ and ρ by

$$\delta = 4(\sqrt{3})\pi(A_1^2 + B_1^2)(A_2^2 - B_1^2)/B_1^2,$$
$$\rho = 48\pi A_1(B_1^2 - A_2^2)\cos(\phi_1 - \theta_1)/B_1. \tag{4.1}$$

For the diagonal case $\rho = \delta = 0$. If we are interested in solutions near the singularity, Ω is very large, and if we assume that A_1, A_2, B_1, and H change relatively slowly (an assumption to be justified later), expressions of the type $e^{-n\Omega}F(A_1, A_2, B_1, H)$, where n is positive, rapidly become ignorable in the equations of motion.

If we drop such terms the equations of motion for A_1, A_2, B_1, θ_1, θ_2, and ϕ_1 can be obtained from (3.6), and in this case it is possible to reduce to quadratures the solutions of the reduced set of equations equivalent to Eqs. (3.6a), (3.6e), and (3.8). Following this procedure we get for A_1, B_1, and γ:

$$A_1 = \left[\frac{D}{2} - \tfrac{1}{4}(D^2 - 4C^2)^{1/2}\sin\left(-6\int \frac{V_S(\beta_\pm)e^{2\Omega}}{H}\,d\Omega\right)\right]^{1/2}, \tag{4.2a}$$

$$B_1 = \left[\frac{D}{2} + \tfrac{1}{4}(D^2 - 4C^2)^{1/2}\sin\left(-6\int \frac{V_S(\beta_\pm)e^{2\Omega}}{H}\,d\Omega\right)\right]^{1/2}, \tag{4.2b}$$

$$\gamma = \cos^{-1}\left(2C\left\{D^2\left[1 - \sin^2\left(-6\int \frac{V_S(\beta_\pm)e^{2\Omega}}{H}\,d\Omega\right)\right] + 4C^2\sin^2\left[-6\int \frac{V_S(\beta_\pm)e^{-2\Omega}}{H}\right]\right\}^{-1/2}\right), \tag{4.2c}$$

with C and D constants. Knowing $V_S(\beta_\pm)/H$ as a function of Ω we can, in principle, integrate these equations and the rest by the procedure described in the preceeding section and obtain the remaining four functions A_2, θ_1, ϕ_1, and θ_2.

Now, inserting δ and ρ in H^2, given in (3.1), we find (if we ignore exponentials in minus Ω) that

$$H^2 = P_+^2 + P_-^2 + \delta^2\coth^2[2(\sqrt{3})\beta_-] + 2\rho V_S(\beta_\pm)e^{-2\Omega} + 9e^{-4\Omega}[V(\beta_\pm) - 1]. \tag{4.3}$$

From Eqs. (3.3)–(3.5) and (4.1) it is possible to show (for the symmetric case) that δ is a constant of the motion and ρ is constant if terms of the form $e^{-3\Omega}/H$ can be neglected (that is, near the singularity). Since δ and ρ are independent of $\beta_\pm$, variations of H with respect to these functions, treating δ and ρ as constants, gives us the correct equations for $\beta_\pm$. Equation (4.3) can be used to calculate $dH/d\Omega$ and we find (ignoring, again, exponentials in minus Ω)

$$\dot{H} = -\frac{2\rho}{H}V_S(\beta_\pm)e^{-2\Omega} - \frac{18}{H}e^{-4\Omega}[V(\beta_\pm) - 1]. \tag{4.4}$$

If we now look at the potentials $V_S(\beta_\pm)$ and $V(\beta_\pm)$, we can justify our assumption that the exponential expressions we have been discarding are indeed ignorable. The form of $V(\beta_\pm)$ is well known, with roughly triangular equipotential curves and exponentially steep walls, with soft channels extending from each corner (see Fig. 1). The function $V_S(\beta_\pm)$ also has triangular equipotentials and exponential walls, but it has no channels at the corners (see Fig. 2). The minimum of $V(\beta_\pm) - 1$ is zero, and that of $V_S(\beta_\pm)$ is 3/2. For large Ω and $\beta_\pm$ such that $V(\beta_\pm) - 1$ and $V_S(\beta_\pm)$ are near their minima, $e^{-2\Omega}V_S(\beta_\pm)$ and $e^{-4\Omega}[V(\beta_\pm) - 1]$ are ignorable and A_1, A_2, B_1, and H are essentially constant, and will remain constant except during brief periods when the universe point is in contact with the exponential walls. Since any change in these quantities will be slow in comparison to the rapid decay of $e^{-n\Omega}$ we are justified in ignoring terms of the type $e^{-n\Omega}F(A_1, A_2, B_1, H)$ near a singularity.

Now, it is possible to use Eqs. (4.3) and (4.4) to calculate $\beta_\pm$ and H as functions of Ω, substitute these functions of Ω in Eqs. (4.2) and obtain all the spinor variables. The Hamiltonian (4.3) can be treated by means of the wall approximation; that is, the strong exponential behavior of all the potential terms allows us to replace them by a series of infinitely hard walls that, because of the explicit Ω dependence of these terms, move with changing Ω. The two triangular potentials, $V(\beta_\pm)$ and $V_S(\beta_\pm)$ give us triangular walls, the *gravitation* and *spinor* walls, respectively, and the triangular symmetry allows us to calculate the velocity of these walls when the universe point is not touching them by looking only at the

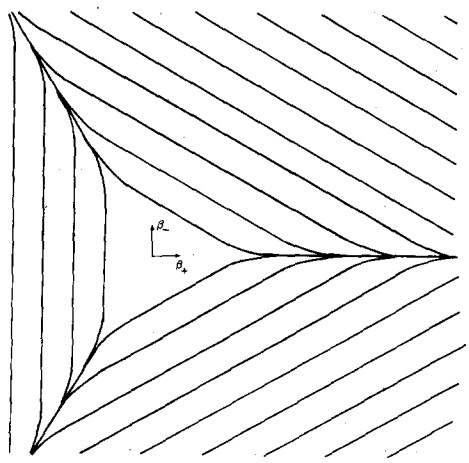


FIG. 1. Equipotentials of $V(\beta_\pm)$ for large values of $\beta_\pm$. The value of V increases exponentially from one contour to the next as one moves out from the center. The walls associated with this potential will coincide with one of the triangles formed by these lines.

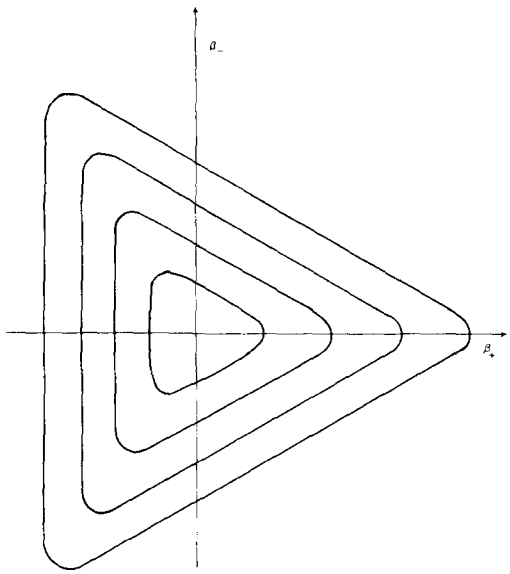


FIG. 2. Equipotentials of $V_S(\beta_\pm)$. The four equipotentials correspond to values of V_S that increase by factors of ten from 10 for the smallest triangle to 10^4 for the largest. The walls associated with this potential will coincide with one of these equipotentials.

wall perpendicular to the β_+ axis. For these walls, the entire potential terms are $3e^{-8\beta_+ - 4\Omega}$ for $V(\beta_\pm)$ and $\rho e^{-4\beta_+ - 2\Omega}$ for $V_S(\beta_\pm)$. Since the position of the walls are given by setting each of these terms equal to a constant and solving for $\beta_{+\,\text{wall}}$ as a function of Ω we see that the velocity in both cases (in the sense of $d\beta_{+\,\text{wall}}/d\Omega$) is 1/2. The last wall, the *centrifugal* wall, blocks off a region near the β_+ axis and is due to the $\delta^2 \coth^2[2(\sqrt{3})\beta_-]$ term in (4.3a). Because there is no explicit Ω dependence in this term, this wall is essentially static and moves slowly away from or toward the β_+ axis as H changes due to Eq. (4.4).

In the Hamiltonian for the diagonal case only the gravitation potential exists, and diagrammatic[2] and piecewise analytic[12] solutions have been found for the vacuum case near the singularity. These solutions can be used in our diagonal case, where we may insert $\beta_\pm$ and H as functions of Ω found from the vacuum solutions into (4.2) to find the behavior of the spinor quantities.

In symmetric case all the walls play a role, and we must study their behavior. Because the velocities of the spinor and gravitation walls are both one-half, we can have two cases, as shown in Fig. 3. In the first case the gravitation wall is inside the spinor wall, and it stays always inside, so that the universe point only interacts with it and the centrifugal wall. In the other case the gravitation wall is outside the spinor wall. Here, the gravitation wall always remain outside, but the character of the interaction depends on the sign of ρ. There seems to be no reason to expect ρ to have one sign or the other, so we must look at both cases. If ρ is positive, the spinor potential is positive, and the reverse of the situation described above for the gravitation wall inside the spinor wall occurs, that is, the universe point always interacts with the spinor wall and never touches the gravitation wall. If however, ρ is negative, the walls perpendicular to the β_+ axis are shown in Fig. 4. In this case, the universe point can interact successively with both walls. If we can describe the trajectory of the universe point far from the walls and give a set of bounce laws for interaction with them, we can build up a qualitative solution.

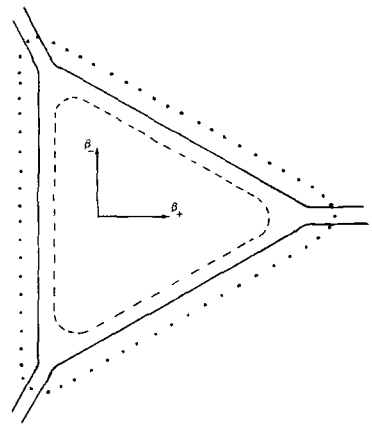


FIG. 3. Two possible positions of the spinor wall with respect to the gravitation wall. The solid line represents an equipotential of $V(\beta_\pm)$, while the dotted and dashed lines are equipotentials of $V_S(\beta_\pm)$, outside and inside the gravitation wall respectively. Since both walls move with velocity $\frac{1}{2}$, their relative position will be preserved.

It is easy to show that far from the walls, H, p_+, and p_- are constants, and the universe point moves in straight lines with velocity, in the sense of $[(\dot{\beta}_+)^2 + (\dot{\beta}_-)^2]^{1/2}$, equal to $(1 - \delta^2/H^2)^{1/2}$. If $\delta/H > \sqrt{3}/2$ this velocity is less than 1/2, and the universe point can only interact once with the centrifugal wall, then moves in a straight line out to infinite values of $\beta_\pm$. If $\delta/H < \sqrt{3}/2$, the universe point can interact at least once with the gravitation and spinor walls. From Eq. (4.4), we can see that if ρ is positive, or the gravitation wall is inside the spinor wall, H will decrease each time the universe point interacts with a wall, and eventually, δ/H will become larger than $\sqrt{3}/2$ and the universe point will stop interacting with the potentials (except for the centrifugal potential). If ρ is negative and the spinor wall is inside the gravitation wall,

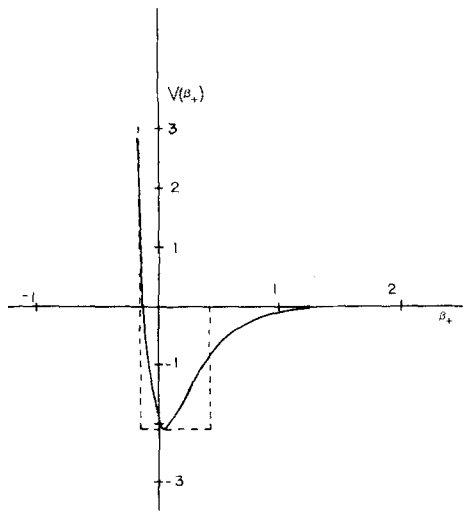


FIG. 4. The potential $V(\beta_\pm) + V_S(\beta_\pm)$ for $\beta_- = 0$ and $\rho = -5$. The dashed line is a hard wall approximating this potential. To the right of the minimum the hard wall is positioned so that the areas above and below the solid curve are equal.

whether or not this type of decoupling takes place will depend on the detailed behavior of the universe point.

We need only give a series of laws for reflection and refraction, for the case where the universe point does interact with the potentials, we have completed our qualitative solution. These laws are simplest for the centrifugal wall, where reflection is specular, and H does not change during a bounce. Reflection from the triangular walls can be calculated using the wall perpendicular to the β_+ axis and extended by symmetry. Reflection or refraction from either the gravitation or spinor walls can be calculated using two constants of motion valid for the walls perpendicular to the β_+ axis, $p_- = \text{const}$, and $p_+ + 2H = \text{const}$. The laws of reflection are

$$\frac{\sin\theta_0}{\sin\theta_i} = \left(\frac{H_i^2 - \delta^2}{H_0^2 - \delta^2}\right)^{1/2}, \tag{4.5a}$$

$$\cot\theta_0 = -\frac{\frac{4}{3}H_i}{(H_i^2 - \delta^2)^{1/2}}\csc\theta_i + \tfrac{5}{3}\cot\theta_i, \tag{4.5b}$$

where θ_i and θ_0 are the angles of incidence and reflection, respectively, and H_i and H_0 are H before and after refection. Equations (4.5) can be used to calculate both θ_0 and H_0. When ρ is negative and the spinor wall is inside the gravitation wall, the universe point passes this wall, is refracted, and eventually reflects from the gravitation wall, returns to the spinor wall and is refracted again. The only new information we need is a law of refraction. In the wall approximation, the potential is given by the dashed line in Fig. 4, where the minimum of the potential can be shown to be $-\rho^2/12$. If the universe point passes form the region of zero potential over the region of $V = -\rho^2/12$, the law for the change of H and the law of refraction are

$$\frac{\sin\theta_0}{\sin\theta_i} = \left(\frac{H_i^2 - \delta^2}{H_0^2 - \delta^2 + \rho^2/12}\right), \tag{4.6a}$$

$$\cot\theta_0 = -\tfrac{1}{3}\cot\theta_i + \tfrac{2}{3}\frac{H_i\csc\theta_i}{(H_i^2 - \delta^2)^{1/2}} + \left(\tfrac{4}{3}\cot\theta_i - \tfrac{2}{3}\frac{H_i\csc\theta_i}{(H_i^2 - \delta^2)^{1/2}}\right) \times\left(1 + \frac{\rho^2/4}{(H_i^2 - \delta^2)[2\cos\theta_i - H_i/(H_i^2 - \delta^2)^{1/2}]^2}\right)^{1/2}. \tag{4.6b}$$

If the universe point passes from $V = -\rho^2/12$ to $V = 0$,

$$\frac{\sin\theta_0}{\sin\theta_i} = \left(\frac{(H_i^2 - \delta^2 + \rho^2/12)}{H_0^2 - \delta^2}\right), \tag{4.7a}$$

$$\cot\theta_0 = -\tfrac{1}{3}\cot\theta_i - \frac{2}{3}\left(\frac{H_i\csc\theta_i}{(H_i^2 - \delta^2 + \rho^2/12)^{1/2}}\right) + \left(\tfrac{4}{3}\cot^2\theta_i + \frac{2}{3}\frac{H_i\csc\theta_i}{(H_i^2 - \delta^2 + \rho^2/12)^{1/2}}\right) \times\left(1 - \frac{\rho^2/4}{(H_i^2 - \delta^2 + \rho^2/12)(2\cos\theta_i + H_i/(H_i^2 - \delta^2 + \rho^2/12)^{1/2})^2}\right)^{1/2}. \tag{4.7b}$$

We now have a complete scheme for computing the qualitative motion of a type IX model filled with a homogeneous spinor field. The universe point moves in straight lines until it encounters one of the walls, and then is reflected or refracted according to one of the bounce laws given above and H changes according to one of the above equations.

V. CONCLUSIONS

The coupling between the Dirac field and the diagonal and symmetric type IX models has been discussed. We gave a Hamiltonian formulation of the problem and showed the incompatibility of the $k = +1$ FRW model with a homogeneous spinor field. We were able to describe the behavior of the universe point near the singularity by means of the qualitative solution method. What we lack in this paper is a quantization scheme.

A number of approaches to the problem of quantizing the gravitational field have appeared[13] from the thirties up to today. However, quantized general relativity is not well enough understood today and compromise models in which homogeneity is imposed before the gravitational field is quantized have been proposed.[14,15] In practice this is what is known as quantum cosmology and quantum models.

Some of the work which has been done in the past has been concerned with quantizing gravity coupled with matter in the sense of square-root Hamiltonian methods,[7] true canonical quantization,[16] the Heisenberg picture,[1] and others. Some of the problems associated with square-root Hamiltonians have been discussed.[16] The "true canonical quantization" scheme could be attempted in our cases. However, as we have seen, unfortunately, the $N_i = 0$ coordinate choice is incompatible with the action (3.1) with the terms in $\dot\psi$ and $\dot{\bar\psi}$, and a "true canonical quantization" could not be easily achieved for the matter field coupled to gravity. Moreover, a Heisenberg picture seems more natural for our cases. However, in order to apply this method one should be able first to solve classically equations (3.3) or equivalently (3.6). Although near the singularity, and especially for the diagonal case, we discussed how to reduce these equations to quadratures, it is not at all clear how to solve them, even in this case, and no obvious quantization scheme springs to mind.

APPENDIX

In studying cosmological models with spinor fields one must be doubly careful about the variational principle one uses. First, one must be sure that the imposition of homo-

geneity before variation still allows one to obtain Einstein's equations, and second, one must be sure that the $T_{\mu\nu}$ that appears in the equations is correct. The purpose of this Appendix is to show that the $T_{\mu\nu}$ we arrive at is valid.

The usual method of obtaining $T_{\mu\nu}$ is to write the spinor action as

$$I = \frac{1}{16\pi}\int (\sqrt{-g})[\tfrac{1}{2}i(\bar{\psi}\gamma^{\alpha}\psi_{;\alpha} - \bar{\psi}_{;\alpha}\gamma^{\alpha}\psi) - m\bar{\psi}\psi]\, d^4x, \tag{A1}$$

where $\psi_{;\alpha}$ is defined in Sec. II, and to vary I with respect to $g_{\mu\nu}$. The problem with this simple prescription is that I is a function of $\hat{h}_{(\mu)}{}^{\alpha}$, sixteen variables instead of the ten $g_{\mu\nu}$. The usual methods for solving this problem are to: (1) Vary both the matter action and the gravitational action with respect to $\hat{h}_{(\mu)}{}^{\alpha}$, obtaining a first order formulation of general relativity coupled to a spinor field (see Nelson and Teitelboim[5]), where a constraint reflecting invariance under Lorentz rotation of this basis $\hat{h}_{(\mu)}{}^{\alpha}$ is obtained; and (2) impose some condition on $\hat{h}_{(\mu)}{}^{\alpha}$ to reduce the number of independent variables in the variation. The usual condition for method (2) is (see, for example, Brill and Wheeler[17])

$$d\gamma^{\alpha} = -\tfrac{1}{2}\gamma^{\lambda\sigma}g^{\tau\alpha}\, dg_{\sigma\tau}, \tag{A2}$$

for both variations and derivatives of γ^{α}.

In general, the variation of (A1) is

$$\begin{aligned}\delta I = {} & \frac{1}{16\pi}\left(\frac{i}{2}\right)\int (\sqrt{-g})\{\bar{\psi}\delta\gamma^{\alpha}\psi_{;\alpha} - \bar{\psi}_{;\alpha}\delta\gamma^{\alpha}\psi\}\, d^4x \\ & + \frac{1}{16\pi}\left(\frac{i}{8}\right)\int (\sqrt{-g})\Big(-\frac{\sqrt{-g}_{,\alpha}}{\sqrt{-g}}\bar{\psi}\Sigma^{\alpha}\psi \\ & - \bar{\psi}_{,\alpha}\Sigma^{\alpha}\psi - \bar{\psi}\Sigma^{\alpha}\psi_{,\alpha} \\ & - \bar{\psi}\Sigma^{\alpha}{}_{\alpha}\psi - \bar{\psi}(\gamma^{\alpha}\delta\gamma^{\beta}\gamma_{\beta;\alpha} - \gamma^{\alpha}\gamma^{\beta}{}_{;\alpha}\delta\gamma_{\beta} \\ & + \delta\gamma^{\beta}\gamma_{\beta;\alpha}\gamma^{\alpha} - \gamma^{\beta}{}_{;\alpha}\delta\gamma_{\beta}\gamma^{\alpha})\psi)\, d^4x,\end{aligned} \tag{A3}$$

where $\Sigma^{\alpha} = \gamma^{\alpha}\delta\gamma^{\beta}\gamma_{\beta} + \delta\gamma^{\beta}\gamma_{\beta}\gamma^{\alpha} - \gamma^{\alpha}g^{\mu\nu}\delta g_{\mu\nu}$, and $\Sigma^{\alpha}{}_{\alpha} = \gamma^{\alpha}{}_{,\alpha}\delta\gamma^{\beta}\gamma_{\beta} + \delta\gamma^{\beta}\gamma_{\beta}\gamma^{\alpha}{}_{,\alpha} - \gamma^{\alpha}{}_{,\alpha}g^{\mu\nu}\delta g_{\mu\nu}$. If we now impose (A2), the second part of (A3) becomes zero, and the first part gives us

$$T_{\mu\nu} = -\tfrac{1}{2}i(\bar{\psi}\gamma_{(\mu}\psi_{;\nu)} - \bar{\psi}_{;(\mu}\gamma_{\nu)}\psi), \tag{A4}$$

the usual spin-1/2 energy–momentum tensor. If, however, the basis we choose does not satisfy (A2), then we must keep the extra terms in (A3). Actually, since there always exists a basis which satisfies (A2), there should exist a spinorial rotation which operating on ψ should allow us to recover (A4) without changing probabilities.

Unfortunately, the bases that have been generally used to study the Bianchi models are of the type $\hat{h}_{(\mu)}{}^{\nu} = q^{-1}{}_{\alpha\mu}\tilde{h}_{(\alpha)}{}^{\nu}$, $\hat{h}^{(\mu)}{}_{\nu} = q_{\mu\alpha}\tilde{h}^{(\alpha)}{}_{\nu}$, where the $\tilde{h}_{(\alpha)}{}^{\nu}$ are an invariant basis which is not varied, and the $q_{\alpha\mu}$ form a "square root" of the metric $\hat{g}_{\mu\nu}$ in the invariant basis, in the sense of $\hat{g}_{\mu\nu} = q_{\alpha\mu}q_{\beta\nu}\eta_{\alpha\beta}$. Also, the $q_{\mu\nu}$ are given as algebraic functions of $\hat{g}_{\mu\nu}$. It is not difficult to show that if $q_{\mu\nu}$ is not diagonal, then the extra terms in (A3) remain.

The basis that is used in the body of the paper is $\tilde{h}^{(0)}{}_{0} = 1$, $\tilde{h}^{(0)}{}_{i} = 0$ $\tilde{h}^{(i)}{}_{0} = 0$, and $\tilde{h}^{(i)}{}_{j}(x^k)$ appropriate to the Bianchi type in question. The invariant basis of one-forms is

$$\omega^0 = dt, \quad \omega^i = \tilde{h}^{(i)}{}_{j}\, dx^j. \tag{A5}$$

The matrix q is

$$\begin{aligned} & q_{00} = N, \quad q_{0i} = 0, \quad q_{i0} = q^{-1}_{ik}N_k, \\ & q_{ij} = q_{ji} = e^{-\Omega}e^{\beta}{}_{ij}, \end{aligned} \tag{A6}$$

where $N_i \equiv \hat{g}_{0i}$ and $\hat{g}_{ij} = e^{-2\Omega}e^{2\beta}{}_{ij}$. With this prescription it is not difficult to show that for Class A models T_{00} and T_{0i} are the same as would be given by (A4) (the fact that T_{0i} is the same is unexpected, but true), while T_{ij} is not, except when β_{ij} is diagonal. Notice that this means that our results agree with those of Isham and Nelson[1] for FRW models. If we accept the T_{ij} given by (A3), we can show that the homogeneous action (2.22) does indeed give us the Einstein equations for class A Models.

Note added in proof: After this paper was submitted, we received a preprint by Marc Henneaux treating some of the same material in a slightly different formalism.

[1] C. J. Isham and J. E. Nelson, Phys. Rev. D **10**, 3226 (1974).
[2] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, N.J., 1975).
[3] R. A. Matzner, L. C. Shepley, and J. B. Warren, Ann. Phys. **57**, 401 (1970).
[4] See, for example, G. Shäfer and H. Dehnen, Astron. Astrophys. **54**, 823 (1976).
[5] J. E. Nelson and C. Teitelboim, Ann. Phys. **116**, 86 (1978).
[6] R. Arnowitt, S. Deser, and C. W. Misner, *The Dynamics of General Relativity* in *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962).
[7] M. P. Ryan, *Hamiltonian Cosmology* (Springer, Berlin, 1972).
[8] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[9] See, for example, T. Davis and J. Ray, Phys. Rev. D **9**, 331 (1974); and C. Collinson and P. Morris, J. Phys. A **6**, 915 (1973).
[10] C. B. Collins and J. M. Stewart, Mon. Not. R. Astron. Soc. **153**, 419 (1971); C. B. Collins, Commun. Math. Phys. **23**, 137 (1971).
[11] O. Bogoyavlenskii and S. Novikov, in *Proceedings of the I. G. Petrovskii Seminar* (Moscow State U. P., Moscow, 1975).
[12] V. A. Belinskii, E. M. Lifshitz, and I. M. Khalatnikov, Usp. Fiz. Nauk **103**, 463 (1971) [Sov. Phys.-Usp. **13**, 745 (1971)].
[13] *Quantum Gravity: An Oxford Symposium*, edited by C. J. Isham, R. Penrose and D. W. Sciama (Clarendon, Oxford, 1975).
[14] B. De Witt, Phys. Rev. **160**, 1113 (1967); **162**, 1195 (1967) **162**, 1239 (1967).
[15] C. W. Misner, Phys. Rev. **186**, 1319 (1969).
[16] W. F. Blyth and C. J. Isham, Phys. Rev. D **11**, 768 (1975).
[17] D. R. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).

ERRATA

Erratum:Towards a factorization of M_4 [J. Math. Phys. 21, 1024(1980)]

Patrick L. Nash
Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514

(Received 21 October 1980; accepted 7 November 1980)

PACS numbers: 02.40.Vh, 99.10. + g

In the first sentence after Eq. (80), X^α should be replaced by $\dot{X}^\alpha$. In the second sentence after Eq. (91), $e^\alpha_{(3,4)}$ should be replaced by $e^\alpha_{(4)}$, and $e^\alpha_{(1,2)}$ replaced by $e^\alpha_{(j)}$.

The minus sign should be deleted from Eq. (99).

Erratum: Inverse Scattering. II. Three Dimensions [J. Math. Phys. 21, 1698 (1980)]

Roger G. Newton
Physics Department, Indiana University Bloomington, Indiana 47405

PACS numbers: 03.65.Nk, 99.10. + g

The *Corollary to Lemma* 2.2 should start: "For almost all x and y, and all $k > 0$..." and in both integrals $\check{K}_k$ should be replaced by $\check{K}^2_k$.

In *Lemma* 3.2 delete the words "for each k_0." Eq. (4.2) should have $-\infty$ as the lower limit on the integral. Eq. (5.1) should read

$$\delta_k = \tfrac{1}{2} \arg \det S_k. \tag{5.1}$$

The equation above (5.24) should read

$$Y^{b\,*}_{\kappa_n} Q \Pi^{-1}_{ik} = 0,$$

Eq. (5.24) should read

$$Y^{b\,*}_{\kappa_n} Q \prod_{m>n} \left(\mathbb{1} - B_m \frac{2\kappa_m}{\kappa_m + \kappa_n}\right)(\mathbb{1} - B_n) = 0, \tag{5.24}$$

and Eq. (5.25) should state:

For real k,

$$\Pi^{\dagger}_k = \Pi^{-1}_k$$

The parenthesis in Eq. (5.26) should be raised to the power N_n. In Eq. (6.4) the second Π_{-k} should be replaced by Π^{-1}_k. In Eq. (6.6) the power N_n should be replaced by $2N_n$.

In *Lemma* 6.1 the hypothesis should read

$$\log \det S_{k=0} = \lim_{k\to\infty} (\log \det S_k + ik\bar{V}/2\pi).$$

The right-hand side of (6.21) should have a factor of Q on its left.

In *Lemma* 6.2 it should be added that the hypotheses are the same as those of *Lemma* 6.1.

In Eq. (8.3) delete one $\int d\theta'$. The left-hand side of the equation on line 16 of Appendix 1 should be replaced by

$$(8\pi)^3 \int_{-\infty}^{\infty} dk\, |I_k(x,y)|^2.$$

On line 4 from the bottom of the right hand column of p. 1712, k should be replaced by K. In footnote 26, (8.1) should read (8.4).

 022-2488/81/030631-01$1.00