On some partitions of hypergraphs and cumulants having applications in statistical mechanics

Antoine Royer
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 28 March 1983; accepted for publication 13 October 1983)

A *hypergraph* $H = \{E_1,E_2,...,E_n\}$ is a set of sets; the sets E_i are called the *edges*, and the elements of $V(H) \equiv \cup_{i=1}^{n} E_i$ the *vertices of* H (if each edge contains only two vertices, we have an ordinary graph). Hypergraphs and *cumulants* over hypergraphs occur in various statistical mechanical problems. A cumulant over H involves a sum over partitions of the set H, whence the motivation for studying partitions of hypergraphs and related notions. Writing $E_i \overset{k\text{-}(v)}{\text{- - - -}} E_j$ if there exist at least $k \in \mathbb{N}$ vertex-disjoint paths between the edges E_i and E_j, we show that $\overset{k\text{-}(v)}{\text{- - - -}}$ is an equivalence relation on H, and denote by $\hat{P}_k H$ the partition of H into its $(\overset{k\text{-}(v)}{\text{- - - -}})$ equivalence classes. H is said to be k-(v) *connected* if every pair of edges is in the relation $\overset{k\text{-}(v)}{\text{- - - -}}$; we show that there exists a *coarsest* partition of H into k-(v) connected blocks, and denote it by $\hat{P}'_k H$. Given a partition $P(H) = \{H_1,H_2,...,H_p\}$ of H, we denote, for any real number κ, $\sigma_\kappa[P(H)] \equiv \Sigma_{j=1}^{p} (|V(H_j)| - \kappa)$, $\sigma'_\kappa[P(H)] \equiv \Sigma_{j=1}^{p} (|V(H_j)| - \kappa|\hat{P}_1 H_j|)$, where $|V(H_j)|$ is the number of vertices and $|\hat{P}_1 H_j|$ the number of connected components of H_j. We introduce "subcumulants" over H, which involve only partitions of H for which σ'_κ has the same value. The subcumulants corresponding to *minimum* values of σ'_κ, $\kappa \in \mathbb{R}$, are of practical interest (especially the cases $\kappa = 0$ and 1). We accordingly study the set $\pi'^{0}_{\kappa}(H)$ of partitions which minimize σ'_κ; this is simply related to the set $\pi^0_\kappa(H)$ of partitions which minimize σ_κ. We show that $P(H) \leqslant \hat{P}'_k H \leqslant \hat{P}_k H$ for all $P(H) \in \pi^0_k(H)$ (where $\leqslant$ signifies "is a subpartition of"), that $P \leqslant P'$ for every $P \in \pi^0_k(H)$ and $P' \in \pi^0_{\kappa-\epsilon}(H)$, $\epsilon > 0$, and that $\pi^0_\kappa(H)$ is a sublattice of the $(\leqslant)$ lattice of all partitions of H. The sets $\pi^0_\kappa(H)$ and $\pi'^{0}_{\kappa}(H)$, and the corresponding minimal values of σ_κ and σ'_κ, are explicitly determined, for any hypergraph if $\kappa \leqslant 1$, and for some special types of hypergraphs if $\kappa > 1$. We introduce a new notion of connectedness for hypergraphs, the σ *connectivity* $\Sigma(H) \equiv \text{Max}\{\kappa, \{H\} \in \pi^0_\kappa(H)\}$, and relate it to the *vertex connectivity* $k_v(H) \equiv \text{Max}\{k, H \text{ is } k\text{-}(v)\text{connected}\}$; we have, in particular, $\Sigma(H) \leqslant k_v(H)$ ($\Sigma = k_v$ if $k_v = 0$ or 1).

PACS numbers: 02.10.Cz, 02.50. + s, 05.20. − y

1. INTRODUCTION

Graph theoretic methods were initially introduced by Mayer[1] in statistical physics, to classify and handle in an efficient manner integrals of products of the form

$$\prod_{i,j} f_{ij}, \tag{1.1}$$

where the f_{ij} are quantities pertaining to pairs (i,j) of particles, and the product is over a subset (graph) of the set of all such pairs. The study and evaluation of integrals of such products over graphs is a subject of continued interest.[2]

When the Mayer method is extended to the case of particles interacting via nonpairwise forces, there occur products of the more elaborate form

$$x_{E_1} x_{E_2} \cdots x_{E_n}, \tag{1.2}$$

wherein each x_{E_i} is a quantity depending on a subset E_i of the N particles constituting the gas. The extension of the theory to such cases was done by Stell,[3] who introduced the notion of simplexes to represent products like (1.2). In a recent work on the quantum virial expansion,[4] we preferred, for reasons of efficiency, rather to use notions and terminology from the theory of *hypergraphs*[5]: A hypergraph

$$H = \{E_1,E_2,...,E_n\} \tag{1.3}$$

is a set of subsets E_i of $\{1,2,...,N\}$; the E_i are called the *edges*, and the elements of $\cup_{i=1}^{n} E_i$ the *vertices*. Thus, to the product (1.2) is associated the hypergraph (1.3).

Another widely used tool in statistical physics is the notion of *cumulant*[6,7]: Given a set $\{x_i, i \in I\}$ of stochastic variables, and denoting by $\langle\ \rangle$ the statistical average, the cumulant

$$\left\langle \prod_{i\in I} x_i \right\rangle_c \equiv \sum_P (-)^{|P|-1} (|P|-1)! \prod_{I' \in P} \left\langle \prod_{i \in I'} x_i \right\rangle, \tag{1.4}$$

where the sum is over all *partitions* P of the set I. Cumulants are mostly used in time-dependent problems.[8] But they have also been applied in equilibrium statistical mechanics[9]: Brout[10] and Kubo[6] combined cumulant with graph theoretical methods to rederive in a concise manner the Mayer expressions[1] for the classical virial coefficients; Brout[11] also obtained, in the quantum fluid case, special kinds of pertur-

 0022-2488/84/051157-19$02.50

bation expansions by using such graph-cumulant methods.

More recently, by combining hypergraph and cumulant methods, and extending the methods of Brout[10] and Kubo,[6] we obtained new expressions for the quantum virial coefficients,[4] of a form much more closely related to the Mayer classical expressions[1] than are the usual Ursell–Kahn–Uhlenbeck expressions.[12] We had previously used similar methods for dealing with pressure broadening by dense gases.[13] In the above works, there appear cumulants of products over hypergraphs; and since a cumulant involves a sum over partitions of its argument set, we were led to consider partitions

$$P(H)=\{H_1,H_2,\ldots,H_p\} \tag{1.5}$$

of a hypergraph H, together with the associated quantity

$$\sigma'_\kappa[P(H)]=\sum_{j=1}^{p}(|V(H_j)|-\kappa|\hat{P}_1H_j|), \tag{1.6}$$

where κ is a real number, $|V(H_j)|$ is the number of vertices of the hypergraph H_j, and $|\hat{P}_1H_j|$ is the number of connected components of H_j. In the problems referred to above,[4,13] the stochastic variables x_{E_i} are such that in the thermodynamic limit (number of particles $N\to\infty$, volume $\mathcal{V}\to\infty$, with $N/\mathcal{V}$ remaining finite), only the partitions of H which *minimize* σ'_κ contribute a finite amount to the cumulants involved, the value of κ depending on the specific problem (κ is 0 or 1 in Refs. 13 and 4). Thus the necessity of identifying the set $\pi'^0_\kappa(H)$ of partitions of H which minimize σ'_κ, as well as determining the corresponding minimal value $\sigma'^0_\kappa(H)$ of σ'_κ.

This was done in Refs. 4 and 13 for the individual cases $\kappa=0$ and 1 required there. However, it is more satisfying to produce these two isolated results as special cases of a more embracing theory covering all values of κ, which can eventually prove to also be of practical utility. This is the purpose of the present paper.

To determine the set $\pi'^0_\kappa(H)$ and value $\sigma'^0_\kappa(H)$,it is simplest to first find the set $\pi^0_\kappa(H)$ of partitions of H which minimize

$$\sigma_\kappa[P(H)]\equiv\sum_{j=1}^{p}(|V(H_j)|-\kappa)$$

and the corresponding minimal value $\sigma^0_\kappa(H)$, the above primed and unprimed objects being related in a simple manner.

Thus, the main theme of this paper is the study of $\pi^0_\kappa(H)$. We interrelate the sets $\pi^0_\kappa(H)$ corresponding to different values of κ, and relate them to partitions associated with notions of vertex-disjoint linkage and connectedness. We explicitly determine $\pi^0_\kappa(H)$ and $\sigma^0_\kappa(H)$, whence also $\pi'^0_\kappa(H)$ and $\sigma'^0_\kappa(H)$, for any hypergraph H in the cases $\kappa\leqslant 1$, and for some special kinds of hypergraphs in the cases $\kappa>1$. The range $\kappa\leqslant 1$ covers the two important values $\kappa=0$ and 1 occurring in the problems alluded to above.

In Sec. 2, we give the basic definitions and concepts we shall be using. A perhaps unusual construction that we introduce is an enlarged lattice[14] of partitions,[5b] which consists of not only the partitions of a given set H, but also of partitions of *subsets* of H.

In Sec. 3, we consider partitions associated with notions of vertex-disjoint linkage and connectedness.

In Sec. 4, we study the partitions which minimize σ_κ or σ'_κ. We introduce a notion of connectedness for hypergraphs which is parametrized by the *real* number κ, rather than by an integer as is the case with most such notions.

In Sec. 5, cumulants over hypergraphs are considered.

We terminate with a brief discussion in Sec. 6. An appendix deals with the extensions of Menger's theorem[15] to hypergraphs.

1.1 Notation

Given a set S, we denote by $|S|$ the number of elements in S, and by $\mathcal{P}[S]$ the set of all subsets of S (the power set of S). Given two sets S and T, we denote by $S-T$ the set of elements of S not included in T. For notational convenience, we often put the elements of a given set in one-to-one correspondence with the elements of some *indexing set*, e.g., $S=\{s_i, i\in I\}$.

Given an element $s\in S$, we denote by $\{s\}$ the subset of S consisting of the single element s. One must be careful to distinguish between s and $\{s\}$; e.g., if $\{S_i, i\in I\}$ is a set of sets, then the unions $\cup_{i\in I}S_i$ and

$$\bigcup_{i\in I}\{S_i\}=\{S_i, i\in I\}$$

are quite different objects.

The union of pairwise disjoint sets is also indicated with a summation sign, i.e.,

$$\bigcup_{i\in I}S_i\equiv\sum_{i\in I}S_i \quad \text{if } i\neq j\Rightarrow S_i\cap S_j=\varnothing \text{ for all } i,j\in I.$$

A set of pairwise disjoint nonempty sets is called a *partition*. *If* $P=\{S_i, i\in I\}$ is a partition and $\Sigma_{i\in I}S_i=S$, we write $P=P(S)$ and say that P is a partition of the set S. The elements of a partition are also called its *parts* or *cells*.

Given two sets S and T, a function $f: S\to T$ defines an *f-equivalence* relation $\underset{f}{\equiv}$ on S: $s\underset{f}{\equiv}s'$ iff $f(s)=f(s')$, $s,s'\in S$. The partition of S into its f-equivalence classes is called the f-partition of S and is denoted $P_f(S)$.

We denote $\mathbb{N}=\{0,1,2,\cdots\}$ and $\mathbb{R}$ the set of real numbers.

2. BASIC DEFINITIONS

2.1 Generalized graphs

A system,[5b] or generalized graph, or Graph (with a capital G) is a triple

$$G=(\mathcal{V},f,\mathcal{N})$$

where $\mathcal{V}$ and $\mathcal{N}$ are mutually disjoint sets, and f is a function from $\mathcal{N}$ into the set $\mathcal{P}[\mathcal{V}]$ of subsets of $\mathcal{V}$. An alternative representation is

$$G=(\mathcal{V},\mathcal{E}), \quad \text{where } \mathcal{E}=\{(\nu,f(\nu)), \nu\in\mathcal{N}\}.$$

The elements of $\mathcal{V}$ are called *vertices*. Each element $E=(\nu,f(\nu))\in\mathcal{E}$ is called an *edge*; $f(\nu)\subseteq\mathcal{V}$ is the *value* of that edge, ν is its *name*. We denote by $\nu(E)$ and $V(E)$ the name and value, respectively, of the edge E. Several different edges may have identical values (multiple edges), but no two edges can have identical names (since $\mathcal{N}$ is a set).[16] To simplify notation, and provided there is no risk of confusion, we

sometimes refer to an edge by its value alone, or by its name alone.

Let $G=(\mathscr{V},\mathscr{E})$ and $G'=(\mathscr{V}',\mathscr{E}')$ be two Graphs. G' is a *subGraph* of G, denoted $G'\subseteq G$, if $\mathscr{V}'\subseteq\mathscr{V}$ and $\mathscr{E}'\subseteq\mathscr{E}$. The union and intersection are defined in the obvious manner:

$$G\cup G'=(\mathscr{V}\cup\mathscr{V}',\ \mathscr{E}\cup\mathscr{E}'),\quad G\cap G'=(\mathscr{V}\cap\mathscr{V}',\ \mathscr{E}\cap\mathscr{E}').$$

There is one restriction however: since our convention is that the edges of a Graph must all have different names, one can take the union of two Graphs only if they are (*union*) *compatible* in the following sense: if two edges $E\in\mathscr{E}$ and $E'\in\mathscr{E}'$ have identical names, then they must also have identical values. Whenever, in the sequel, we take the union of two or more Graphs, it shall be understood that they are compatible in the above sence; this may be conveniently achieved by supposing that the Graphs considered are all subGraphs of some "large" Graph $\mathbb{G}=(\mathbb{V},\mathbb{E})$.

We represent a Graph $G=(\mathscr{V},\mathscr{E})$ by a diagram wherein each vertex $v\in\mathscr{V}$ is drawn as a small circle labeled v, and each edge $E\in\mathscr{E}$ as a dot labeled $v(E)$, with lines joining it to each vertex it contains [Fig. 1(a)]. The more standard[5] representation of a Graph is as shown in Fig. 1(b), each edge being drawn as a closed curve encircling the vertices it contains, except for edges containing just two vertices, which are usually drawn as a line joining the two vertices. Another possible representation is in terms of simplexes.[3] However, the circle-dot representation of Fig. 1(a) is easier to visualize. Also, it makes manifest the symmetry existing between edges and vertices. This symmetry often allows one to *dualize* definitions and theorems by simply interchanging edges and vertices.

A Graph whose edges all contain two or fewer vertices is called a graph (with a small g). A graph is *m-partite* if its vertex set can be partitioned into m subsets such that edges exist only between vertices belonging to different subsets. The circle-dot diagram representing a Graph G may also be viewed as a bipartite graph, whose vertices are the dots and circles, and edges the lines joining them; we call $B[G]$ that bipartite graph.

Given a Graph $G=(\mathscr{V},\mathscr{E})$ we let

$$\tilde{G}\equiv\mathscr{V}\cup\mathscr{E}=\{h_1,h_2,...,h_m\},\tag{2.1}$$

so that each h_i stands for either an edge or a vertex. We write h_i—h_j if h_i is a vertex and h_j an edge containing it, or vice versa; h_i and h_j are then said to be *incident* on each other.

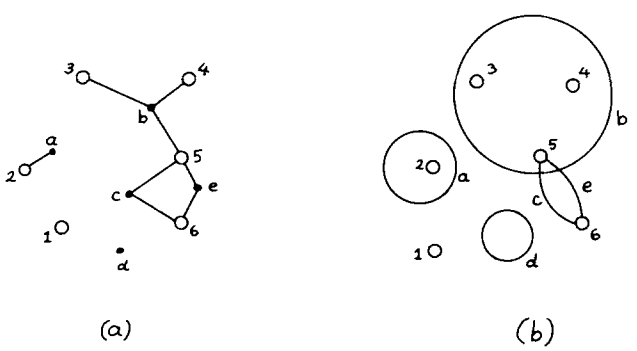


FIG. 1. (a) Circle-dot representation of the Graph $G=(\mathscr{V},\mathscr{E})$, where $\mathscr{V}=\{1,2,3,4,5,6\}$ and $\mathscr{E}=\{(a,\{2\}),(b,\{3,4,5\}),(c,\{5,6\}),(d,\varnothing),(e,\{5,6\})\}$. (b) Standard representation of the same Graph.

The *degree* $|h_i|$ is the number of elements of $\tilde{G}$ incident on h_i; h_i is *isolated* if $|h_i|=0$, *dangling* if $|h_i|=1$. Two edges (vertices) are *adjacent* if there is a vertex (edge) incident on both.

A *path* of length k, $k\in\mathbb{N}$, is an ordered sequence of k *distinct* elements h_1—h_2—···—h_k, each incident on the preceding. A *cycle* is a closed path $\overbrace{h_1—h_2—\cdots—h_k}$. We write $h_i \overset{G}{\text{- - -}} h_j$ if there exists a path in G between h_i and h_j; this is an equivalence relation, whose equivalence classes define the *connected components* of G. G is *connected* if it has a single connected component. (We write G in $h_i \overset{G}{\text{- - -}} h_j$ to distinguish from $h_i \overset{G'}{\text{- - -}} h_j$, where G' may be any other Graph also containing h_i and h_j.)

We shall speak of *removing* a vertex or an edge; this means that in the diagram of G, we remove the corresponding circle or dot, *and* the lines emanating from it. An element $h_i\in\tilde{G}$ is an *articulation* of order m, $2\leqslant m\in\mathbb{N}$, if upon its removal, the number of connected components of G increases by $m-1$; h_i is a *pure articulation* if $m=|h_i|$.

A Graph is called a *tree* if it is connected and contains no cycles. A Graph is a *forest* if all its connected components are trees (equivalently if it contains no cycles, or if all its edges and vertices are either pure articulations, dangling, or isolated) (Fig. 2).

2.2 Hypergraphs[5,16]

A *hypergraph* $H=(\mathscr{V},\mathscr{E})$ is a Graph which has no isolated edges or vertices $[E\in\mathscr{E}\Rightarrow V(E)\neq\varnothing,\ \mathscr{V}=\cup_{E\in\mathscr{E}}V(E)]$. Since every vertex is contained in an edge, a hypergraph is completely characterized by its edge set, and we shall identify

$$H\equiv\mathscr{E}.\tag{2.2}$$

The Graph notions $H\subseteq H'$, $H\cup H'$, $H\cap H'$, then have their usual set meanings. We denote by

$$V(H)=\bigcup_{E\in H}V(E)\tag{2.3}$$

the vertex set of H. Often, in order that the hypergraphs we consider be (union) compatible, they shall be assumed to be subhypergraphs of some "large" hypergraph $\mathbb{H}$.

A hypergraph is *ordered* if its elements are in one-to-one correspondence with the elements of an ordered set, e.g.,

$$H=\{E_1,E_2,...,E_n\}.$$

We say that H is *properly ordered* if

$$H^{(j)}\equiv\{E_1,E_2,...,E_j\},\ 1\leqslant j\leqslant n$$

is connected for each $j=1,...,n$ (thus if H is drawn by succes-

FIG. 2. Examples of a tree (a) and of a forest (b).

sive addition of each edge in proper order, there are no disconnected intermediate subhypergraphs). Proper order can obviously be achieved (not uniquely)iff H is connected (i.e., choose E_1 arbitrarily; among the remaining edges, at least one is adjacent to E_1 or else H would not be connected; call it E_2; etc.).

Lemma 2.1: Let the connected hypergraph $H = \{E_1,E_2,...,E_n\}$ be in proper order. Then H is a tree iff $H^{(j)} = \{E_1,E_2,...,E_j\}$ and E_{j+1} have exactly one common vertex, i.e.,

$$|V(E_{j+1}) \cap V(H^{(j)})| = 1 \tag{2.4}$$

for each $j = 1,2,...,n-1$.

Proof: (if) Suppose H is not a tree, i.e., it contains a cycle $C = \overbrace{v_a — E_a — v_b — E_b — \cdots — v_z — E_z}$. Each edge $E_x \in C$ shares at least two vertices with the union of the other edges in C. Thus, in the properly ordered sequence $\{E_1,E_2,...,E_n\}$, the right-most member of C shares at least two vertices with the union of the edges on its left, i.e., (2.4) does not hold for all $j = 1,...,n-1$.

(only if) Suppose E_{j+1} and $H^{(j)}$ share two vertices v_a and v_b for some $j \in \{1,2,...n-1\}$. Since $H^{(j)}$ is connected (H being in proper order), there exists a path $v_a - - - - v_b$ in $H^{(j)}$; then $\overbrace{E_{j+1} — v_a - - - - v_b}$ is a cycle, and H is not a tree. Q.E.D.

Lemma 2.2: For any hypergraph H, we have

$$\text{(i)} \sum_{E \in H} |E| \geqslant |\tilde{H}| - |\hat{P}_1 H| = |V(H)| + |H| - |\hat{P}_1 H|,$$

$$\text{(ii)} \sum_{v \in V(H)} |v| \geqslant |\tilde{H}| - |\hat{P}_1 H|,$$

the equalities holding iff H is a forest. [$|E| = |V(E)|$ is the number of vertices incident on edge E, $|v|$ the number of edges incident on vertex v; $|V(H)|$ is the number of vertices, $|H|$ the number of edges, and $|\hat{P}_1 H|$ the number of connected components of H. $\tilde{H} = H \cup V(H)$, so that $|\tilde{H}| = |V(H)| + |H|$].

Proof: Let us first assume that $H = \{E_1,E_2,...,E_n\}$ is connected and in proper order. Denote $H^{(j)} = \{E_1,E_2,...,E_j\}, j \leqslant n$. We have $|\tilde{H}^{(1)}| = |E_1| + 1$, and $|\tilde{H}^{(j+1)}| \leqslant |\tilde{H}^{(j)}| + |E_{j+1}|$, the equality holding iff $|V(H^{(j)}) \cap V(E_{j+1})| = 1$. Whence (i) (with $|\hat{P}_1 H| = 1$) in view of Lemma (2.1). The case $|\hat{P}_1 H| \geqslant 1$ follows immediately: for let $H_1,H_2,...H_c$ be the connected components of H; then $\Sigma_{E \in H} |E| = \Sigma_{i=1}^{c} \Sigma_{E \in H_i} |E| \geqslant \Sigma_{i=1}^{c} (|\tilde{H}_i| - 1) = |\tilde{H}| - c$. (ii) follows from (i) by the vertex-edge duality. Q.E.D.

2.3 Partitions of hypergraphs

A partition of a hypergraph H is a partition of the (edge) set H:

$$P(H) = \{H_i, i \in I\}, H_i \neq \emptyset,$$

$$i \neq j \Rightarrow H_i \cap H_j = \emptyset, \quad \sum_{i \in I} H_i = H,$$

i.e., a set of pairwise edge-disjoint subhypergraphs of H whose union is H. A partition is illustrated as in Fig. 3(b), with dashed lines delineating the different parts.

We denote by $\pi(H)$ the set of all partitions of H, and put

$$\Pi(H) = \sum_{H' \subseteq H} \pi(H')$$

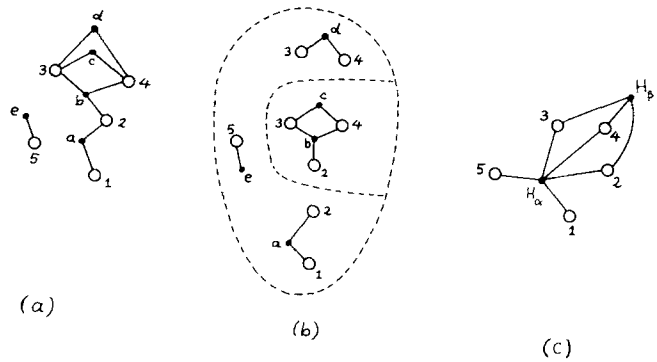


FIG. 3. (a) Hypergraph $H = \{(a,\{1,2\}), (b,\{2,3,4\}), (c,\{3,4\}), (d,\{3,4\}), (e,\{5\})\}$. (b) Partition $P(H) = \{H_\alpha, H_\beta\}$, where $H_\alpha = \{a,d,e\}$ and $H_\beta = \{b,c\}$. (c) Hypergraph $F[P(H)] = \{(H_\alpha, V(H_\alpha)), (H_\beta, V(H_\beta))\} = \{(H_\alpha,\{1,2,3,4,5\}), (H_\beta,\{2,3,4\})\}$.

the set of all partitions of all subhypergraphs of H. We denote by $P_{\text{dis}}(H)$ the *discrete* partition of H, i.e., if $H = \{E_i, i \in I\}$, then

$$P_{\text{dis}}(H) = \{\{E_i\}, i \in I\}.$$

The partition $\{H\}$ is called the *trivial* partition of H. We call

$$\pi'(H) \equiv \pi(H) - \{H\} = \{P, P \in \pi(H), |P| \geqslant 2\}$$

the set of all nontrivial partitions of H.

2.3.1 Graphs associated with partitions

A partition $P(H)$ of a hypergraph H naturally defines a new hypergraph

$$F[P(H)] = \{(H_i, V(H_i)), H_i \in P(H)\}.$$

The diagram of $F[P(H)]$ is obtained from that of H by coalescing together the dots representing the edges of H_i for each $H_i \in P(H)$ [Fig. 3(c)].

Given a partition P of a set H, and a partition P' of a set H', we define their *intersection graph* as the bipartite graph $I[P,P'] = (\mathscr{V}_I, \mathscr{E}_I)$ where

$$\mathscr{V}_I = P \cup P',$$

$$\mathscr{E}_I = \{(H_\alpha \cap H_\beta, \{H_\alpha, H_\beta\}) | H_\alpha, H_\beta \in P \cup P', H_\alpha \neq H_\beta, H_\alpha \cap H_\beta \neq \emptyset\}.$$

Clearly, in every edge value $\{H_\alpha, H_\beta\}$, one element belongs to P and the other to P' (Fig. 4).

Remark: in $F[P(H)]$ and $I[P,P']$, the edge names are sets.[17] To simplify notation, we shall usually omit the names, e.g., write $F[P] = \{V(H_i), H_i \in P\}$.

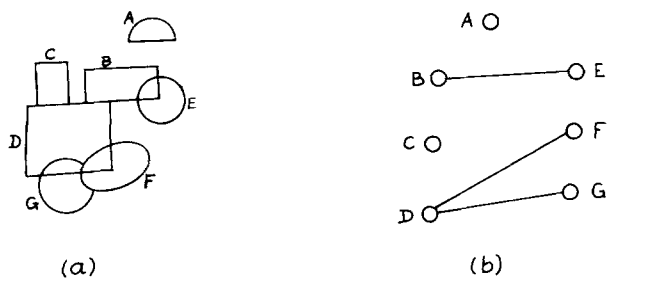


FIG. 4. (a) The closed curves (rectangles, ellipses, and truncated ellipses) represent sets. (b) The intersection graph $I[P,P']$ for $P = \{A,B,C,D\}$ and $P' = \{A,E,F,G\}$.

2.3.2 Partial ordering of partitions

Let $P(H)$ be a partition of H, and $P'(H')$ a partition of H'. If each $H_i \in P(H)$ is contained in some $H'_j \in P'(H')$, then $P(H)$ is called a *Subpartition* (with a capital S) of $P'(H')$, and conversely, $P'(H')$ is called a *Superpartition* of $P(H)$; this is abbreviated

$$P(H) \leqslant P'(H') \Leftrightarrow \{\forall\, H_i \in P(H),\ \exists\, H'_j \in P'(H') \ni H_i \subseteq H'_j\}.$$

If, moreover, $P(H) \neq P'(H')$, then $P(H)$ is called a *proper* Subpartition of $P'(H')$, abbreviated $P(H) \lneqq P'(H')$. If $P(H)$ and $P'(H')$ are partitions of the *same* set $H = H'$, we use the words subpartition and superpartition with a small *s*.

Remark: If $P(H) \leqslant P'(H')$, then, for all $H_i \in P(H)$ and $H'_j \in P'(H')$,

$$H_i \cap H'_j \neq \emptyset \Rightarrow H_i \subseteq H'_j.$$

The relation $\leqslant$ is reflexive ($P \leqslant P$), transitive ($P \leqslant P'$ and $P' \leqslant P'' \Rightarrow P \leqslant P''$), and antisymmetric[18] ($\{P \leqslant P'$ and $P' \leqslant P \Leftrightarrow P = P'\}$), and is thus a partial ordering. Thus the set of partitions $\Pi(\mathbb{H})$ is a $(\leqslant)$ poset.[19]

2.3.3 Lattice structure of $\Pi(\mathbb{H})$

Given two partitions, $P(H) = \{H_i, i \in I\}$ of the set $H \subseteq \mathbb{H}$, and $P'(H') = \{H'_j, j \in J\}$ of $H' \subseteq \mathbb{H}$, let us denote

$$P(H) \wedge P'(H') = \{H_i \cap H'_j, (i,j) \in I \times J, H_i \cap H'_j \neq \emptyset\} \in \pi(H \cap H'), \tag{2.5}$$

$$P(H) \vee P'(H') = \left\{ \bigcup_{H'' \in V(C_m)} H'',\ m \in M \right\} \in \pi(H \cup H')$$

where C_m, $m \in M$, are the connected components of the intersection graph $I[P(H), P'(H')]$.

Example: In Fig. 4, $P \wedge P' = \{A, B \cap E, D \cap F, D \cap G\}$ and $P \vee P' = \{A, B \cup E, C, D \cup F \cup G\}$.

Lemma 2.3: $P(H) \wedge P'(H')$ [resp. $P(H) \vee P'(H')$] is the coarsest Subpartition (resp. finest Superpartition) of both $P(H)$ and $P'(H')$, i.e.:

(i) $P \wedge P' \leqslant P$, $P \wedge P' \leqslant P'$ and $\{P'' \leqslant P$ and $P'' \leqslant P'\} \Leftrightarrow P'' \leqslant P \wedge P'$,

(ii) $P \vee P' \geqslant P$, $P \vee P' \geqslant P'$ and $\{P'' \geqslant P$ and $P'' \geqslant P'\} \Leftrightarrow P'' \geqslant P \vee P'$

for any partition $P''(H'')$. In other words, $P \wedge P'$ and $P \vee P'$ are the *meet* and *join*, respectively, of P and P', so that $\Pi(\mathbb{H})$ is a $(\leqslant)$ lattice.[14]

Proof: The first two relations of (i) and (ii) are obvious. To show the third, let $P''(H'') = \{H''_k, k \in K\}$:

(i): $\{P'' \leqslant P$ and $P'' \leqslant P'\} \Rightarrow \{$for each $k \in K$, $H''_k \subseteq H_{i_k}$ and $H''_k \subseteq H'_{j_k}$ for some $i_k \in I$ and $j_k \in J\} \Rightarrow H''_k \subseteq H_{i_k} \cap H'_{j_k} \Rightarrow P'' \leqslant P \wedge P'$.

(ii): Suppose $P'' \geqslant P$ and $P'' \geqslant P'$, and let H_a be an element of $P(H)$; let H_x be in the same connected component of $I[P,P']$ as H_a, i.e., there is a sequence $H_a, H'_a, H_b, H'_b, \ldots, H_x$, with $H'_a \in P'(H')$, $H_b \in P(H)$, etc., wherein each element intersects its predecessor. Since $P'' \geqslant P$, there exists $H''_a \in P''(H'')$ such that $H''_a \supseteq H_a$. Now: $H''_a \supseteq H_a \Rightarrow H''_a \cap H'_a \neq \emptyset$ (since $H_a \cap H'_a \neq \emptyset$) $\Rightarrow H''_a \supseteq H'_a$ (since $P'' \geqslant P'$) $\Rightarrow \cdots \Rightarrow H''_a \supseteq H_x$. By repeating the same reasoning for each element of either $P(H)$ or $P'(H')$ lying in the same connected component of $I[P,P']$ as H_a, C_a say, we obtain $H''_a \supseteq \bigcup_{H''' \in V(C_a)} H'''$. Whence $P'' \geqslant P \vee P'$. Q.E.D.

Remark: The set $\pi(H)$ of partitions of H is a sublattice of $\Pi(H)$. In the literature, it is usually just $\pi(H)$ which is considered.

A Partition $P(H)$ induces an equivalence relation $\overset{P}{\equiv}$ on H, and vice versa:

$$E_a \overset{P}{\equiv} E_b \Leftrightarrow \{E_a \text{ and } E_b \text{ belong to the same cell of } P(H)\}$$

for any two edges E_a and E_b of H. Let $P(H)$ and $P'(H')$ be two partitions, with the associated equivalence relations $\overset{P}{\equiv}$ on H and $\overset{P'}{\equiv}$ on H'. We denote by $(\overset{P}{\equiv}) \wedge (\overset{P'}{\equiv})$ on $H \cap H'$, and $(\overset{P}{\equiv}) \vee (\overset{P'}{\equiv})$ on $H \cup H'$, the equivalence relations associated with $P \wedge P'$ and $P \vee P'$, respectively[20]:

$$E_a (\overset{P}{\equiv}) \wedge (\overset{P'}{\equiv}) E_b \Leftrightarrow \{E_a \overset{P}{\equiv} E_b \text{ and } E_a \overset{P'}{\equiv} E_b\}; \tag{2.6}$$

$$E_a (\overset{P}{\equiv}) \vee (\overset{P'}{\equiv}) E_b \Leftrightarrow \{\text{there exists in } H \cup H' \text{ a sequence } E_0, E_1, \ldots, E_n, \text{ with } E_0 = E_a, E_n = E_b, \text{ such that } E_i \overset{P}{\equiv} E_{i-1} \text{ and/or } E_i \overset{P'}{\equiv} E_{i-1}, \text{ for each } i = 1, \ldots, n\}. \tag{2.7}$$

2.3.4 Operators $\mathcal{O}$: partitions→partitions

We now consider operators $\mathcal{O}$: partitions→partitions. For any partition $P \in$ domain $(\mathcal{O})$, we denote by $\mathcal{O}^{-1}P$ the preimage of P in $\mathcal{O}$, i.e., the set of partitions P' such that $\mathcal{O}P' = P$. Given sets of partitions $S = \{P_i, i \in I\} \subseteq$ domain$(\mathcal{O})$, and $S' = \{P'_j, j \in J\}$, we denote

$$\mathcal{O}S \equiv \{\mathcal{O}P_i, i \in I\},\quad \mathcal{O}^{-1}S' \equiv \bigcup_{j \in J} \mathcal{O}^{-1}P'_j.$$

We have

$$\mathcal{O}^{-1}\mathcal{O}S \supseteq S,\quad \mathcal{O}\mathcal{O}^{-1}S' = S' \cap \text{Range}(\mathcal{O}). \tag{2.8}$$

An operator $\mathcal{O}$ is *linear* if $\mathcal{O}(P + P') = (\mathcal{O}P) \cup (\mathcal{O}P')$ for all pairs $P(H)$, $P'(H')$ with $H \cap H' = \emptyset$. Given two operators $\mathcal{O}_1$ and $\mathcal{O}_2$, we define $\mathcal{O}_1\mathcal{O}_2$ and $\mathcal{O}_1 * \mathcal{O}_2$, where $*$ may stand for $\cup$, $\cap$, $\wedge$ or $\vee$, by

$$(\mathcal{O}_1\mathcal{O}_2)P = \mathcal{O}_1(\mathcal{O}_2 P),\quad (\mathcal{O}_1 * \mathcal{O}_2)P = (\mathcal{O}_1 P) * (\mathcal{O}_2 P).$$

2.3.5 Raising and lowering operators

We call a *raising* (resp. *lowering*) operator an operator $\mathcal{O}$: partitions→partitions, such that $\mathcal{O}P \geqslant P$ (resp. $\mathcal{O}P \leqslant P$) for all $P \in$ domain$(\mathcal{O})$.

With a partition $P(H)$ is naturally associated a raising operator $P(H)^{\uparrow}$ and a lowering operator $P(H)^{\downarrow}$, defined by

$$\begin{aligned} P(H)^{\uparrow} P'(H') &\equiv P(H) \vee P'(H'), \\ P(H)^{\downarrow} P'(H') &\equiv P(H) \wedge P'(H') \end{aligned} \tag{2.9}$$

for any partition $P'(H')$. The operator $P(H)^{\downarrow}$ is easily shown to be linear. With a hypergraph H, we associate

$$H^{\uparrow} \equiv \{H\}^{\uparrow}, \quad H^{\downarrow} \equiv \{H\}^{\downarrow}, \tag{2.10}$$

or more explicitly, letting $P'(H') = \{H'_j, j \in J\}$ and $J' = \{j \mid j \in J, H \cap H'_j \neq \emptyset\}$ (see Fig.5),

$$H^{\uparrow} P'(H') = \left\{H \cup \left(\bigcup_{j \in J'} H'_j\right)\right\} \cup \{H'_j, j \in J - J'\} \in \pi(H \cup H'), \tag{2.11a}$$

$$H^{\downarrow} P'(H') = \{H \cap H'_j, j \in J'\} \in \pi(H \cap H'). \tag{2.11b}$$

Lemma 2.4: (i) $P(H)^{\uparrow} = \Pi_{H_i \in P(H)} H_i^{\uparrow}$, (ii) $P(H)^{\downarrow} = \Sigma_{H_i \in P(H)} H_i^{\downarrow}$ (the ordering in the product is immaterial).

Proof: (i) follows from the associativity and commutativity of $\vee$,[14] and the fact that $\{H_1\} \vee \{H_2\} \vee \cdots \vee \{H_n\} = \{H_1, H_2, \ldots, H_n\}$ if the H_i, $i = 1, \ldots, n$, are pairwise disjoint. (ii) is obvious from comparing with (2.5). Q.E.D.

Remark 2.1: *If* $H \subseteq H'$, then $H \subseteq \cup_{j \in J'} H'_j$, so that $H \cup (\cup_{j \in J'} H'_j) = \cup_{j \in J'} H'_j$ in (2.11a).

Let $\mathscr{O}^{\downarrow}$ be a lowering operator, $P(H)$ a partition of a finite hypergraph H: $|H| < \infty$. Consider $(\mathscr{O}^{\downarrow})^n P(H)$, $n \in \mathbb{N}$. For some $n \leqslant |\Pi(H)|$, we will have $\mathscr{O}^{\downarrow}(\mathscr{O}^{\downarrow})^n P(H) = (\mathscr{O}^{\downarrow})^n P(H)$ [indeed, each application of $\mathscr{O}^{\downarrow}$, if it changes something, creates a *proper* Subpartition of its operand; but the poset $\Pi(H)$ is finite]. We define $(\mathscr{O}^{\downarrow})^{\infty}$ by

$$(\mathscr{O}^{\downarrow})^{\infty} P(H) = (\mathscr{O}^{\downarrow})^{|\Pi(H)|} P(H) \text{ for any partition } P(H), \quad |H| < \infty. \tag{2.12}$$

2.3.6 Partition operators

We call *partition operator* an operator $\hat{P}$: hypergraphs $\to$partitions, such that for any $H \in \text{domain}(\hat{P})$, $\hat{P}H \leqslant \{H\}$ is a partition of H or of a subhypergraph of H. We shall identify such operators with a hat.

Example 2.1: We denote by $\hat{P}_1 H$ the partition of any hypergraph H into its connected components.

With a lowering operator $\mathscr{O}^{\downarrow}$ is naturally associated a partition operator $\mathscr{O}^{\downarrow\wedge}$, and conversely, with a partition operator $\hat{P}$ a lowering operator $\hat{P}^{\downarrow}$, defined by

$$\mathscr{O}^{\downarrow\wedge} H = \mathscr{O}^{\downarrow}\{H\}, \quad \hat{P}^{\downarrow} P(H) = \sum_{H_i \in P(H)} \hat{P} H_i. \tag{2.13}$$

We have $\hat{P}^{\downarrow\wedge} = \hat{P}$, and $\mathscr{O}^{\downarrow\wedge\downarrow} = \mathscr{O}^{\downarrow}$ if $\mathscr{O}^{\downarrow}$ is linear.[21] We shall usually not distinguish notationally between $\mathscr{O}^{\downarrow}$ and $\mathscr{O}^{\downarrow\wedge}$, nor between $\hat{P}$ and $\hat{P}^{\downarrow}$, i.e., $\mathscr{O}^{\downarrow} H$ shall be understood to mean $\mathscr{O}^{\downarrow\wedge} H$, and $\hat{P} P(H)$ to mean $\hat{P}^{\downarrow} P(H)$. For instance, if $P(H) = \{H_i, i \in I\}$, then

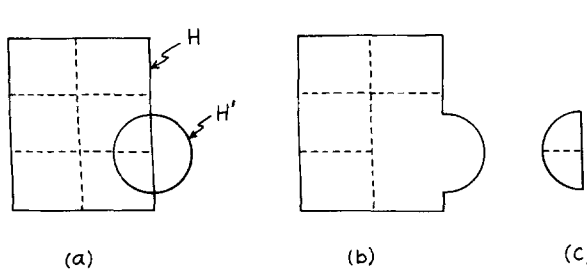


FIG. 5. (a) The large rectangle represents a set H, partitioned (dashed lines) into six subsets; we call $P(H)$ that partition of H. The circle represents a set H'. (b) The partition $P(H) \vee \{H'\} = P(H)^{\uparrow}\{H'\} = H'^{\uparrow} P(H)$ of $H \cup H'$. (c) The partition $P(H) \wedge \{H'\} = P(H)^{\downarrow} H' = H'^{\downarrow} P(H)$ of $H \cap H'$.

$$P(H)^{\downarrow} H' = P(H) \wedge \{H'\} = \{H_i \cap H', i \in I, H_i \cap H' \neq \emptyset\} \tag{2.14}$$

is the "projection" of $P(H)$ on H' (Fig. 5). Also $\hat{P}^{\infty}$ stands for $(\hat{P}^{\downarrow})^{\infty\wedge}$.

Given two partition operators $\hat{P}$ and $\hat{P}'$, we write

$$\hat{P} \leqslant \hat{P}' \text{ iff } \hat{P}H \leqslant \hat{P}'H \text{ for all } H \in \text{domain}(\hat{P}) \cap \text{domain}(\hat{P}').$$

Clearly, for any $\hat{P}$, $\hat{P}H \leqslant \hat{1}H$, i.e., $\hat{P} \leqslant \hat{1}$ where $\hat{1}$ is the trivial partition operator $\hat{1}H = \{H\}$. We define $\hat{P} \wedge \hat{P}'$ and $\hat{P} \vee \hat{P}'$ by their actions

$$(\hat{P} \wedge \hat{P}')H = (\hat{P}H) \wedge (\hat{P}'H), \quad (\hat{P} \vee \hat{P}')H = (\hat{P}H) \vee (\hat{P}'H).$$

2.3.7 c-partitions

A partition $P(H)$ is called a *c-partition* if all its parts are connected hypergraphs, i.e.,

$$P(H) \text{ is a } c\text{-partition} \Leftrightarrow \hat{P}_1 P(H) = P(H).$$

We denote by $\pi_c(H)$ the set of all c-partitions of H. Clearly, every c-partition of H is a subpartition of $\hat{P}_1 H$ [but the converse is not necessarily true (see, e.g., $\hat{P}_3 H_1$ in Fig. 9)].

Remark: $\hat{P}_1^{-1} P(H) \neq \emptyset \Leftrightarrow P(H) \in \pi_c(H)$ [thus, if $P(H) \notin \pi_c(H)$, then $\emptyset = \hat{P}_1 \hat{P}_1^{-1} P(H) \neq \hat{P}_1^{-1} \hat{P}_1 P(H) \neq \emptyset$].

A partition is *ordered* if its elements are in one-to-one correspondence with those of an ordered set, e.g.,

$$P(H) = \{H_1, H_2, \ldots, H_p\}, \quad p \in \mathbb{N}.$$

We say that $P(H)$ is in *proper order* if the hypergraphs $\Sigma_{i=1}^{j} H_i$ are connected for each $j = 1, \ldots, p$. Obviously, proper order can always be achieved if $P(H)$ is a c-partition and iff H is connected.

Lemma 2.5: (i) For any two partitions $P(H)$ and $P'(H')$,

$$P'(H') \geqslant P(H) \Rightarrow \hat{P}_1 P'(H') \geqslant \hat{P}_1 P(H).$$

(ii) If $P(H)$ is a c-partition, then

$$P'(H') \geqslant P(H) \Leftrightarrow \hat{P}_1 P'(H') \geqslant P(H).$$

Proof: (i) Let $P(H) = \{H_i, i \in I\}$, $P'(H') = \{H'_j, j \in J\}$, $\hat{P}_1 H_i = \{H_{im}, m \in M_i\}$, $\hat{P}_1 H'_j = \{H'_{jn}, n \in N_j\}$. $P'(H') \geqslant P(H) \Rightarrow H_i \subseteq H'_{j_i} \Rightarrow H_{im} \subseteq H'_{j_i}$ for some $j_i \in J$, for each $m \in M_i$ and each $i \in I$. But since H_{im} is connected, it must be entirely contained in a single connected component of H'_{j_i}, i.e., $H_{im} \subseteq H'_{j_i n}$ for some $n \in N_{j_i}$, whence $\hat{P}_1 P(H) \leqslant \hat{P}_1 P'(H')$.

(ii) ($\Rightarrow$) follows from (i), since $\hat{P}_1 P(H) = P(H)$. ($\Leftarrow$) is obvious since $P'(H') \geqslant \hat{P}_1 P'(H')$. Q.E.D.

Given a partition $P(H)$ of the hypergraph H, let us denote by $\mathscr{S}[P(H)] \subseteq \pi(H)$ the set of all superpartitions of $P(H)$, $\mathscr{S}_c[P(H)] \subseteq \pi_c(H)$ the set of all c-superpartitions of $P(H)$.

Lemma 2.6: The following statements are equivalent:

(i) $P(H)$ is a c-partition,

(ii) $\mathscr{S}_c[P(H)] = \hat{P}_1 \mathscr{S}[P(H)]$,

(iii) $\mathscr{S}[P(H)] = \hat{P}_1^{-1} \mathscr{S}_c[P(H)]$.

Proof: (i) $\Rightarrow$ (ii): Let $P(H)$ be a c-partition; $P'(H) \in \mathscr{S}[P(H)] \Rightarrow P'(H) \geqslant P(H) \Rightarrow \hat{P}_1 P'(H) \geqslant P(H)$ [by Lemma 2.5] $\Rightarrow \hat{P}_1 P'(H) \in \mathscr{S}_c[P(H)]$ [since $\hat{P}_1 P'(H)$ is a c-partition]; thus $\hat{P}_1 \mathscr{S} \subseteq \mathscr{S}_c$; but $\mathscr{S}_c \subseteq \hat{P}_1 \mathscr{S}$ since $\mathscr{S}_c \subseteq \mathscr{S}$ and $\hat{P}_1 \mathscr{S}_c = \mathscr{S}_c$. Whence $\hat{P}_1 \mathscr{S} = \mathscr{S}_c$. (ii) $\Rightarrow$ (i): If $P(H)$ is not a c-partition, then $\hat{P}_1 P(H) \not\leqslant P(H)$; but $\hat{P}_1 P(H) \in \hat{P}_1 \mathscr{S}$ [since $P(H) \in \mathscr{S}$]; thus $\hat{P}_1 \mathscr{S} \neq \mathscr{S}_c$ since $\hat{P}_1 \mathscr{S}$

contains $\hat{P}_1 P(H)$ which is a proper subpartition of $P(H)$. (ii-)⇒(iii): (ii)⇒$\mathscr{S} \subseteq \hat{P}_1^{-1}\mathscr{S}_c$ by (2.8); but $\hat{P}_1^{-1}\mathscr{S}_c \subseteq \mathscr{S}$, since $\mathscr{S}$ is the set of all superpartitions of $P(H)$; thus $\hat{P}_1^{-1}\mathscr{S}_c = \mathscr{S}$. (iii)⇒(ii) by application of $\hat{P}_1$ on (iii) and use of (2.8), noticing that $\mathscr{S}_c \subseteq \text{Range}(\hat{P}_1)$. Q.E.D.

3. (VERTEX) CONNECTEDNESS AND RELATED PARTITIONS OF A HYPERGRAPH

3.1 Partition operators associated with vertex sets

Let $W \subseteq V(\mathbb{H})$ be a set of vertices, $H \subseteq \mathbb{H}$ a hypergraph. We denote by $H - W$ the Graph (it may contain isolated edges) obtained from H by removing the vertices belonging to $V(H) \cap W$. Let E_a and E_b be two edges of H: we have

$$E_a \underset{H-W}{\text{- - - - -}} E_b$$

iff there exists in H a path between edges E_a and E_b disjoint from W (i.e., containing no vertices belonging to W). This is an equivalence relation between edges.[22] We denote by $\hat{P}_W H$ the partition of H into the equivalence classes of $\underset{H-W}{\text{- - - - -}}$. Thus, with each subset $W \subseteq V(\mathbb{H})$ is associated a partition operator $\hat{P}_W$ on $\mathscr{P}[\mathbb{H}]$. Note the case $W = \varnothing$ (the empty set): $\hat{P}_\varnothing H = \hat{P}_1 H$ is the partition of H into its connected components.

Remark: Let $\hat{P}_W H = \{H_i, i \in I\}$. We have

$$\hat{P}_\varnothing (H - W) = \{H_i - W, i \in I\}, \tag{3.1}$$

i.e., the connected components of $H - W$ are the subGraphs $H_i - W$. Thus, one may construct $\hat{P}_W H$ by first removing the vertices belonging to W, separating out the connected components of the resulting Graph $H - W$, and then reinstating the vertices that were removed (Fig. 6).

Remark: Not every partition is of the type $\hat{P}_W H$ (Fig. 7).

Lemma 3.1: Let W and W' be subsets of $V(\mathbb{H})$. We have

$$\hat{P}_{W \cup W'} \leqslant \hat{P}_W \hat{P}_{W'} \leqslant \hat{P}_W \wedge \hat{P}_{W'}, \tag{3.2}$$

$$\hat{P}_{W \cap W'} = \hat{P}_W \vee \hat{P}_{W'}. \tag{3.3}$$

Proof: To show that $\hat{P}' \leqslant \hat{P}''$, it suffices to show that if two edges E_a and E_b of a hypergraph H belong to the same element of $\hat{P}'H$, they also belong to the same element of $\hat{P}''H$. Let $\hat{P}_{W'} H = \{H'_m, m \in M\}$. Since $\hat{P}_W \hat{P}_{W'} H = \Sigma_{m \in M} \hat{P}_W H'_m$, we have $\{E_a$ and E_b in same element of $\hat{P}_W \hat{P}_{W'} H\} \Leftrightarrow \{E_a \underset{H'_m - W}{\text{- - - - -}} E_b$ for some $m \in M\}$. (3.2) is then obvious, for $E_a \underset{H-(W \cup W')}{\text{- - - - - - - -}} E_b \Rightarrow E_a \underset{H'_m - W}{\text{- - - - -}} E_b$ for some $m \in M$ [i.e., any path between E_a and E_b disjoint from $W \cup W'$ lies completely in a H'_m (since disjoint from W') and is disjoint from W]; furthermore, $E_a \underset{H'_m - W}{\text{- - - - -}} E_b \Rightarrow \{E_a \underset{H-W}{\text{- - - - -}} E_b$ and $E_a \underset{H-W'}{\text{- - - - -}} E_b\} \Leftrightarrow E_a (\underset{H-W}{\text{- - - - -}}) \wedge (\underset{H-W'}{\text{- - - - -}}) E_b$. To prove (3.3), we show that (i) $P_W \vee P_{W'} \leqslant P_{W \cap W'}$ and (ii) $P_W \vee P_{W'} \geqslant P_{W \cap W'}$. (i) is obvious since $E_a \underset{H-W}{\text{- - - - -}} E_b \Rightarrow E_a \underset{H-(W \cap W')}{\text{- - - - - - -}} E_b$, i.e., $\hat{P}_W \leqslant \hat{P}_{W \cap W'}$, and likewise $\hat{P}_{W'} \leqslant \hat{P}_{W \cap W'}$. To show (ii), we note that $E_a \underset{H-(W \cap W')}{\text{- - - - - - - -}} E_b \Rightarrow$ {there exists a path $(E_0, v_1, E_1, v_2, \ldots, E_{n-1}, v_n, E_n)$, $E_0 = E_a$, $E_n = E_b$, such that $v_i \notin W$ or $v_i \notin W'$, i.e., $E_{i-1} \underset{H-W}{\text{- - - - -}} E_i$ or $E_{i-1} \underset{H-W'}{\text{- - - - -}} E_i$, $i = 1, \ldots, n\} \Rightarrow E_a (\underset{H-W}{\text{- - - - -}}) \vee (\underset{H-W'}{\text{- - - - -}}) E_b$. Q.E.D.

Remark: $\leqslant$ cannot be replaced by $=$ in (3.2) (Fig. 8).

Lemma 3.2: Let $\hat{P}_W H = \{H_i, i \in I\}$. Then

$$V(H_i) \cap V(H_j) \subseteq W \text{ for all } i, j \in I, i \neq j. \tag{3.4}$$

Proof: Obvious, for if (3.4) did not hold, there would exist a path from some edge $E_a \in H_i$ to some $E_b \in H_j$ disjoint from W. Q.E.D.

Definition: A subset $C \subseteq V(H)$ is called a (*vertex*) *cut* of size $|C|$ of the hypergraph H if the number of connected components of H increases upon removal of the vertex set C, but not upon removal of a proper subset of C, i.e., if $|\hat{P}_C H| > |\hat{P}_\varnothing H|$, but $|\hat{P}_W H| = |\hat{P}_\varnothing H|$ for all $W \subsetneq C$.

Lemma 3.3: Let H be connected, C a (v) cut of H with $\hat{P}_C H = \{H_i, i \in I\}$. Then

$$V(H_i) \cap V(H_j) = C \quad \text{for all } i, j \in I, i \neq j. \tag{3.5}$$

Proof: It suffices to show that $C \subseteq V(H_i)$ for each $i \in I$; (3.5) will then follow by (3.4). Suppose $C \not\subseteq V(H_i)$, i.e.,

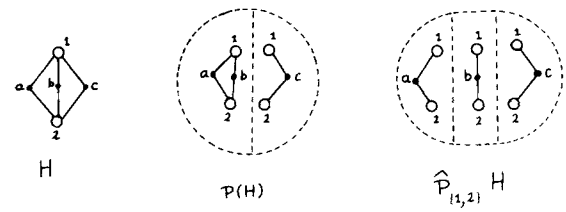


FIG. 7. The partition $P(H)$ is not of the type $\hat{P}_W H$ [in particular $P(H) \neq \hat{P}_{\{1,2\}} H$].

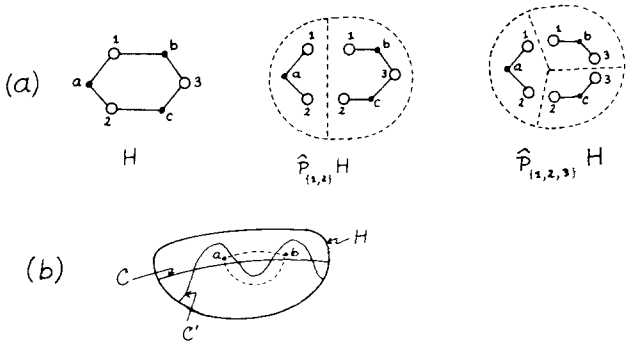


FIG. 8. (a) Here we have $\hat{P}_{\{3\}} H = \{H\}$, $\hat{P}_{\{1,2\}} \hat{P}_{\{3\}} H = \hat{P}_{\{1,2\}} H$, $\hat{P}_{\{3\}} \hat{P}_{\{1,2\}} H = \hat{P}_{\{1,2,3\}} H$ and $\hat{P}_{\{3\}} \wedge \hat{P}_{\{1,2\}} H = \hat{P}_{\{1,2\}} H$. Thus, putting $W = \{1,2\}$ and $W' = \{3\}$, we have $\hat{P}_{W'} \hat{P}_W H \neq \hat{P}_W P_{W'} H \neq \hat{P}_{W \cup W'} H \neq \hat{P}_W \wedge \hat{P}_{W'} H \neq \hat{P}_{W'} \hat{P}_W H$. (b) Schematic drawing to illustrate how $\hat{P}_C \wedge \hat{P}_{C'} H$ may differ from $\hat{P}_{C \cup C'} H$, where C and C' are (v) cuts of H; we have $E_a \underset{H-C}{\text{- - - - -}} E_b$ and $E_a \underset{H-C'}{\text{- - - - -}} E_b$, i.e., $E_a (\underset{H-C}{\text{- - - - -}}) \wedge (\underset{H-C'}{\text{- - - - -}}) E_b$, but not $E_a \underset{H-C \cup C'}{\text{- - - - -}} E_b$.

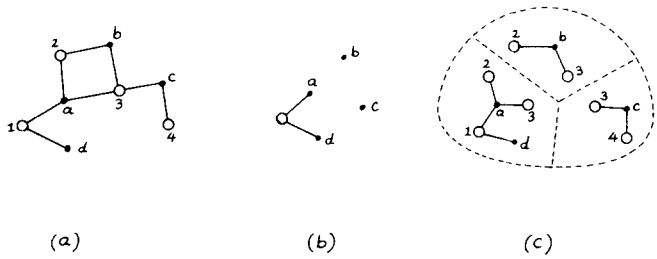


FIG. 6. (a) Hypergraph H. (b) Graph $H - W$, where $W = \{2,3,4\}$. (c) Partition $\hat{P}_W H$.

$C \cap V(H_i) = W_i \subsetneq C$. Call $H^{(i)} = H - H_i$; suppressing W_i disconnects H_i from $H^{(i)}$ [since suppressing C does, and $C - W_i$ is not in $V(H_i)$]. But this contradicts the fact that C is a (v) cut. Q.E.D.

3.2 k-(v) linked edges

Menger's theorem for hypergraphs: Let E_a and E_b be two edges of a hypergraph H, and let $k \in \mathbb{N}$. Then, the following two statements are equivalent:

there exist in H at least k pairwise vertex-disjoint paths between E_a and E_b; (3.6a)

at least k vertices of H must be removed to disconnect E_a and E_b. (3.6b)

The above is a straightforward extension of Menger's theorem for graphs[15] (an explicit proof, with some further discussion, is given in the Appendix).

Definition: Two edges E_a and E_b of H are said to be k-(v) *linked* (in H), denoted

$$E_a \overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}} E_b ,$$

if (3.6) is true.

Remark: A dual notion, k-(edge) linkage, and a stronger notion, k-linkage, are discussed in the Appendix.

Lemma 3.4: Let $C_1, C_2, \ldots, C_m$ be all the (v) cuts of size less than k of the hypergraph H. Then,

$$\left(\overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}}\right) = \left(\underset{H - C_1}{\text{- - - - -}}\right) \wedge \left(\underset{H - C_2}{\text{- - - - -}}\right) \wedge \cdots \wedge \left(\underset{H - C_m}{\text{- - - - -}}\right)$$

as relations between the edges of H. Thus $\overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}}$ is an *equivalence relation* between edges.[23]

Proof: Let E_a and E_b be two edges of H. $E_a \overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}} E_b$

$\Leftrightarrow$ {there exists no (v) cut of H of size $< k$ which disconnects E_a and E_b} $\Leftrightarrow$ $\{E_a \underset{H - C_i}{\text{- - - - -}} E_b,$

$i = 1, \ldots, m\} \Leftrightarrow E_a \left(\underset{H - C_1}{\text{- - - - -}}\right) \wedge \cdots \wedge \left(\underset{H - C_m}{\text{- - - - -}}\right) E_b$. Q.E.D.

We denote by $\hat{P}_k H$ the partition of H into its $\left(\overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}}\right)$-equivalence classes. Obviously, $\hat{P}_{k+1} \leqslant \hat{P}_k$, $\hat{P}_0 = \hat{1}$ and $\hat{P}_1$ induces the partition into connected components. In view of Lemma (3.4) (Fig. 9),

$$\hat{P}_k H = (\hat{P}_{C_1} \wedge \hat{P}_{C_2} \wedge \cdots \wedge \hat{P}_{C_m}) H .$$

3.3 k-(v) blocks and k-(v) partitions

Definitions: A hypergraph is k-(v) *connected*, or is a k-(v) *block*, if all edge pairs are in the relation $\overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}}$, or, equivalently, if $\hat{P}_k H = H$. The *vertex connectivity* $k_v(H)$ is the maximum value of k such that H is k-(v) connected; equivalently, $k_v(H)$ is the minimum size of the (v) cuts of H.[24]

Example: A hypergraph is a 2-(v) block iff it is connected and contains no articulation vertices (but it may contain articulation edges) (Fig. 10). $k_v(H) = 1 \Leftrightarrow H$ is connected and has at least one articulation vertex.

Lemma 3.5: If H is k-(v) connected, then so is $F[P(H)]$, and

$$k_v(F[P(H)]) \geqslant k_v(H),$$

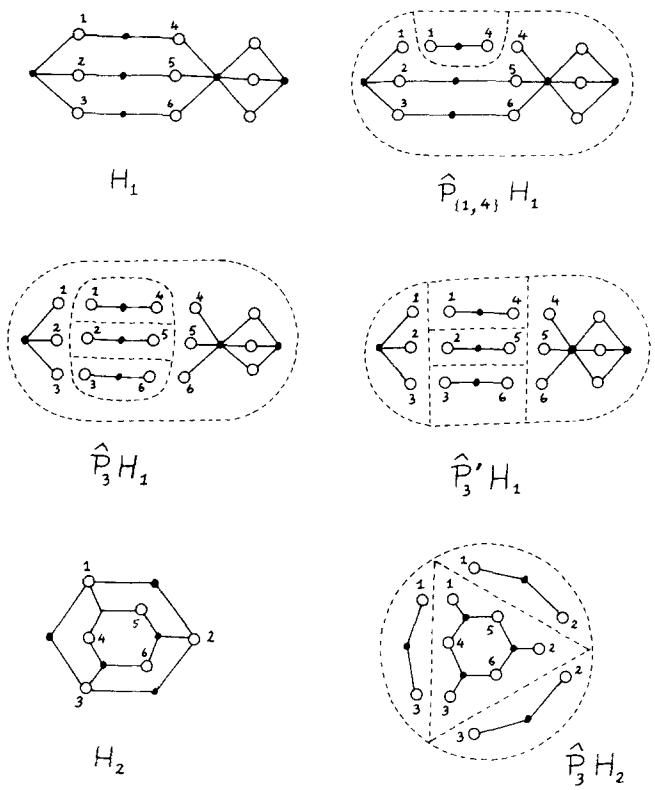


FIG. 9. $\hat{P}_3 H_1 = (\hat{P}_{C_1} \wedge \hat{P}_{C_2} \wedge \hat{P}_{C_3}) H_1$, where $C_1 = \{1,4\}$, $C_2 = \{2,5\}$, $C_3 = \{3,6\}$. $\hat{P}_3 H_2 = (\hat{P}_{C'_1} \wedge \hat{P}_{C'_2} \wedge \hat{P}_{C'_3}) H_2$, where $C'_1 = \{1,2\}$, $C'_2 = \{2,3\}$, $C'_3 = \{1,3\}$. $\hat{P}'_3 H_2 = P_{\text{dis}}(H_2)$.

for any partition $P(H)$, where $F[P(H)] = \{V(H_i), H_i \in P(H)\}$ (see Sec. 2.3.1).

Proof: $F[P(H)]$ is obtained from H by coalescing together edges (Fig. 3); but this does not affect vertex-disjoint paths. Q.E.D.

Definition: A partition $P(H)$ of a hypergraph H is called a k-(v) *partition* if each one of its parts is a k-(v) block, or, equivalently, if $\hat{P}_k P(H) = P(H)$.

In general, there exist several different k-(v) partitions of a given hypergraph; e.g., if H is k-(v) connected, then $\{H\}$ and $P_{\text{dis}}(H)$ are both k-(v) partitions.

If H is not k-(v) connected, then $\hat{P}_k H$ is not in general a k-(v) partition [any two edges of $H_i \in \hat{P}_k H$ are k-(v) linked in H, but not necessarily in H_i (see $\hat{P}_3 H_{1,2}$ in Fig. 9)]. However, repeated application of $\hat{P}_k$ eventually yields a k-(v) partition—the coarsest such partition of H in fact. Indeed, denoting

$$\hat{P}'_k \equiv (\hat{P}_k)^{\infty}$$

we have

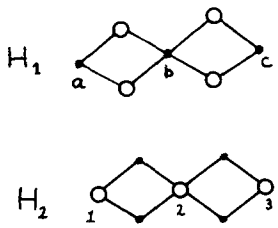


FIG. 10. H_1 is 2-(v) connected, even though it contains an articulation edge (E_b). H_2 is not 2-(v) connected, since it contains an articulation vertex (v_2).

Lemma 3.6: $\hat{P}'_k H$ is a k-(v) partition, and all other k-(v) partitions of H are subpartitions of $\hat{P}'_k H$. Obviously, $\hat{P}'_{k+1} \leqslant \hat{P}'_k \leqslant \hat{P}_k$.

Proof: $\hat{P}'_k H$ is a k-(v) partition since $\hat{P}_k(\hat{P}_k)^\infty H = (\hat{P}_k)^\infty H$. Let $P(H) = \{G_j, j\in J\}$ be any k-(v) partition of H. We have $P(H) \leqslant \hat{P}_k H$, since $E_a \overset{k\text{-}(v)}{\underset{G_j}{\text{- - - - -}}} E_b$ implies $E_a \overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}} E_b$, for any two edges E_a and E_b of G_j. Let $\hat{P}_k H = \{H'_m, m\in M\}$, and let $G_j, j\in J_m \subseteq J$, be the elements of $P(H)$ contained in H'_m, i.e., $P(H)^\downarrow H'_m = \{G_j, j\in J_m\}$; this is a partition of H'_m into k-(v) blocks, so again $P(H)^\downarrow H'_m \leqslant \hat{P}_k H'_m$ for each $m\in M$, implying $P(H) \leqslant (\hat{P}_k)^2 H$. Repeating this argument $|H(H)|$ times, we obtain $P(H) \leqslant (\hat{P}_k)^\infty H$. Q.E.D.

Repeated application of $\hat{P}_k$ is not the simplest manner of obtaining $\hat{P}'_k H$. It is simpler to perform cuts of size $<k$, until there remain only k-(v) blocks (Fig. 11)[25]:

Lemma 3.7: Let C_1 be a cut of H of size $<k$, C_2 a cut of one or more elements of $\hat{P}_{C_1} H$, C_3 a cut of one or more elements of $\hat{P}_{C_2}\hat{P}_{C_1} H$, etc., all of sizes $<k$. The partition of H obtained after performing such cuts, until there remains only k-(v) blocks, is $\hat{P}'_k H$.

Proof: It suffices to show that $\hat{P}'_k H \leqslant \hat{P}_{C_n}\hat{P}_{C_{n-1}} \cdots \hat{P}_{C_2}\hat{P}_{C_1} H$ for any $n\in\mathbb{N}$. Let $\hat{P}'_k H = \{H_i, i\in I\}$. We have $\hat{P}'_k H \leqslant \hat{P}_{C_1} H$, since $E_a \overset{k\text{-}(v)}{\underset{H_i}{\text{- - - - -}}} E_b \Rightarrow E_a \overset{}{\underset{H - C_1}{\text{- - - - -}}} E_b$. Thus, $\hat{P}'_k H = \hat{P}'_k \hat{P}_{C_1} H$. But again, $\hat{P}'_k \hat{P}_{C_1} H \leqslant \hat{P}_{C_2}\hat{P}_{C_1} H$, etc. Q.E.D.

3.4 The cases $k \leqslant 2$

In general, $\hat{P}'_k H \neq \hat{P}_k H$, because the paths between k-(v) linked edges are not necessarily contained in the $\overset{k\text{-}(v)}{\underset{H}{\text{- - - - -}}}$ equivalence class of these edges (see H_1 and H_2 in Fig. 9); however, they are when $k \leqslant 2$, and we have

Lemma 3.8: $\hat{P}'_0 = \hat{P}_0 = \hat{1}$, $\hat{P}'_1 = \hat{P}_1$, $\hat{P}'_2 = \hat{P}_2$.

Proof: $\hat{P}'_0 = \hat{P}_0 = \hat{1}$ is obvious. To demonstrate that $\hat{P}_k H = \hat{P}'_k H$, it suffices to show that if an edge E belongs to one of k vertex-disjoint paths between two edges E_a and E_b, then E is k-(v) linked to E_a or E_b. This is obvious in the case $k = 1$. Case $k = 2$: Let π_1 and π_2 be two vertex-disjoint paths between E_a and E_b, and let E belong to π_1 say; we then have the two vertex-disjoint paths $E \overset{\pi_1}{\text{- - - - -}} E_a$ and $E \overset{\pi_1}{\text{- - - - -}} E_b \overset{\pi_2}{\text{- - - - -}} E_a$, where $E \overset{\pi_1}{\text{- - - - -}} E_a$ designates that part of the path π_1 lying between E and E_a [Fig. 12(a)]. Q.E.D.

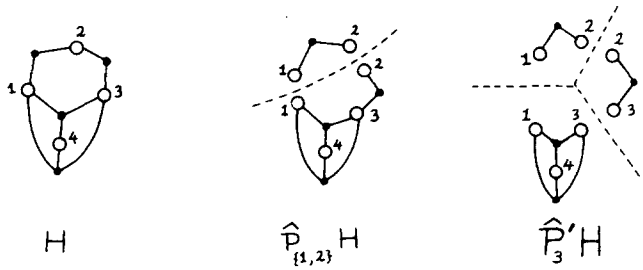


FIG. 11. $\hat{P}'_3 H = \hat{P}_{|3|}\hat{P}_{|1,2|} H = \hat{P}_{|2|}\hat{P}_{|1,3|} H = \hat{P}_{|1|}\hat{P}_{|2,3|} H$.

Remark: In view of Lemma (3.7), we have[26]

$$\hat{P}_2 H = \hat{P}'_2 H = \hat{P}_{\{v_1\}}\hat{P}_{\{v_2\}} \cdots \hat{P}_{\{v_m\}} H = \hat{P}_{\{v_1, v_2, \ldots, v_m\}} H,$$

where $v_1, v_2, \ldots, v_m$ are all the articulation vertices of H.

Lemma 3.9: Let H' be a connected subhypergraph of H. Then $(\hat{P}_2 H)^\downarrow H'$ is a *c-partition of* H'.

Proof: Let $\hat{P}_2 H = \{S_j, j\in J\}$. We must show that each (nonempty) $S_j \cap H'$ is connected. Suppose $H' \cap S_j$ is not connected, and let E_a and E_b be two edges of $H' \cap S_j$, unlinked in $H' \cap S_j$. Since H' is connected, there exists a path[27] $\pi_1 \subset H'$ between E_a and E_b, $\pi_1 \not\subset H' \cap S_j$ [Fig. 12(b)]. Since S_j is connected, there exists a path $\pi_2 \subset S_j$ between E_a and E_b, $\pi_2 \not\subset H' \cap S_j$. Then $\pi_1 \cup \pi_2$ contains a cycle with edges from both S_j and $H' - S_j$, i.e., S_j is not a complete $\overset{2\text{-}(v)}{\text{(- - - - -)}}$ equivalence class. Q.E.D.

Lemma 3.10: Let $P(H)$ be a c-partition of the hypergraph H. The following two statements are equivalent:

(i) $P(H) \geqslant \hat{P}_2 H$;

(ii) Each $H_i \in P(H)$ shares at most one vertex with each connected component of $H - H_i$. [In fact, (ii)$\Rightarrow$(i) for *any* partition $P(H)$.]

Proof: (i)$\Rightarrow$(ii): Let $P(H) \geqslant \hat{P}_2 H$. Suppose $H_i \in P(H)$ shares vertices v_1 and v_2 with a connected component H' of $H - H_i$. Since H_i is connected, there is a path π_1 between v_1 and v_2 contained in H_i; likewise, since H' is connected, there is a path π_2 between v_1 and v_2 contained in H' [Fig. 12(c)]. Let $E_a \in \pi_1$ and $E_b \in \pi_2$, where E_a and E_b are edges of H. Clearly, $E_a \overset{2\text{-}(v)}{\underset{H}{\text{- - - - -}}} E_b$ [the two paths are $E_a \overset{\pi_1}{\text{- - - - -}} v_1 \overset{\pi_2}{\text{- - - - -}} E_b$ and $E_a \overset{\pi_1}{\text{- - - - -}} v_2 \overset{\pi_2}{\text{- - - - -}} E_b$; but E_b is not in H_i, whence a contradiction, since H_i must contain complete $\overset{2\text{-}(v)}{\underset{H}{\text{(- - - - -)}}}$ equivalence classes.

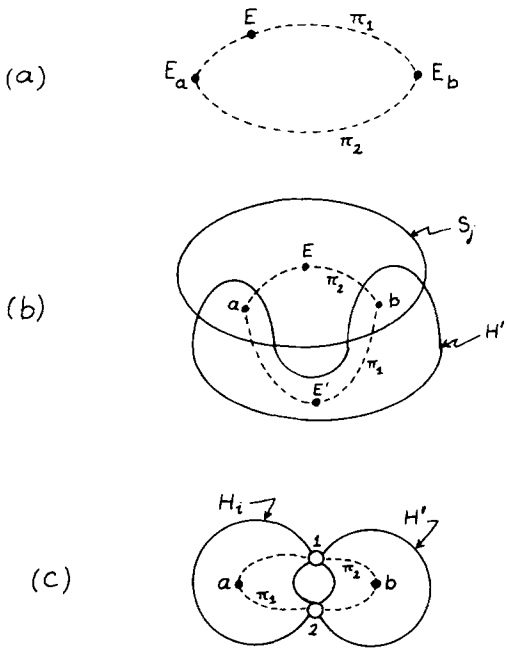


FIG. 12. (a) See proof of Lemma 3.8. (b) See proof of Lemma 3.9. (c) See proof of Lemma 3.10.

(ii)⇒(i): Suppose (ii) is true. Since all paths between any edge $E_a \in H_i$ and any edge $E_b \in H - H_i$ must pass through the single vertex shared by H_i and the connected component of $H - H_i$ containing E_b, these two edges are at most 1-(v) linked; thus each H_i contains complete ($\overset{2\text{-}(v)}{\underset{H}{\text{- - - - -}}}$) equivalence classes, implying (i). Q.E.D.

It is easy to see that for any hypergraph H, $F[\hat{P}_2 H]$ is a forest (let $\hat{P}_2 H = \{S_j, j \in J\}$; every cycle in H is contained inside one of the S_j; it disappears when the edges of S_j are coalesced together, i.e., $F[\hat{P}_2 H]$ contains no cycles). More generally we have [Fig. 13(a)]:

Lemma 3.11: Let H be a connected hypergraph, $P(H) = \{H_i, i \in I\}$ a partition of H. Then $F[P(H)] = \{V(H_i), i \in I\}$ is a tree iff $P(H)$ is a c-superpartition of $\hat{P}_2 H$.

Proof: (if) Let $P(H) = \{H_1, H_2, \ldots, H_p\}$ be a c-superpartition of $\hat{P}_2 H$, in proper order. By Lemma 3.10, H_{j+1} shares a single vertex with $\Sigma_{i=1}^{j} H_i$ [which is connected since $P(H)$ is in proper order], i.e., $|V(H_{j+1}) \cap (\cup_{i=1}^{j} V(H_i))| = 1$ for each $j = 1, \ldots, p-1$, so $\{V(H_i), i = 1, \ldots, p\}$ is a tree by Lemma 2.1.

(only if): Let $P(H) = \{H_i, i \in I\}$ be such that $\{V(H_i), i \in I\}$ is a tree. Suppose that H_1, say, is not connected. Let E_a and E_b be two edges of H_1 not linked in H_1, whence nonadjacent. Since H is connected, there exists in H a path between E_a and E_b, not contained in H_1; this path becomes a cycle in $F[P(H)]$ (wherein the dots representing E_a and E_b are coalesced together), contradicting the hypothesis. Thus $P(H)$ is a c-partition; furthermore, since $F[P(H)]$ is a tree, $P(H)$ satisfies (ii) of Lemma 3.10, i.e., it is a superpartition of $\hat{P}_2 H$. Q.E.D.

Lemma 3.12: Let H be any hypergraph, and $P(H) \leqslant \hat{P}_1 H$. Then $F[P(H)]$ is a forest iff $P(H)$ is a c-superpartition of $\hat{P}_2 H$.

Proof: Let $\hat{P}_1 H = \{H_i, i \in I\}$. $P(H) \leqslant \hat{P}_1 H \Rightarrow P(H) = \Sigma_{i \in I} P(H)^{\downarrow} H_i$, whence $\hat{P}_1 F[P(H)] = \{F[P^{\downarrow} H_i], i \in I\}$ (since, in view of Lemma 3.5, each $F[P^{\downarrow} H_i]$ is connected, and moreover the vertex sets $V(F[P^{\downarrow} H_i]) = V(H_i)$, $i \in I$, are pairwise disjoint). By Lemma 3.11, the $F[P^{\downarrow} H_i]$, $i \in I$, are trees iff each $P^{\downarrow} H_i$ is a c-superpartition of $\hat{P}_2 H_i$, i.e., iff $P(H) = \Sigma_{i \in I} P^{\downarrow} H_i$ is a c-superpartition of $\hat{P}_2 H$. Q.E.D.

Remark: If we do not specify $P(H) \leqslant \hat{P}_1 H$ in Lemma 3.12, the "only if" part must be replaced by: $F[P(H)]$ is a forest only if $P(H) \geqslant \hat{P}_2 H$ [Fig. 13(b)].

4. A CONTINUOUS NOTION OF CONNECTEDNESS AND RELATED PARTITIONS

Throughout this section, k, κ, and ϵ will denote positive integers ($k \in \mathbb{N}$), real numbers ($\kappa \in \mathbb{R}$), and strictly positive real numbers ($\epsilon \in \mathbb{R}$, $\epsilon > 0$), respectively.

Given a hypergraph H, let us denote, for any real number κ,

$$\bar{\sigma}_\kappa(H) \equiv \sum_{E \in H} (|E| - \kappa) = \sum_{E \in H} |E| - \kappa |H|.$$

Lemma 4.1: For any hypergraph H,
$\bar{\sigma}_\kappa(H) \geqslant |V(H)| - (\kappa - 1)|H| - |\hat{P}_1 H|$,
the equality holding iff H is a forest.

Proof: Immediate from Lemma 2.2.

Given a partition $P(H) = \{H_i, i \in I\}$, let us denote

$$\sigma_\kappa[P(H)] \equiv \bar{\sigma}_\kappa(F[P(H)]) = \sum_{i \in I} (|V(H_i)| - \kappa) = \sum_{i \in I} |V(H_i)| - \kappa |P(H)|.$$

Obviously

$$P = \sum_{j \in J} P_j \Rightarrow \sigma_\kappa[P] = \sum_{j \in J} \sigma_\kappa[P_j], \tag{4.1}$$

$$\sigma_\kappa[P_{\text{dis}}(H)] = \bar{\sigma}_\kappa(H), \quad \sigma_\kappa[\{H\}] = |V(H)| - \kappa. \tag{4.2}$$

The function $\sigma_\kappa : \pi(H) \to \mathbb{R}$, defines an equivalence relation on $\pi(H)$ (see Sec. 1.1). We denote by

$$\pi_\kappa^i(H), \quad i = 0, 1, \ldots, n_\kappa(H)$$

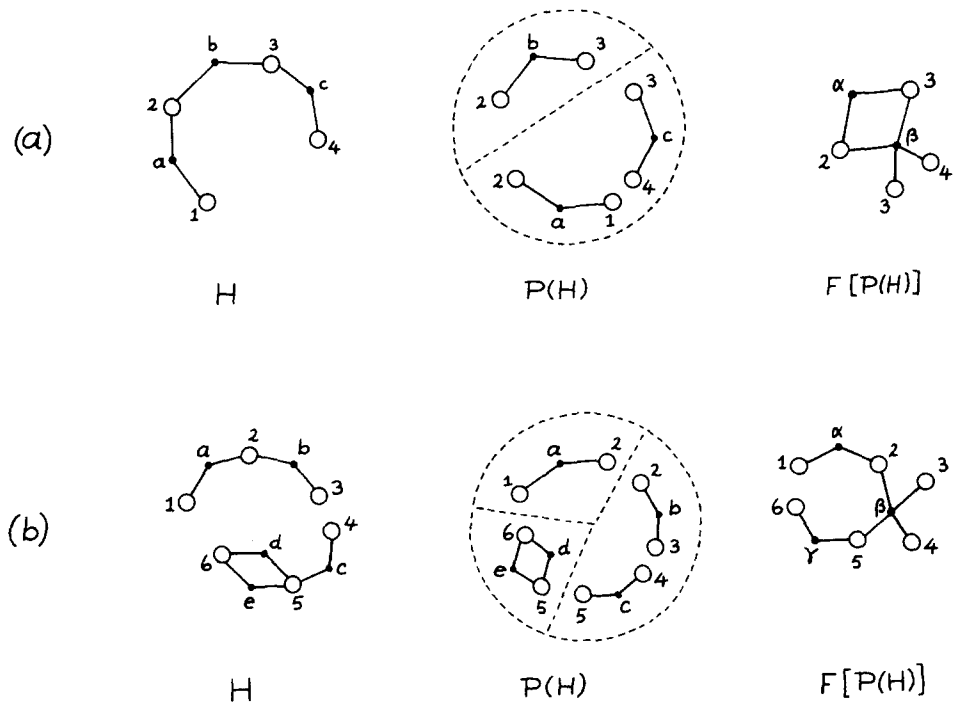


FIG. 13. Illustrating Lemmas 3.11 and 3.12. (a) $P(H) = \{H_\alpha, H_\beta\}$, where $H_\alpha = \{b\}$ and $H_\beta = \{a,c\}$, so that $F[P(H)] = \{(\alpha,\{2,3\}), (\beta,\{1,2,3,4\})\}$. We here have $\hat{P}_1 H \geqslant P(H) \geqslant \hat{P}_2 H$, whence $P(H)$ not a c-partition $\Rightarrow F[P(H)]$ is not a tree. (b) $P(H) = \{H_\alpha, H_\beta, H_\gamma\}$, where $H_\alpha = \{a\}$, $H_\beta = \{b,c\}$, and $H_\gamma = \{d,e\}$, so that $F[P(H)] = \{(\alpha,\{1,2\}), (\beta,\{2,3,4,5\}), (\gamma,\{5,6\})\}$. We here have $P(H) \geqslant \hat{P}_2 H$, but not $P(H) \leqslant \hat{P}_1 H$, so that $F[P(H)]$ can be a tree, even if $P(H)$ is not a c-partition.

the σ_κ-equivalence classes of $\pi(H)$, and call

$$\sigma_\kappa^i(H) = \sigma_\kappa[P(H)], \quad P(H) \in \pi_\kappa^i(H).$$

We index the $\pi_\kappa^i(H)$ such that $\sigma_\kappa^i(H) < \sigma_\kappa^j(H)$ if $i < j$. Thus, $\sigma_\kappa^0(H)$ is the minimum value of σ_κ over $\pi(H)$, and $\pi_\kappa^0(H)$ is the set of partitions of H which minimize σ_κ.

Given a partition $P(H)$, let us denote

$$\Delta[P(H)] \equiv \sum_{H_i \in P(H)} |V(H_i)| - |V(H)|.$$

This measures the degree of mutual overlap of the vertex sets $V(H_i)$. Obviously,

$$\Delta[P(H)] \geqslant 0, \tag{4.3}$$

the equality holding iff the sets $V(H_i)$, $H_i \in P(H)$, are pairwise disjoint, i.e., iff $P(H) \geqslant \hat{P}_1 H$.

Given a nontrivial partition $P(H)$ [i.e., $|P(H)| \geqslant 2$], we denote

$$R[P(H)] \equiv \Delta[P(H)]/(|P(H)| - 1), \quad P(H) \in \pi'(H).$$

4.1 κ-(σ) connectedness and σ-connectivity

Definition: We say that the hypergraph H is κ-(σ) connected if $\{H\} \in \pi_\kappa^0(H)$, i.e., if

$$\sigma_\kappa[P(H)] \geqslant \sigma_\kappa[\{H\}] \tag{4.4a}$$

or, equivalently, if

$$R[P(H)] \geqslant \kappa \tag{4.4b}$$

for all nontrivial partitions $P(H)$. If $\pi_\kappa^0(H) = \{\{H\}\}$, i.e., if $\geqslant$ is replaced by $>$ in (4.4), we say that H is κ-(σ) *Connected*, with a capital C.

Lemma 4.2: (i) If H is κ-(σ) connected, it is also $(\kappa - \epsilon)$-(σ) Connected. (ii) If H is κ-(σ) Connected, there exists $\epsilon > 0$ such that H is $(\kappa + \epsilon)$-(σ) connected.

Proof: (i) is obvious from considering (4.4b). (ii): If H is κ-(σ) Connected, then for any $P \in \pi'(H)$, $R[P] = \kappa + \epsilon_P$ for some $\epsilon_P > 0$. Letting $\epsilon = \min\{\epsilon_P, P \in \pi'(H)\}$, we have $P \in \pi'(H) \Rightarrow R[P] \geqslant \kappa + \epsilon$. Q.E.D.

Definition: The *σ-connectivity* $\Sigma(H)$ of a hypergraph H is the largest value of κ for which H is κ-(σ) connected, or, equivalently,

$$\Sigma(H) \equiv \mathrm{Min}\{R[P], P \in \pi'(H)\}.$$

Remark: The partitions which minimize R are always c-partitions, since for any $P(H)$, $R[\hat{P}_1 P(H)] \leqslant R[P(H)]$ (from $\Delta[\hat{P}_1 P] = \Delta[P]$ and $|\hat{P}_1 P| \geqslant |P|$), the equality holding iff $\hat{P}_1 P(H) = P(H)$. Thus, to determine $\Sigma(H)$, if suffices to sample $\pi_c'(H)$.

We now relate (σ) connectedness with (v) connectedness:

Lemma 4.3: For any (v) cut C of a hypergraph H, we have

(i) $R[\hat{P}_C H] = |C|$,

(ii) $\sigma_\kappa[\hat{P}_C H] = \sigma_\kappa[\{H\}] + (|\hat{P}_C H| - 1)(|C| - \kappa)$

(if $C = \varnothing$, then $\hat{P}_C = \hat{P}_1$).

Proof: From Lemma 3.3, we deduce

$$\Delta[\hat{P}_C H] = (|\hat{P}_C H| - 1)|C|$$

whence (i) and (ii). Q.E.D.

Lemma 4.4: (i) If H is k-(σ) Connected or $(k + \epsilon)$-(σ) connected, then it is also $(k + 1)$-(v) connected.

(ii) $\Sigma(H) \leqslant k_v(H)$.

Proof: If H is not $(k + 1)$-(v) connected, then there exists a (v) cut C with $|C| \leqslant k < k + \epsilon$. But $R[\hat{P}_C H] = |C| \,(\lesseqgtr) \kappa \Rightarrow H$ is not κ-(σ)($\overset{c}{C}$)onnected. (ii) follows from (i) or Lemma 4.3(i). Q.E.D.

The converse of Lemma 4.4 is not true in general [e.g., in Fig. 14, H_4 is 3-(v) connected but not 2-(σ) Connected, and $\Sigma(H_4) < k_v(H_4)$], despite the fact that no (v) cut can decrease σ_κ if $\kappa < k_v(H)$ [by Lemma 4.3(ii)], and $R[\hat{P}_C H] \geqslant k_v(H)$ [by Lemma 4.3(i)]. What may happen, e.g., is that, because the parts of $\hat{P}_C H$ are not necessarily $k_v(H)$-(v) connected, one may decrease σ_κ and R by performing further cuts (see the sequence of cuts of H_4 in Fig. 14). However, the parts of $\hat{P}_C H$ are necessarily 1-(v) connected (by definition of $\hat{P}_C$), so that [in view of (4.1)] a second cut cannot decrease σ_κ if $\kappa \leqslant 1$, and likewise for any further cuts. This suggests the following lemma (but does not prove it, since not every partition is obtainable as a succession of cuts):

Lemma 4.5: (i) All hypergraphs are 0-(σ) connected [as well as 0-(v) connected];

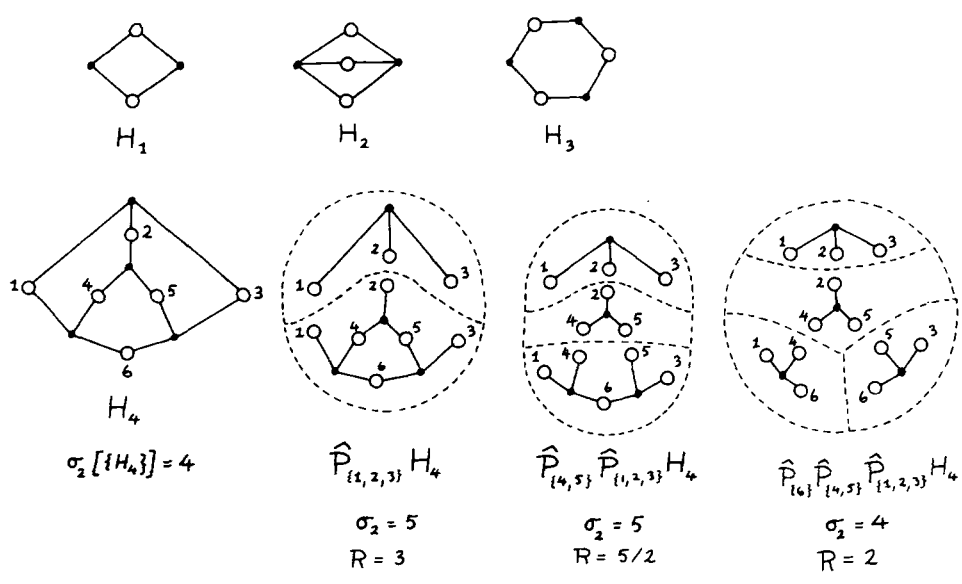


FIG. 14. $\Sigma(H_1) = k_v(H_1) = 2$, $\Sigma(H_2) = k_v(H_2) = 3$, $\Sigma(H_3) = \frac{3}{2}$, $k_v(H_3) = 2$, $\Sigma(H_4) = 2$, $k_v(H_4) = 3$.

(ii) If a hypergraph is 1-(v) connected, it is also 1-(σ) connected;

(iii) If a hypergraph is 2-(v) connected, it is also 1-(σ) Connected.

Proof: (i) follows from (4.3). (ii): Let H be connected, and let $P(H) = \{H_i, i\in I\} \in \pi'(H)$. We have, from applying Lemma 4.1 to $F[P(H)] = \{V(H_i), i\in I\}$, which is connected by Lemma 3.5, and using $V(F) = V(H)$,

$$\sigma_1[P(H)] \equiv \tilde{\sigma}_1(F[P(H)]) \geqslant |V(H)| - 1, \tag{4.5}$$

i.e., H is 1-(σ) connected. (iii): The equality in (4.5) holds iff $F[P(H)]$ is a tree. If H is 2-(v) connected, then so is $F[P(H)]$, by Lemma 3.5, and $F[P(H)]$ is not a tree (since moreover $|I| \geqslant 2$); whence $\sigma_1[P(H)] > |V(H)| - 1$, i.e., H is 1-(σ) Connected. Q.E.D.

Remark: A 2-(v) connected hypergraph is not necessarily 2-(σ) connected, e.g., H_3 in Fig. 14.

Lemma 4.6: (i) For any hypergraph H[28]:

$$k_v(H) = 0 \Leftrightarrow \Sigma(H) = 0,$$

$$k_v(H) = 1 \Leftrightarrow \Sigma(H) = 1,$$

$$k_v(H) \geqslant 2 \Leftrightarrow \Sigma(H) > 1.$$

(ii) There exists no hypergraph H with $\Sigma(H) < 0$ or $0 < \Sigma(H) < 1$.

Proof: (i) follows from Lemmas 4.4 and 4.5 and Eq. (4.3) [noticing that $\Delta[\hat{P}_1 H] = 0$, and $|\hat{P}_1 H| \geqslant 2$ if $k_v(H) = 0$]. (ii) follows from (i) and the fact that $k_v(H) \in \{0,1,2,\cdots\}$. Q.E.D.

Examples: (i) Let $H = \{(a,\{1,2,\ldots,n\}),(b,\{1,2,\ldots,n\})\}$ (e.g., H_1 and H_2 in Fig. 14); then $\Sigma(H) = R[P_{\text{dis}}(H)] = (2n-n)/(2-1) = n = k_v(H)$.

(ii) Let $H = \{(a_1,\{1,2\}), (a_2,\{2,3\}), (a_3,\{3,4\}),\ldots,(a_n,\{n,1\})\}$, i.e., a ring (e.g., H_1 and H_3 in Fig. 14). It is easy to show that[29] $\Sigma(H) = R[P_{\text{dis}}(H)] = (2n-n)/(n-1) = n/(n-1)$. Here, $k_v(H) = 2$.

Remark: In these examples, R is minimized by $P_{\text{dis}}(H)$; this is not always the case (e.g., in Fig. 15, $\Sigma(H) \neq R[P_{\text{dis}}(H)]$).

The (σ) connectivity of a hypergraph H sets a lower bound on the degree of mutual overlap of its edges, since we have [setting $P(H) = P_{\text{dis}}(H)$ in (4.4b)]

$$\sum_{E\in H} |V(E)| - \left| \bigcup_{E\in H} V(E) \right| \geqslant \Sigma(H)(|H| - 1).$$

We can indeed see, on the examples above, that for a given value of $k_v(H)$, a hypergraph is the more "aerated" the smaller $\Sigma(H)$ is.

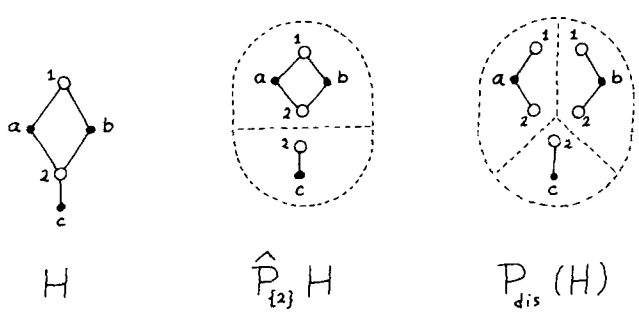


FIG. 15. Example of a hypergraph where $\Sigma(H) \neq R[P_{\text{dis}}(H)]$: $\Sigma(H) = k_v(H) = R[\hat{P}_{|2|} H] = 1$, $R[P_{\text{dis}}(H)] = \frac{3}{2}$.

4.2 Partitions minimizing σ_κ

Definition: We say that a partition $P(H) \in \pi(H)$ is κ-(σ) *minimal* if it belongs to $\pi^0_\kappa(H)$, i.e., if

$$\sigma_\kappa[P(H)] \leqslant \sigma_\kappa[P'(H)] \tag{4.6}$$

for all partitions $P'(H) \in \pi(H)$.

Remark 4.1: If $P(H)$ is κ-(σ) minimal, so must be any subset of $P(H)$ in view of (4.1). In particular, each element of $P(H)$ must be a κ-(σ) connected hypergraph [since $\{H'\}$ is κ-(σ) minimal iff H' is κ-(σ) connected].

We will now show that $\pi^0_\kappa(H)$ is a sublattice of the ($\leqslant$) lattice $\pi(H)$ of partitions of H, and that every partition in $\pi^0_\kappa(H)$ is a subpartition of every partition in $\pi^0_{\kappa-\epsilon}(H)$, for any $\epsilon > 0$. We shall require the following:

Lemma 4.7: Let V_i, W_i, $i\in I$, be sets, with $V_i \subseteq W_i$ for each $i\in I$. Denote $V = \cup_{i\in I} V_i$ and $W = \cup_{i\in I} W_i$. We have

$$\sum_{i\in I} |V_i| - |V| \leqslant \sum_{i\in I} |W_i| - |W|. \tag{4.7}$$

Proof: Call $\Delta_i = W_i - V_i$, so that $W_i = V_i + \Delta_i$, $V_i \cap \Delta_i = \emptyset$, and $|W_i| = |V_i| + |\Delta_i|$. Denote $\Delta = \cup_{i\in I} \Delta_i$. We have $W = V \cup \Delta$, $|W| = |V| + |\Delta| - |V\cap\Delta|$. Thus

$$\begin{aligned}
&\sum_{i\in I} |W_i| - |W| \\
&\quad = \sum_{i\in I} (|V_i| + |\Delta_i|) - (|V| + |\Delta| - |V\cap\Delta|) \\
&\quad = \left(\sum |V_i| - |V|\right) + \left(\sum |\Delta_i| - |\Delta|\right) + |V\cap\Delta| \\
&\quad \geqslant \sum |V_i| - |V|,
\end{aligned}$$

since $\Sigma_{i\in I} |\Delta_i| - |\Delta| \geqslant 0$ and $|V\cap\Delta| \geqslant 0$. Q.E.D.

Lemma 4.8: Let $P(H) = \{H_i, i\in I\}$ be a partition of H; $G \subseteq H$ a subhypergraph of H such that $G\cap H_i \neq \emptyset$ for each $i\in I$. Then

$$\sigma_\kappa[P(H)] \underset{\geqslant}{\overset{\leqslant}{}} \sigma_\kappa[\{H\}] \underset{\Leftarrow}{\overset{\Rightarrow}{}} \sigma_\kappa[P(H)^{\mathrm{I}} G] \underset{\geqslant}{\overset{\leqslant}{}} \sigma_\kappa[\{G\}]. \tag{4.8$^{\text{a}}_{\text{b}}$}$$

Proof: Let $P(H)^{\mathrm{I}} G = \{G_i, i\in I\}$, $G_i = G\cap H_i$, and rewrite (4.8) as

$$\sum_{i\in I} |V(H_i)| - |V(H)| \underset{\geqslant}{\overset{\leqslant}{}} \kappa(|I| - 1) \qquad \underset{\Leftarrow}{\overset{\Rightarrow}{}} \sum_{i\in I} |V(G_i)| - |V(G)| \underset{\geqslant}{\overset{\leqslant}{}} \kappa(|I| - 1). \tag{4.9$^{\text{a}}_{\text{b}}$}$$

But this is immediate from (4.7), since $V(G_i) \subseteq V(H_i)$ for each $i\in I$. Q.E.D.

Lemma 4.9: Let $G \subseteq H$ be a subhypergraph of H, and $P(H)$ a partition of H.

(i) If $P(H)$ is κ-(σ) minimal, then

$$\sigma_\kappa[P(H)^{\mathrm{I}} G] \leqslant \sigma_\kappa[\{G\}].$$

(ii) If G is κ-(σ) connected, then

$$\sigma_\kappa[G^{\mathrm{I}} P(H)] \leqslant \sigma_\kappa[P(H)].$$

Proof: Let $P(H) = \{H_i, i\in I\}$. Denote $I_a = \{i | i\in I, H_i \cap G \neq \emptyset\}$, $I_b = I - I_a$, $H^a = \cup_{i\in I_a} H_i$, $H^b = \cup_{i\in I_b} H_i = H - H^a$. Denote $P^a \equiv P(H)^{\mathrm{I}} H^a = \{H_i, i\in I_a\}$, $P^b = P(H)^{\mathrm{I}} H^b = \{H_i, i\in I_b\}$. We have

$$\sigma_\kappa[P(H)] = \sigma_\kappa[P^a] + \sigma_\kappa[P^b], \tag{4.10}$$

and since $G \subseteq H^a$,

$$P^{a\downarrow} G = P(H)^\downarrow G. \tag{4.11}$$

(i) If $P(H)$ is κ-(σ) minimal, so must be P^a and P^b in view of (4.10), whence $\sigma_\kappa[P^a] \leqslant \sigma_\kappa[\{H^a\}]$. Since $G \cap H_i \neq \varnothing$ for each $i \in I_a$, we have by (4.8a), $\sigma_\kappa[P^{a\downarrow} G] \leqslant \sigma_\kappa[\{G\}]$; whence Lemma 4.9(i) in view of Eq. (4.11).

(ii) If G is κ-(σ) connected, we have $\sigma_\kappa[P^{a\downarrow} G] \geqslant \sigma_\kappa[\{G\}]$. Since $H_i \cap G \neq \varnothing$ for each $i \in I_a$, we have, by (4.8b), $\sigma_\kappa[P^a] \geqslant \sigma_\kappa[\{H^a\}]$; adding $\sigma_\kappa[P^b]$ to both sides yields Lemma 4.9(ii) [recall $G^\uparrow P(H) = \{H^a\} \cup \{H_i, i \in I_b\}$ if $G \subseteq H$]. Q.E.D.

Lemma 4.10: Let $P(H)$ be a κ-(σ) minimal partition of H, and $P'(H')$ any partition of H'[30]:

(i) if $H \supseteq H'$, then $\sigma_\kappa[P \wedge P'] \leqslant \sigma_\kappa[P']$,

(ii) if $H \subseteq H'$, then $\sigma_\kappa[P \vee P'] \leqslant \sigma_\kappa[P']$.

Proof: Let $P(H) = \{H_1, H_2, \ldots, H_p\}$, $P'(H') = \{G_j, j \in J\}$.

(i) $\sigma_\kappa[P \wedge P'] = \Sigma_{j\in J} \sigma_\kappa[P(H)^\downarrow G_j] \leqslant \Sigma_{j\in J} \sigma_\kappa[\{G_j\}]$ $= \sigma_\kappa[P'(H')]$.

(ii) $\sigma_\kappa[P \vee P'] = \sigma_\kappa[H_1^\uparrow H_2^\uparrow \cdots H_p^\uparrow P'(H')]$ $\leqslant \sigma_\kappa[H_2^\uparrow H_3^\uparrow \cdots H_p^\uparrow P'] \leqslant \cdots \leqslant \sigma_\kappa[P'(H')]$.

We used Lemma 2.4, Eq. (4.1), and Lemma 4.9 [noticing that $H_i^\uparrow P''(H') \in \pi(H')$ if $H_i \subseteq H'$, and that each $H_i \in P(H)$ is κ-(σ) connected, in view of Remark 4.1. Q.E.D.

Theorem 4.1: If $P_1(H)$ and $P_2(H)$ are κ-(σ) minimal partitions of H, so are $P_1 \wedge P_2$ and $P_1 \vee P_2$.

Proof: Since $P_1(H)$ is κ-(σ) minimal, $\sigma_\kappa[P_1 \wedge P_2] \leqslant \sigma_\kappa[P_2]$ and $\sigma_\kappa[P_1 \vee P_2] \leqslant \sigma_\kappa[P_2]$, by Lemma 4.10. But $\sigma_\kappa[P_2]$ is the minimum value of σ_κ. Q.E.D.

Thus, the set $\pi_\kappa^0(H)$ of κ-(σ) minimal partitions of a hypergraph H is a sublattice of the $(\leqslant)$ lattice $\pi(H)$ of all partitions of H. We denote by $\hat{\hat{P}}_\kappa H$ the l.u.b., and by $\hat{\underline{P}}_\kappa H$ the g.l.b. of $\pi_\kappa^0(H)$, i.e., $\hat{\hat{P}}_\kappa H$ is the coarsest κ-(σ) minimal partition of H, while $\hat{\underline{P}}_\kappa H$ is the finest.

Lemma 4.11: (i) Every element of $P(H) \in \pi_\kappa^0(H)$ is κ-(σ) connected.

(ii) Every element of $\hat{\underline{P}}_\kappa H$ is κ-(σ) Connected.

(iii) Every element of $\pi_{k+\epsilon}^0(H)$ is a $(k+1)$-(v) partition.

(iv) $\hat{\underline{P}}_k H$ is a $(k+1)$-(v) partition.

(v) $\hat{\hat{P}}_{k+\epsilon} \leqslant \hat{\underline{P}}_k \leqslant \hat{P}'_{k+1}$.

Proof: (i) was already argued in Remark 4.1. (ii): Suppose that $H_i \in \hat{\underline{P}}_k H$ is not κ-(σ) Connected; then σ_κ does not increase under some further partitioning of H_i [in view of Eq. (4.1)], i.e., $\hat{\underline{P}}_k H$ is not the finest κ-(σ) minimal partition of H. (iii) and (iv) follow from (i), (ii), and Lemma (4.4). (v) follows from (iv), Lemma 3.6, and Theorem 4.2 below. Q.E.D.

Remark: If $P(H)$ is k-(σ) minimal, no two of its parts can share more than k vertices, for otherwise σ_k would decrease when they are lumped together; likewise, two elements of $\hat{\hat{P}}_k H$ cannot share more than $k-1$ vertices, for otherwise σ_k would not increase when they are lumped together.

Theorem 4.2: Every partition in $\pi_{\kappa+\epsilon}^0(H)$ is a subpartition of every partition in $\pi_\kappa^0(H)$, which we express as

$$\pi_{\kappa+\epsilon}^0(H) \leqslant \pi_\kappa^0(H).$$

Proof: Let $P(H) = \{H_i, i \in I\} \in \pi_\kappa^0(H)$, $P'(H) = \{G_j, j \in J\} \in \pi_{\kappa+\epsilon}^0(H)$. Denote $\Theta_{ij} = H_i \cap G_j$, $I_j = \{i | i \in I, \Theta_{ij} \neq \varnothing\}$, $K_j = \Sigma_{i\in I_j} H_i$, $P_j = P_j(K_j) = \{H_i, i \in I_j\}$. We will show that $|I_j| = 1$ for each $j \in J$, implying $P'(H) \leqslant P(H)$. We have, noticing that $\{\Theta_{ij}, i \in I_j\} = P(H)^\downarrow G_j \in \pi(G_j)$,

$$\Delta[P_j] = \sum_{i\in I_j} |V(H_i)| - |V(K_j)| \geqslant \sum_{i\in I_j} |V(\Theta_{ij})| - |V(G_j)|$$
$$= \Delta[P(H)^\downarrow G_j] \geqslant (\kappa+\epsilon)(|I_j| - 1); \tag{4.12}$$

the first $\geqslant$ is because $V(\Theta_{ij}) \subseteq V(H_i)$ for each $i \in I_j$, and Lemma 4.7; the second $\geqslant$ because G_j is $(\kappa+\epsilon)$-(σ) connected, by Lemma 4.11(i) [cf. Eq. (4.4b)]. In view of Remark (4.1), P_j is κ-(σ) minimal, so that

$$\sigma_\kappa[P_j] \leqslant \sigma_\kappa[\{K_j\}] \Leftrightarrow \Delta[P_j] \leqslant \kappa(|I_j| - 1). \tag{4.12'}$$

Combining Eqs. (4.12) and (4.12'), we get

$$(\kappa+\epsilon)(|I_j| - 1) \leqslant \kappa(|I_j| - 1),$$

i.e., $|I_j| \leqslant 1$, whence $|I_j| = 1$ since $|I_j| \geqslant 1$. Q.E.D.

We do not know whether a simple formula exists for $\sigma_\kappa^0(H)$, nor whether it is possible to simply characterize (other than by its definition) the sublattice $\pi_\kappa^0(H)$ for arbitrary values of κ. However, the cases $\kappa \leqslant 1$ are easily solved.

4.2.1 Cases $\kappa \leqslant 1$

Lemma 4.12[31]: (i) $\pi_\kappa^0(H) = \{\{H\}\}$ and $\sigma_\kappa^0(H) = |V(H)| - \kappa$ if $\kappa < 0$.

(ii) $\pi_0^0(H) = \mathscr{S}[\hat{P}_1 H]$, the set of all superpartitions of $\hat{P}_1 H$, and $\sigma_0^0(H) = |V(H)|$.

(iii) $\pi_\kappa^0(H) = \{\hat{P}_1 H\}$ and $\sigma_\kappa^0(H) = |V(H)| - \kappa|\hat{P}_1 H|$ if $0 < \kappa < 1$.

(iv) $\pi_1^0(H) = \mathscr{S}_c[\hat{P}_2 H]$, the set of all c-superpartitions of $\hat{P}_2 H$, and $\sigma_1^0(H) = |V(H)| - |\hat{P}_1 H|$.

(v) $\hat{P}_0 = \hat{P}'_0 = \hat{\hat{P}}_0 = \hat{1}$, $\hat{P}_1 = \hat{P}'_1 = \hat{\underline{P}}_0 = \hat{\hat{P}}_1$, $\hat{P}_2 = \hat{P}'_2 = \hat{\underline{P}}_1$.

Proof: (ii): For any partition $P(H) = \{H_i, i \in I\}$, we have $\sigma_0[P(H)] = \Sigma_{i\in I} |V(H_i)| \geqslant |V(H)|$, the equality holding iff the sets $V(H_i)$, $i \in I$, are pairwise disjoint, i.e., iff $P(H) \geqslant \hat{P}_1 H$. (iv): In view of Lemma 4.1, we have, for any $P(H)$,

$$\sigma_1[P(H)] \equiv \bar{\sigma}_1(F[P(H)]) \geqslant |V(F)| - |\hat{P}_1 F|$$
$$\geqslant |V(H)| - |\hat{P}_1 H|, \tag{4.13}$$

the first equality holding iff F is a forest; the second inequality follows from $V(F) = V(H)$ and $|\hat{P}_1 F| \leqslant |\hat{P}_1 H|$, the equality holding iff $P(H) \leqslant \hat{P}_1 H$, in view of lemma 3.5. Thus, in view of Lemma 3.12, both equalities in Eq. (4.13) hold iff $P(H)$ is a c-superpartition of $\hat{P}_2 H$. (v) follows from (ii), (iv), and Lemma 3.8; (i) and (iii) from (v), Lemma 4.11, and Theorem 4.2.[32] Q.E.D.

Remark: $\sigma_\kappa^0(H) = |V(H)| - \kappa|\hat{P}_1 H|$ for $0 \leqslant \kappa \leqslant 1$, but not in general for $\kappa \geqslant 1$ [e.g., in Fig. 16, $\sigma_2^0(H) = 3 < |V(H)| - 2|\hat{P}_1 H| = 4$].

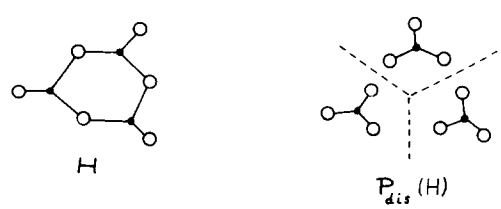


FIG. 16. $\pi_2^0(H) = \{P_{\text{dis}}(H)\}$, $\sigma_2^0(H) = \sigma_2[P_{\text{dis}}(H)] = 3$.

4.3 Recapitulation

We considered three kinds of partitions of a hypergraph H: $\hat{P}_k H$ the partition into subsets of edges pairwise k-(v) linked in H; $\hat{P}'_k H$ the coarsest partition into k-(v) connected blocks; $\pi^0_\kappa(H)$ the sublattice of κ-(σ) minimal partitions.

We have the partial ordering shown in Fig. 17. For any particular hypergraph, the above sequence of partitions becomes uninteresting after the discrete partition is reached [since $P_{\rm dis}(H)$ is the finest partition of H, so that all further partitions in the sequence are also equal to $P_{\rm dis}(H)$].

4.4 Special cases

If H is a forest, then $\hat{P}_2 H = P_{\rm dis}(H)$, and we have

Lemma 4.13: If H is a forest, then $\pi^0_\kappa(H) = \{P_{\rm dis}(H)\}$ for $\kappa > 1$, and $\sigma^0_\kappa(H) = |V(H)| - (\kappa-1)|H| - |\hat{P}_1 H|$ for $\kappa \geqslant 1$.

Proof: Follows from $\pi^0_\kappa(H) \leqslant \{\hat{P}_2 H\}$ if $\kappa > 1$, $\hat{\underline{P}}_1 H = \hat{P}_2 H = P_{\rm dis}(H)$, Eq. (4.2), and Lemma 4.1.

If H is a graph, we have $\hat{P}_3 H = P_{\rm dis}(H)$, since each edge contains at most two vertices, and is therefore at most 2-(v) linked to any other edge; thus

Lemma 4.14: If H is a graph (without isolated edges or vertices), we have (i) $\hat{\underline{P}}_2 H = P_{\rm dis}(H)$; (ii) $\pi^0_\kappa(H) = \{P_{\rm dis}(H)\}$ for $\kappa > 2$; (iii) $\sigma^0_\kappa(H) = 2n_2 + n_1 - \kappa|H|$ for $\kappa \geqslant 2$, where n_i, $i = 1,2$, is the number of edges containing i vertices [if H contains only edges of degree 2, then $n_1 = 0$, $n_2 = |H|$, and $\sigma^0_\kappa(H) = (2-\kappa)|H|$].

Proof: (i) follows from $\hat{\underline{P}}_2 H \leqslant \hat{P}_3 H = P_{\rm dis}(H)$, (ii) from $\pi^0_\kappa(H) \leqslant \{\hat{\underline{P}}_2 H\} = \{P_{\rm dis}(H)\}$ if $\kappa > 2$, and (iii) from Eq. (4.2). Q.E.D.

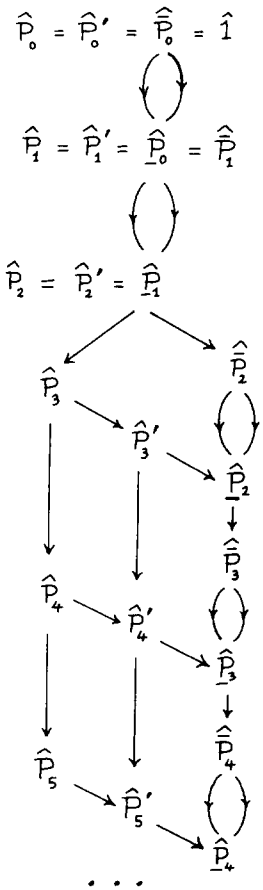


FIG. 17. Partial ordering of the partition operators $\hat{P}_k$, $\hat{P}'_k$, $\hat{\underline{P}}_k$, and $\hat{\underline{P}}_k$. $\hat{P}_a \to \hat{P}_b \Leftrightarrow P_a \geqslant P_b$. $\hat{\underline{P}}_k \supset \hat{\underline{P}}_k$ symbolizes the lattice π^0_k of k-(σ) minimal partitions. The lattices π^0_κ for $k < \kappa < k+1$ are located between $\hat{\underline{P}}_k$ and $\hat{\underline{P}}_{k+1}$.

4.5 Partitions minimizing σ'_κ

Given a partition $P(H) = \{H_i, i \in I\}$ of a hypergraph H, let us denote

$$\sigma'_\kappa[P(H)] \equiv \sigma_\kappa[\hat{P}_1 P(H)] = \sum_{i \in I} (|V(H_i)| - \kappa|\hat{P}_1 H_i|)$$

(the second line follows from $\hat{P}_1 P(H) = \Sigma_{i \in I} \hat{P}_1 H_i$ $\Rightarrow \sigma_\kappa[\hat{P}_1 P(H)] = \Sigma_{i \in I} \sigma_\kappa[\hat{P}_1 H_i]$, and $\sigma_\kappa[\hat{P}_1 H] = |V(H)| - \kappa|\hat{P}_1 H|$ for any H).

Lemma 4.15: (i) $\sigma'_\kappa[\hat{P}_1 P(H)] = \sigma'_\kappa[P(H)]$ for any partition $P(H)$;

(ii) $P = \Sigma_{j \in J} P_j \Rightarrow \sigma'_\kappa[P] = \Sigma_{j \in J} \sigma'_\kappa[P_j]$;

(iii) $\sigma'_\kappa[\{H\}] = \Sigma_{H' \in \hat{P}_1 H} \sigma'_\kappa[\{H'\}]$ for any hypergraph H.

Proof: (i) since $\hat{P}_1 \hat{P}_1 = \hat{P}_1$; (ii) from Eq. (4.1) and $P = \Sigma_{j \in J} P_j \Rightarrow \hat{P}_1 P = \Sigma_{j \in J} \hat{P}_1 P_j$; (iii) from $\sigma'_\kappa[\{H\}] = \sigma'_\kappa[\hat{P}_1 H]$ and (ii) applied to $\hat{P}_1 H = \Sigma_{H' \in \hat{P}_1 H} \{H'\}$. Q.E.D.

Let us denote by

$$\pi'^i_\kappa(H), \quad i = 0,1,\ldots,n'_\kappa(H)$$

the σ'_κ-equivalence classes of $\pi(H)$, and call

$$\sigma'^i_\kappa(H) = \sigma'_\kappa[P(H)], \quad P(H) \in \pi'^i_\kappa(H),$$

the indexing being such that $\sigma'^i_\kappa(H) < \sigma'^j_\kappa(H)$ if $i < j$. Thus, $\pi'^0_\kappa(H)$ is the minimum value of σ'_κ over $\pi(H)$ and $\pi'^0_\kappa(H)$ is the set of partitions of H which minimize σ'_κ.

Remark: If $P(H) \in \pi'^i_\kappa(H)$, then $\hat{P}_1^{-1} \hat{P}_1(H) \subseteq \pi'^i_\kappa(H)$, in view of Lemma 4.15(i).[33]

Lemma 4.16: For every $i \in \{0,1,\ldots,n'_\kappa(H)\}$, there is a $j \in \{0,1,\ldots,n_\kappa(H)\}$ such that

$$\pi'^i_\kappa(H) = \hat{P}_1^{-1} \pi^j_\kappa(H) \quad \text{and} \quad \sigma'^i_\kappa(H) = \sigma^j_\kappa.$$

Thus, the σ'_κ-partition of $\pi(H)$ is a superpartition of the σ_κ-partition of $\pi(H)$.

Proof: For every $P(H) \in \pi'^i_\kappa(H)$, we have $\hat{P}_1^{-1} \hat{P}_1 P(H) \subseteq \pi'^i_\kappa$; but $\hat{P}_1 P(H) \in \pi^j_\kappa(H)$ for some $j \in \{0,\ldots,n_\kappa(H)\}$, whence $\pi'^i_\kappa \subseteq P_1^{-1} \pi^j_\kappa$ and $\sigma'^i_\kappa(H) = \sigma'_\kappa[P(H)] = \sigma_\kappa[\hat{P}_1 P(H)] = \sigma^j_\kappa(H)$. But for every $P(H) \in \hat{P}_1^{-1} \pi^j_\kappa$, $\sigma'_\kappa[P(H)] = \sigma_\kappa[\hat{P}_1 P(H)] = \sigma^j_\kappa = \sigma'^i_\kappa$, implying $\hat{P}_1^{-1} \pi^j_\kappa \subseteq \pi'^i_\kappa$. Q.E.D.

Remark: In general, $\hat{P}_1^{-1} \pi^j_\kappa(H)$ will be empty for some values of $j \in \{0,\ldots,n_\kappa(H)\}$ [note that $\hat{P}_1^{-1} \pi^j_\kappa(H) \neq \emptyset$ iff $\pi^j_\kappa(H)$ contains at least one c-partition].

Lemma 4.17: (i) $\pi'^0_\kappa(H) = \mathscr{S}[\hat{P}_1 H]$ and $\sigma'^0_\kappa(H) = |V(H)| - \kappa|\hat{P}_1 H|$ if $\kappa < 0$.

(ii) $\pi'^0_\kappa(H) = \hat{P}_1^{-1} \pi^0_\kappa(H)$ and $\sigma'^0_\kappa(H) = \sigma^0_\kappa(H)$ if $\kappa \geqslant 0$.

Proof: In view of Lemma 4.11, $\pi^0_\kappa(H)$ contains at least one c-partition if $\kappa \geqslant 0$, so that $\hat{P}_1^{-1} \pi^0_\kappa$ is not empty, whence (ii) by Lemma 4.16. (i): Let $\kappa < 0$, so that $\pi^0_\kappa(H) = \{\{H\}\}$ by Lemma 4.12. (a) If H is connected, then $\hat{P}_1^{-1}\{\{H\}\} \neq \emptyset$ so that $\pi'^0_\kappa(H) = \hat{P}_1^{-1}\{\{H\}\} = \{\{H\}\}$. (b) If H is not connected, then $\hat{P}_1^{-1}\{\{H\}\} = \emptyset$ so that here $\pi'^0_\kappa = \hat{P}_1^{-1} \pi^j_\kappa$ for some $j \geqslant 1$. Let us find the partitions $P(H) = \{H_i, i \in I\}$ which minimize $\sigma'_\kappa[P(H)] = \Sigma_{i \in I} |V(H_i)| - \kappa|I|$ with $\kappa < 0$. In view of Lemma 4.15(i), the set of partitions minimizing σ'_κ equals $\hat{P}_1^{-1}$ applied on the set of c-partitions minimizing σ'_κ. Now, $P(H) \in \pi_c(H)$ implies $P(H) \leqslant \hat{P}_1 H$, so that $-\kappa|I|$ is minimized

within $\pi_\kappa(H)$ by $\hat{P}_1 H$; but so is $\Sigma_{i\in I}|V(H_i)|$. Whence $\pi'^0_\kappa(H) = \hat{P}_1^{-1}\hat{P}_1 H$. Combining (a) and (b), we have $\pi'^0_\kappa(H) = \hat{P}_1^{-1}\hat{P}_1 H = \mathscr{S}[\hat{P}_1 H]$ and

$$\sigma'^0_\kappa(H) = \sigma'_\kappa[\{H\}] = |V(H)| - \kappa|\hat{P}_1 H|. \qquad \text{Q.E.D.}$$

Lemma 4.18: $\pi'^0_\kappa(H) = \mathscr{S}[\hat{P}_1 H]$ if $\kappa < 1$; $\pi'^0_1(H) = \mathscr{S}[\hat{P}_2 H]$.

Proof: Follows from Lemmas 4.12, 4.17, and 2.6(iii) [noticing that $\hat{P}_2 H \in \pi_c(H)$].

Remark 4.2: In the cases where $\pi^0_\kappa(H) = \{P_{\text{dis}}(H)\}$ (see Sec. 4.4), $\pi'^0_\kappa(H) = \hat{P}_1^{-1}\{P_{\text{dis}}(H)\}$ is the set of all the partitions of H whose parts H_i consist of pairwise vertex-disjoint edges, i.e., any vertex appears at most once in each $H_i \in P(H) \in \pi'^0_\kappa(H)$.

5. CUMULANTS OVER HYPERGRAPHS

5.1 Cumulants[6,7]

Let $\{x_i, i\in I\}$ be a set of stochastic variables, and denote by $\langle(\cdots)\rangle$ the statistical average, i.e., $\langle f(\{x_i, i\in I\})\rangle$ is the expectation of the function $f(\{x_i\})$; in particular, $\langle 1\rangle = 1$. Two sets of stochastic variables $\{x_i, i\in I\}$ and $\{y_j, j\in J\}$ are statistically *independent* if, for any two functions $f(\{x_i, i\in I\})$ and $g(\{y_j, j\in J\})$, we have

$$\langle f(\{x_i\})g(\{y_j\})\rangle = \langle f(\{x_i\})\rangle\langle g(\{y_j\})\rangle. \tag{5.1}$$

The following combination of multivariate moments is called a *cumulant*:

$$\left\langle \prod_{i\in I} x_i \right\rangle_{c\{x\}} \equiv \sum_{P\in\pi(I)} (-)^{|P|-1}(|P|-1)! \prod_{J\in P} \left\langle \prod_{j\in J} x_j \right\rangle, \tag{5.2}$$

the sum being over all partitions of the set I. E.g., denoting $\langle x_i\rangle = M_i$, $\langle x_i x_j\rangle = M_{ij}$, etc.,

$$\begin{aligned} &\langle x_i\rangle_{c\{x\}} = \langle x_i\rangle, \quad \langle x_i x_j\rangle_{c\{x\}} = \langle x_i x_j\rangle - \langle x_i\rangle\langle x_j\rangle, \\ &\langle x_1 x_2 x_3\rangle_{c\{x\}} \\ &\quad = M_{123} - M_{12}M_3 - M_1 M_{23} - M_{13}M_2 + 2M_1 M_2 M_3. \end{aligned} \tag{5.3}$$

The $\{x\}$ in subscript indicates the variables with respect to which the cumulant is built; e.g., if we set $f_1 = x_1 x_2$ and $f_2 = x_3$, then

$$\langle f_1 f_2\rangle_{c\{f\}} = \langle f_1 f_2\rangle - \langle f_1\rangle\langle f_2\rangle$$

is not the same as $\langle f_1 f_2\rangle_{c\{x\}} = \langle x_1 x_2 x_3\rangle_{c\{x\}}$.

Cumulants occur in a wide variety of probabilistic and statistical mechanical problems. They owe their usefulness to the following three properties:

Theorem 5.1[6,7]: (i) A cumulant is invariant under addition of constants to its arguments, i.e.,

$$\left\langle \prod_{i\in I} x_i \right\rangle_{c\{x\}} = \left\langle \prod_{i\in I} (x_i + c_i) \right\rangle_{c\{x+c\}}, \quad |I|\geqslant 2,$$

where c_i, $i\in I$, are constants.

(ii) A cumulant vanishes if its arguments separate into two or more statistically independent subsets, i.e., if Eq. (5.1) holds, then

$$\left\langle \prod_{i\in I} x_i \prod_{j\in J} y_j \right\rangle_{c\{x,y\}} = 0.$$

(iii)

$$\begin{aligned} &\ln\left\{\left\langle \exp\left(\sum_{i\in I} x_i\right)\right\rangle\right\} \\ &\quad = \left\langle \exp\left(\sum_{i\in I} x_i\right) - 1\right\rangle_{c\{x\}} \\ &\quad = \sum_{i\in I}\langle x_i\rangle_{c\{x\}} + \frac{1}{2}\sum_{i,j}\langle x_i x_j\rangle_{c\{x\}} + \cdots. \end{aligned}$$

The notion of cumulant can be generalized as follows: given a set

$$\{M_{I'}, I'\subseteq I, I'\neq\varnothing\}$$

of objects indexed by subsets of a set I, the "cumulant" $(M_I)_c$ is defined as

$$(M_I)_c \equiv \sum_{P\in\pi(I)} (-)^{|P|-1}(|P|-1)! \prod_{I'\in P} M_{I'}. \tag{5.4}$$

If $M_{I'} = \langle \Pi_{i\in I'} x_i\rangle$, then $(M_I)_c = \langle \Pi_{i\in I} x_i\rangle_{c\{x\}}$. Adapting Theorem 5.1(ii) we have

Lemma 5.1: If $M_{I'+J'} = M_{I'}M_{J'}$ for all nonempty $I'\subseteq I$ and $J'\subseteq J$, then $(M_{I+J})_c = 0$.

Example 5.1: Let $M_{I'} = \langle \Pi_{i\in I'} x_i\rangle$ where $\langle\ \rangle$ is an unormalized averaging operation, i.e., $\langle 1\rangle \neq 1$. Denoting $(M_I)_c \equiv \langle \Pi_{i\in I} x_i\rangle_{c\{x\}}$, we have[34] $\langle(\Pi_{i\in I} x_i)(\Pi_{j\in J}\ y_j)\rangle_{c\{x,y\}} = 0$ if $\langle(\Pi_{i\in I'} x_i)(\Pi_{j\in J'} y_j)\rangle = \langle \Pi_{i\in I'}\ x_i\rangle\ \langle \Pi_{j\in J'}\ y_j\rangle$ for all nonempty $I'\subseteq I$ and $J'\subseteq J$.

5.2 Cumulants of hypergraphs

Let $\{M_{H'}, H'\subseteq H\}$ be a set of algebraic objects indexed by subhypergraphs of a hypergraph H. The construct (5.4) here becomes

$$(M_H)_c = \sum_{P\in\pi(H)} (-)^{|P|-1}(|P|-1)! \prod_{H'\in P} M_{H'}. \tag{5.5}$$

Let $\{x_E, E\in H\}$ be a set of stochastic variables indexed by the edges of a hypergraph H. Let us denote for all $H'\subseteq H$

$$H'^x \equiv \prod_{E\in H'} x_E, \quad M_{H'} \equiv \langle H'^x\rangle, \quad \langle H^x\rangle_{c\{x\}} \equiv (M_H)_c. \tag{5.6}$$

Such objects occur in many statistical-mechanical problems. The vertices of H usually correspond to particles, and $\langle\ \rangle$ to the phase space average; the x_E may represent, e.g., multiparticle interactions. Usually, two or more quantities having no particles in common will be statistically independent, i.e., $M_{H'}$ factorizes according to the connected components of H':

$$M_{H'} = \prod_{H_i\in\hat{P}_1 H'} M_{H_i} \quad \text{for all } H'\subseteq H \tag{5.7}$$

(i.e., the sets $\{x_E, E\in H_i\}$, $H_i\in\hat{P}_1 H$, are independent). We have

Lemma 5.2: If Eq. (5.7) holds, then $(M_H)_c = 0$ if H is not connected.

Proof: Let $H = H_1 + H_2$ where $V(H_1)\cap V(H_2) = \varnothing$; Eq. (5.7) implies that $M_{H'_1+H'_2} = M_{H'_1}M_{H'_2}$ for all $H'_1\subseteq H_1$ and $H'_2\subseteq H_2$, whence $(M_{H_1+H_2})_c = 0$ by Lemma 5.1. Q.E.D.

5.3 c'_κ-subcumulants

Let us now denote

$$(M_H)_{c'_\kappa} \equiv \sum_{P\in\pi'^0_\kappa(H)} (-)^{|P|-1}(|P|-1)! \prod_{H'\in P} M_{H'}, \tag{5.8}$$

where $\{\pi'^{i}_{\kappa}(H), i=0,1,\ldots,n'_{\kappa}(H)\}$ is the σ'_{κ}-partition of $\pi(H)$. Thus

$$(M_H)_c = \sum_{i=0}^{n'_\kappa} (M_H)_{c^i_\kappa}. \tag{5.9}$$

Theorem 5.2: If the condition (5.7) holds, then $(M_H)_{c^i_\kappa} = 0$ for each $i \in \{0,1,\ldots,n^i_\kappa(H)\}$ if H is not connected.

Proof: Let ϵ be a real number, and define for all $H' \subseteq H$,

$$Q_{H'} \equiv M_{H'} \epsilon^{\sigma'_\kappa[\{H'\}]}. \tag{5.10}$$

For any partition $P(H)$, we have, in view of Lemma 4.15(ii),

$$\prod_{H' \in P(H)} Q_{H'} = \epsilon^{\sigma'_\kappa[P(H)]} \prod_{H' \in P(H)} M_{H'}, \tag{5.11}$$

whence

$$(Q_H)_c = \sum_{i=0}^{n'_\kappa} (Q_H)_{c^i_\kappa} = \sum_{i=0}^{n'_\kappa} \epsilon^{\sigma'^{i}_{\kappa}(H)} (M_H)_{c^i_\kappa}. \tag{5.12}$$

Now, if Eq. (5.7) holds, then, because of Lemma 4.15(iii),

$$Q_{H'} = \prod_{H'' \in \hat{P}_1 H'} Q_{H''}.$$

This implies, by Lemma (5.2), that $(Q_H)_c = 0$ if H is not connected; but since ϵ can assume arbitrary values, each term on the rhs of Eq. (5.12) must vanish separately. Q.E.D.

The objects $(M_H)_{c^i_\kappa}$ may be called *subcumulants*, since they consist of only part of the terms constituting the cumulant $(M_H)_c$, yet still possess the important property of Lemma 5.2. If $M_H = \langle H^x \rangle$ we use the notation

$$\langle H^x \rangle_{c^i_\kappa\{x\}} = (M_H)_{c^i_\kappa}.$$

Relations of the kind (5.10) actually occur in certain statistical-mechanical problems.[4,13] More explicitly, we have

$$\langle H'^x \rangle = \epsilon^{|V(H')| - \kappa|\hat{P}_1 H'|} \langle H'^x \rangle \quad \text{for all } H' \subseteq H, \tag{5.13}$$

where ϵ is a small quantity eventually tending to zero, and $\langle\ \rangle$ is an *unormalized* averaging operation such that $\langle H'^x \rangle$ remains finite as $\epsilon \to 0$. We thus have [see Eq. (5.12)]

$$\langle H^x \rangle_{c\{x\}} = \sum_{i=0}^{n'_\kappa} \epsilon^{\sigma'^{i}_{\kappa}(H)} \langle H^x \rangle_{c^i_\kappa\{x\}}. \tag{5.14}$$

In the problems referred to, the cumulants $\langle H^x \rangle_{c\{x\}}$ appear with multiplicative factors $\epsilon^{-\sigma'^{0}_{\kappa}(H)}$, so that when the limit $\epsilon \to 0$ is taken only the term $i=0$ in Eq. (5.14) survives, i.e.,

$$\lim_{\epsilon \to 0} \epsilon^{-\sigma'^{0}_{\kappa}(H)} \langle H^x \rangle_{c\{x\}} = \langle H^x \rangle_{c^0_\kappa\{x\}}, \tag{5.15}$$

whence the necessity of determining the subcumulants $\langle H^x \rangle_{c^0_\kappa\{x\}}$.

Examples: (i) $\kappa < 1$: Here, $\pi'^{0}_{\kappa}(H) = \mathscr{S}[\hat{P}_1 H]$. Let $\hat{P}_1 H = \{C_i, i \in I\}$; then

$$\langle H^x \rangle_{c^0_\kappa\{x\}} = \sum_{P \in \pi(I)} (-)^{|P|-1} (|P|-1)! \prod_{I' \in P} \left\langle \prod_{i \in I'} C_i^x \right\rangle = \left\langle \prod_{i \in I} C_i^x \right\rangle_{c\{C^x\}} \equiv \langle H^x \rangle_{c\{C^x\}} \tag{5.16}$$

is the cumulant built on the connected components of H.

(ii) $\kappa = 1$: $\pi'^{0}_{1}(H) = \mathscr{S}[\hat{P}_2 H]$. Let $\hat{P}_2 H = \{S_j, j \in J\}$; then

$$\langle H^x \rangle_{c^0_1\{x\}} = \left\langle \prod_{j \in J} S_j^x \right\rangle_{c\{S^x\}} \equiv \langle H^x \rangle_{c\{S^x\}} \tag{5.17}$$

is the cumulant built on the 2-(v) blocks (or "stars") of H.

In the above two examples it is obvious that the subcumulants $\langle H^x \rangle_{c^0_\kappa\{x\}}$ have the property of Theorem 5.2, being equal to ordinary cumulants.

(iii) In the special cases where $\pi^0_\kappa(H) = \{P_{\text{dis}}(H)\}$ (see Sec. 4.4), we can write, in view of Remark 4.2,

$$\langle H^x \rangle_{c^0_\kappa\{x\}} = \langle H^x \rangle\!\rangle_{c\{x\}}, \tag{5.18}$$

where

$$\langle\!\langle (\cdots) \rangle\!\rangle = \langle L(\cdots) \rangle \tag{5.19}$$

and L is a "leveling" operator[6,4] which *eliminates* all products containing any vertex more than once (L is understood to act only *inside* $\langle\ \rangle$).

(iv) Let T be a tree. Since T is connected, Eq. (5.16) gives

$$\langle T^x \rangle_{c^0_\kappa\{x\}} = \langle T^x \rangle \quad \text{for } \kappa < 1. \tag{5.20}$$

Since $\hat{P}_2 T = P_{\text{dis}}(T)$, we have $\pi'^{0}_{1}(T) = \mathscr{S}[P_{\text{dis}}(T)]$ and

$$\langle T^x \rangle_{c^0_1\{x\}} = \langle T^x \rangle_{c\{x\}}. \tag{5.20'}$$

When $\kappa > 1$, $\pi'^{0}_{\kappa}(T) = \hat{P}_1^{-1} P_{\text{dis}}(T)$, so that

$$\langle T^x \rangle_{c^0_\kappa\{x\}} = \langle T^x \rangle\!\rangle_{c\{x\}}, \quad \kappa > 1, \tag{5.20''}$$

with $\langle\!\langle\ \rangle\!\rangle$ given by Eq. (5.19). If Eq. (5.7) holds, then, by Lemma B.1 of Ref. 4,

$$\langle T^x \rangle\!\rangle_{c\{x\}} = \alpha(T) \prod_{E \in T} \langle x_E \rangle,$$

where

$$\alpha(T) = (-)^{|T|-1} \prod_{v \in V(T)} (|v|-1)!.$$

6. CONCLUSION

In the statistical-mechanical problems referred to in Sec. 5.3, ϵ is the inverse of the volume, so that $\epsilon \to 0$ is the thermodynamic limit [the multiplicative factors preceding the cumulants $\langle H^x \rangle_{c\{x\}}$ are, more specifically, $(n\epsilon^{-1})^{\sigma'^{0}_{\kappa}(H)}$ where n is the particle number density]. As to κ, it has the values 0 or 1 [$\kappa = 0$ (and $\sigma'^{0}_{\kappa}(H) = |V(H)|$) in the theory of pressure broadening of spectral lines by foreign gas[13]; $\kappa = 1$ (and $\sigma'^{0}_{\kappa}(H) = |V(H)| - |\hat{P}_1 H|$) in the theory of self-broadening of spectral lines, and in the quantum virial expansion problem[4]].

The present paper is in fact an outgrowth of the demonstration of Eq. (5.15) and determination of the subcumulants $\langle H^x \rangle_{c^0_\kappa\{x\}}$, for the special cases $\kappa = 0$ and 1 required in the above-mentioned problems. These special cases can be, and have previously been,[4,13] solved in much less (though still substantial) space than the present paper. But it is more satisfying to see these isolated results set within a more general theory, which may eventually turn out to be useful for other applications, besides being of mathematical interest.

ACKNOWLEDGMENTS

I wish to thank Professor A. Kotzig and Professor I. Rosenberg for several discussions on this subject, and the latter in particular for reading the manuscript and suggesting several improvements.

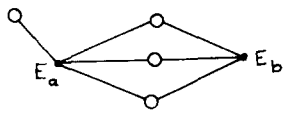


FIG. 18. $H = \{E_a, E_b\}$.

APPENDIX: MENGER'S THEOREM FOR HYPERGRAPHS

Theorem A1: Let H be a hypergraph, E_a and E_b two edges of H, $k \in \mathbb{N}$. The following two statements are equivalent:

(i) there exist at least k vertex-disjoint paths between E_a and E_b;

(ii) one must remove at least k vertices to unlink E_a and E_b.

Note that $k \leqslant \text{Min}\{|E_a|, |E_b|\}$.

Proof: (i)$\Rightarrow$(ii) is obvious. (ii)$\Rightarrow$(i): (The following proof by induction on the number of edges is a simple adaptation of the proof of Menger's theorem for graphs given in Ref. 15.) If $H = \{E_a, E_b\}$, the result is obvious (Fig. 18). Let then H have $m > 2$ edges and $k = k(E_a, E_b; H)$ be the minimum number of vertices that must be removed to unlink E_a and E_b. We assume that (ii)$\Rightarrow$(i) is true for all hypergraphs of less than m edges (induction hypothesis).

We consider two cases: Case (a): there exists a k-(v) cut[35] C unlinking E_a and E_b, not contained in E_a or E_b, i.e., $C \not\subseteq E_a$ and $C \not\subseteq E_b$. Let $\hat{P}_C H = \{H_1, H_2, \ldots, H_p\}$, with $E_a \in H_1$, $E_b \in H_2$, say. By Lemma 3.3, $V(H_1) \supseteq C$ and $V(H_2) \supseteq C$, whence $|H_1| > 1$ and $|H_2| > 1$ [$E_a \in H_1$, $H_1 \supseteq C$, and $C \not\subseteq E_a$ imply that H_1 contains other edges besides E_a]. Consider the hypergraphs $H'_1 = H_1 \cup \{E'_b\}$ and $H'_2 = \{E'_a\} \cup H_2$, where $E'_a = V(H_1)$, $E'_b = V(H_2)$; thus H'_1 is obtained from H by coalescing together the edges of H_2 into a single edge E'_b; clearly, $k(E_a, E'_b; H'_1) = k(E'_a, E_b; H'_2) = k(E_a, E_b; H) = k$. Since $|H'_1| < m$ and $|H'_2| < m$ (implied by $|H_2| > 1$ and $|H_1| > 1$), there exist, by the induction hypothesis, k vertex-disjoint paths in H'_1 between E_a and E'_b, each one passing through a vertex of C; and likewise between E'_a and E_b in H'_2. By joining at vertices of C the parts of the paths between E_a and E'_b lying inside H_1 with the parts of the paths between E'_a and E_b lying inside H_2, we obtain k vertex-disjoint paths in H between E_a and E_b.

Case (b): Each k-(v) cut unlinking E_a and E_b is completely contained in either E_a or E_b. We can assume without loss of generality that each vertex of H belongs to a k-(v) cut, since otherwise its removal would not affect the value of $k = k(E_a, E_b; H)$. It follows that the vertex sets E_a and E_b are both k-(v) cuts [since each vertex of E_a and E_b belongs to a k-(v) cut which is itself contained in either E_a or E_b]. Thus E_a and E_b both contain k vertices. Each vertex of E_a must be linked to E_b (since it belongs to a cut), whence to at least one vertex of E_b, and vice versa. This implies that the vertices $v_1, v_2, \ldots, v_k$ of E_a and the vertices $w_1, w_2, \ldots, w_k$ of E_b can be ordered in such a manner that v_i is linked to w_i for each $i = 1, \ldots, k$ (matching). But the path between v_i and w_i contains no other vertices (besides v_i and w_i), since we excluded vertices not belonging to k-(v) cuts, i.e., to E_a or E_b (we have either $v_i = w_i$ or v_i—E'—w_i, where E' is an edge different from E_a or E_b). We thus have k vertex-disjoint paths E_a—v_i—E_b or E_a—v_i—E'—w_i—E_b, $i = 1, \ldots, k$. Q.E.D.

Corollary A1: Let h_i and h_j be any two distinct elements of $H \cup V(H)$, except a pair of adjacent edges, or a vertex and an edge incident on each other. The following two statements are equivalent:

(i) there exist at least k vertex-disjoint paths between h_i and h_j;

(ii) one must remove at least k vertices to unlink h_i and h_j.

Proof: (i)$\Rightarrow$(ii) is obvious. (ii)$\Rightarrow$(i): let $v \in V(H)$, $E_b \in H$, nonincident on each other. Coalesce together all the edges incident on v, and call the resulting edge E_v. The number of vertices that must be removed to unlink E_v and E_b is the same as that needed to unlink v and E_b; likewise the number of vertex-disjoint paths between E_v and E_b is the same as that between v and E_b (Fig. 19). Apply Theorem A1 to E_v and E_b. If v and w are two nonadjacent vertices, form E_v and E_w and use the same argument. Q.E.D.

Duals to Theorem A1 and Corollary A1 are obtained by interchanging edges and vertices. When applied to graphs, Theorem A1, Corollary A1, and their duals contain the edge form and vertex forms of Menger's theorem for graphs.

Remark: Let us write $h_i \overset{k\text{-}(e)}{-\!-\!-\!-\!-} h_j$ if there exist k edge-disjoint paths between h_i and h_j (or equivalently, if one must remove at least k edges to unlink h_i and h_j). By duality with Lemma 3.4, $\overset{k\text{-}(e)}{-\!-\!-\!-\!-}$ is an equivalence relation between vertices. The relations $\overset{k\text{-}(e)}{-\!-\!-\!-\!-}$ and $\overset{k\text{-}(v)}{-\!-\!-\!-\!-}$ are different in general; however, $\overset{2\text{-}(e)}{-\!-\!-\!-\!-}$ and $\overset{2\text{-}(v)}{-\!-\!-\!-\!-}$ are equivalent relations in hypergraphs whose edges are all of degree $\leqslant 3$ (because if an edge is an articulation, then at least one of its incident vertices is also an articulation).

We say that H is k-(e) connected iff all vertex pairs are in the relation $\underset{H}{\overset{k\text{-}(e)}{-\!-\!-\!-\!-}}$.

Corollary A2: Let h_i and h_j be any two elements of $H \cup V(H)$ not incident on each other. The following two statements are equivalent:

(i) there exist at least k disjoint (i.e., sharing neither vertices nor edges) paths between h_i and h_j;

(ii) one must remove at least k elements (edges or vertices) of $H \cup V(H)$ to unlink h_i and h_j.

The above statements are abbreviated $h_i \underset{H}{\overset{k}{-\!-\!-\!-\!-}} h_j$.

Proof: This is just a statement of Corollary A1 as applied to the bipartite graph $B[H]$ (wherein all vertices correspond to elements of $V(H) \cup H$, so that vertex-disjoint paths in $B[H]$ correspond to disjoint paths in H). Q.E.D.

Remark: The relation $\overset{k}{-\!-\!-\!-\!-}$ is not an equivalence relation in general (Fig. 20). However, it is an equivalence relation between edges in a hypergraph whose edges are all of degree $\leqslant 3$, in which case vertex-disjoint paths are also edge

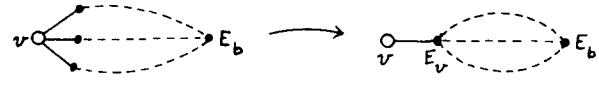


FIG. 19. The set of edges incident on vertex v are coalesced into a single edge E_v.

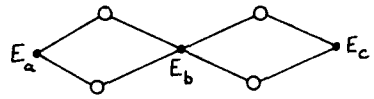


FIG. 20. Example of a hypergraph on which $\overset{2}{\text{----}}$ is not an equivalence relation; we have $E_a \overset{2}{\text{---}} E_b$ and $E_b \overset{2}{\text{---}} E_c$, but not $E_a \overset{2}{\text{---}} E_c$.

disjoint (but not the converse) (Fig. 21) [note then that all the equivalence classes of $\overset{k}{\text{-----}}$ each consist of a single edge, unless $k \leqslant 3$ (since $E_a \overset{k}{\text{-----}} E_b$ implies $k \leqslant \text{Min}\{|E_a|, |E_b|\}$)].

Lemma A1: Let E_a and E_b be two edges of the hypergraph H. If $E_a \overset{k\text{-}(v)}{\text{-----}} E_b$, then also $v_1 \overset{k\text{-}(v)}{\text{-----}} E_b$, $E_a \overset{k\text{-}(v)}{\text{-----}} v_2$, and $v_1 \overset{k\text{-}(v)}{\text{-----}} v_2$ for any vertices $v_1 \in E_a$ and $v_2 \in E_b$.

Proof: Let $\pi_1, \pi_2, \ldots, \pi_k$ be k vertex-disjoint paths between E_a and E_b. If v_1 is on one of the paths between E_a and E_b, π_1 say, we have the k vertex-disjoint paths $\pi_1' = v_1 \overset{\pi_1}{\text{-----}} E_b$ and $\pi_i' = v_1 — E_a \overset{\pi_i}{\text{-----}} E_b$, $i = 2,\ldots,k$, between v_1 and E_b [Fig. 22(a)]. If v_1 is not on one of the paths, we have the k vertex-disjoint paths $\pi_i' = v_1 — E_a \overset{\pi_i}{\text{-----}} E_b$, $i = 1,\ldots,k$. To show $v_1 \overset{k\text{-}(v)}{\text{----}} v_2$, we apply the same argument as above, with v_2 playing the role of v_1, v_1 that of E_b, and π_i' that of π_i. Q.E.D.

Remark: The converse of Lemma A1 is not true in general [Fig. 23(b)].

Corollary A3: H is k-(v) connected iff $h_i \overset{k\text{-}(v)}{\underset{H}{\text{----}}} h_j$ for all pairs of elements h_i and h_j in $\tilde{H} = H \cup V(H)$. Dually, H is k-(e) connected iff $h_i \overset{k\text{-}(e)}{\underset{H}{\text{----}}} h_j$ for all $h_i, h_j \in \tilde{H}$.

Let us say that H is k connected iff there exist k disjoint paths between every pair of elements of $\tilde{H}$. Let us also say that H is k-(vv) connected if there exist k vertex-disjoint paths between every pair of vertices in $V(H)$. Clearly, H is k connected iff $B[H]$ is k-(vv) connected. Also, in view of Lemma A1, if H is k-(v) connected, it is also k-(vv) connected (but not the converse in general).

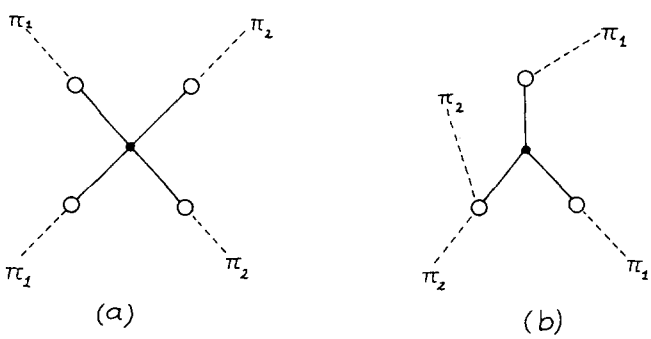


FIG. 21. (a) Two vertex-disjoint paths (π_1 and π_2) can both pass through an edge of degree >4, but (b) not through an edge of degree $\leqslant 3$.

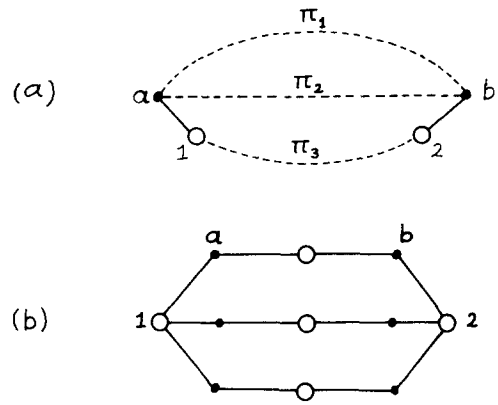


FIG. 22. (a) $E_a \overset{3\text{-}(v)}{\text{----}} E_b \Rightarrow v_1 \overset{3\text{-}(v)}{\text{----}} v_2$, but (b) $v_1 \overset{3\text{-}(v)}{\text{----}} v_2 \not\Rightarrow E_a \overset{3\text{-}(v)}{\text{----}} E_b$ [E_a and E_b are only 2-(v) linked].

[1] J. E. Mayer, J. Chem. Phys. **5**, 67 (1937); J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940).

[2] Recent works are: U. Mohanty, J. Math. Phys. **23**, 1489 (1982); K. W. Kratky, J. Stat. Phys. **27**, 533 (1982); **29**, 129 (1982); M. Lavaud, J. Stat. Phys. **27**, 593, 745 (1982).

[3] G. Stell, *Generating Functionals and Graphs*, in *Graph Theory and Theoretical Physics*, edited by F. Harary (Academic, New York, 1967); *Correlation Functions and Their Generating Functionals*, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 5B.

[4] A. Royer, J. Math. Phys. **24**, 897 (1983).

[5] (a) C. Berge, *Graphes et Hypergraphes* (Dunod, Paris, 1970) [English translation: *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1973)]; (b) J. E. Graver and M. E. Watkins, *Combinatorics with Emphasis on the Theory of Graphs* (Springer, Berlin, 1977).

[6] R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).

[7] For more mathematical discussions of cumulants, see e.g., H. Cramer, *Random Variables and Probability Distributions*, 3rd ed. (Cambridge U.P., Cambridge, 1970); R. N. Bhattacharya and R. Ranga Rao, *Normal Approximations and Asymptotic Expansions* (Wiley, New York, 1976); D. R. Brillinger, *Time Series, Data Analysis and Theory*, expanded edition (Holden-Day, San Francisco, 1981).

[8] See e.g., R. Kubo, J. Math. Phys. **4**, 174 (1963); N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981); A. Royer, J. Math. Phys. **24**, 380 (1983).

[9] See e.g., R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley, New York, 1975).

[10] R. Brout, Phys. Rev. **115**, 824 (1959).

[11] R. Brout and P. Carruthers, *Lectures on the Many Electron Problem* (Gordon and Breach, New York, 1969), Sec. 2.5 and references therein.

[12] H. D. Ursell, Proc. Cambridge Philos. Soc. **23**, 685 (1927); B. Kahn and G. E. Uhlenbeck, Physica **5**, 399 (1938).

[13] A. Royer, Phys. Rev. A **6**, 1741 (1972); **22**, 1625 (1980).

[14] See e.g., S. MacLane and G. Birkhoff, *Algebra* (MacMillan, New York, 1967).

[15] See e.g., R. J. Wilson, *Introduction to Graph Theory*, 2nd ed. (Longman, London, 1978).

[16] In Ref. 4, hypergraphs are dealt with as multisets, as this is the most convenient, notationally and operationally, for the purposes of that paper. In the present paper, however, we consider the lattice structure of the set of partitions of hypergraphs, and this requires that we deal with sets [it is possible to define partitions of a multiset, as well as the relation $\leqslant$ between such partitions (see Ref. 4), but one runs into ambiguities when trying to define the meet and join of two partitions of a multiset].

[17] The name sets are $\mathcal{N}_F = \{H_i, H_i \in P(H)\}$ and $\mathcal{N}_I = \{H_\alpha \cap H_\beta | H_\alpha, H_\beta \in P \cup P', H_\alpha \neq H_\beta, H_\alpha \cap H_\beta \neq \emptyset\}$; the vertex sets are $\mathcal{V}_F = V(H)$ and $\mathcal{V}_I = P \cup P'$, and the functions $f: \mathcal{N} \to \mathcal{P}[\mathcal{V}]$ are defined by $f_F(H_i) = V(H_i) \subseteq V(H)$ and $f_I(H_\alpha \cap H_\beta) = \{H_\alpha, H_\beta\} \subseteq P \cup P'$.

[18] Transitivity and reflexivity are obvious. Antisymmetry: $P \leqslant P' \Rightarrow H_i \subseteq H_{j_i}'$; $P' \leqslant P \Rightarrow H_{j_i}' \subseteq H_{i_{j_i}} \Rightarrow H_i \subseteq H_{i_{j_i}} \Rightarrow H_i = H_{i_{j_i}}$ (since P is a partition) $\Rightarrow H_i = H_{j_i}' \Rightarrow P = P'$.

[19] (Poset is an acronym for "partially ordered set"; see Ref. 14.) There exist

several different partial orderings of $\Pi(H)$. Examples are $P(H) \leqslant P'(H')$ if:

(i) each $H_i \in P(H)$ is contained in some $H'_j \in P'(H')$;

(ii) each $H'_j \in P'(H')$ contains some $H_i \in P(H)$;

(iii) both (i) and (ii) hold;

(iv) (i) and/or (ii) hold.

All these correspond to the usual partial ordering of partitions if $H = H'$ [i.e., on $\pi(H)$], but define different partial orderings on $\Pi(H)$. That chosen in the text [(i) above] is the most convenient for our purposes, as it leads to Lemma 2.4 and 4.10 which are used for demonstrating Theorem 4.1.

[20] If we write $(\underset{P}{=}) \leqslant (\underset{P'}{=})$ if $P(H) \leqslant P'(H')$ (implying $H \subseteq H'$), then the set of equivalence relations on subsets of $\mathbb{H}$ constitute a $(\leqslant)$ lattice.

[21] $\hat{P}^{\downarrow}\hat{H} = \hat{P}^{\downarrow}\{H\} = \hat{P}H$; $\mathscr{O}^{\downarrow}\hat{}^{\downarrow}P = \Sigma_{H_i \in P}\,\mathscr{O}^{\downarrow}\hat{}H_i = \Sigma_{H_i \in P}\,\mathscr{O}^{\downarrow}\{H_i\}$ $= \mathscr{O}^{\downarrow}\,\Sigma_{H_i \in P}\{H_i\} = \mathscr{O}^{\downarrow}P$, the third equality holding if $\mathscr{O}^{\downarrow}$ is linear.

[22] In fact, $\underset{H\text{-}W}{\text{- - - - -}}$ is an equivalence relation on $H \cup (V(H) - W)$.

[23] $\underset{H}{\overset{k\text{-}(v)}{\text{- - - - -}}}$ is an equivalence relation on H, but in general not on $H \cup V(H)$ (the argument in the proof of Lemma 3.4 does not work if E_a and/or E_b can belong to one of the (v) cuts]. E.g., in Fig. 10, $E_a \overset{2\text{-}(v)}{\underset{H_1}{\text{- - - -}}} E_b$ and $H_b \overset{2\text{-}(v)}{\underset{H_1}{\text{- - - -}}} E_c$ imply $H_a \overset{2\text{-}(v)}{\underset{H_1}{\text{- - - -}}} E_c$; but we have $v_1 \overset{2\text{-}(v)}{\underset{H_2}{\text{- - - -}}} v_2$ and $v_2 \overset{2\text{-}(v)}{\underset{H_2}{\text{- - - -}}} v_3$, yet not $v_1 \overset{2\text{-}(v)}{\underset{H_2}{\text{- - - -}}} v_3$.

[24] *Remark*: Let $P(H) = \{H_1, H_2, \ldots, H_p\}$. If $p = 2$, then $V(H_1) \cap V(H_2)$ is obviously a (v) cut of H, so that $k_v(H) \leqslant |V(H_1) \cap V(H_2)|$. But if $p \geqslant 2$, we do not necessarily have $|V(H_i) \cap V(H_j)| \geqslant k_v(H)$. E.g., let H be as in Fig. 8(a), and let $P(H) = \hat{P}_{\{1,2,3\}}H = \{\{a\},\{b\},\{c\}\}$; we have $|V(H_i) \cap V(H_j)| = 1 < k_v(H) = 2$ [note that $\{1,2,3\}$ is not a (v) cut of H, so that Eq. (3.5) does not apply].

[25] Alternatively, if we define an "operator" $\hat{Q}_k$ by $\hat{Q}_k H = \hat{P}_C H$ where C is *any* (v) cut of H of size less than k, then $\hat{P}'_k = \hat{Q}_k^{\infty}$.

[26] We have $\hat{P}_{C_1} \wedge \hat{P}_{C_2} \wedge \cdots \wedge \hat{P}_{C_m}H = \hat{P}_{C_1}\hat{P}_{C_2}\cdots\hat{P}_{C_m}H = \hat{P}_{C_1 \cup C_2 \cup \cdots \cup C_m}H$ if the sets C_i, $i = 1,\ldots,m$, are all (v) cuts of size 1; otherwise, only the weaker result (3.2) holds in general [e.g., in Fig. 9, $\pi_c(H_1) \not\ni \hat{P}_3 H_1 = \hat{P}_{C_1} \wedge \hat{P}_{C_2} \wedge \hat{P}_{C_3}H_1 \neq \hat{P}_{C_1 \cup C_2 \cup C_3}H_1 \in \pi_c(H_1)$ (by definition, $\hat{P}_W H \in \pi_c(H)$ for any vertex set W)]. The above equalities are due to the fact that when a (v) cut of size 1 is performed, no new cuts are created, i.e., the 2-(v) blocks of H get separated from one another without getting "opened up." Note also that the relation $\hat{P}'_k H = \hat{P}_{C_1 \cup C_2 \cup \cdots \cup C_m}$ where C_i, $i = 1,\ldots,m$, are all the (v) cuts of H of size $< k$, true for $k \leqslant 2$, does not hold in general if $k \geqslant 3$ [e.g., in Fig. 9, $\hat{P}'_3 H_1 = \hat{P}_{C_1 \cup C_2 \cup C_3}H_1$, but $\hat{P}'_3 H_2 = P_{\mathrm{dis}}(H_2) \neq \hat{P}_{C'_1 \cup C'_2 \cup C'_3}H_2 = \hat{P}_3 H_2$].

[27] Given a path π, we write $\pi \subset H$ if each edge in π is also in H.

[28] If $k_v(H) = 1$, then H has at least one articulation vertex, v say, and $\Sigma(H) = R[\hat{P}_{\{v\}}H] = 1$.

[29] Consider the discrete partition of a ring hypergraph. If we form a new c-partition by lumping together two adjacent edges, then Δ and $|P|$ both decrease by 1, so that R increases; likewise if we lump a third edge with the preceding pair, and so on. Thus, $\Sigma(H) = R[P_{\mathrm{dis}}(H)]$

[30] Note that the condition $H \supseteq H'$ (resp. $H \subseteq H'$) implies that $P \wedge P'$ (resp. $P \vee P'$) is a partition of H'.

[31] The singularities of the function $\pi^0_\kappa(H)$: $\mathbb{R} \to \pi(H)$ at $\kappa = 0$ and 1 appear less surprising when we observe that in $\sigma_\kappa[P(H)] = \sigma_0[P(H)] - \kappa|P(H)|$, the first term is constant inside $\pi^0_0(H) = \mathscr{S}[\hat{P}_1 H]$, while $-\kappa|P(H)|$ is minimized inside $\pi^0_0(H)$ by $\hat{P}_1 H$ alone if $\kappa > 0$; likewise, in $\sigma_\kappa[P(H)] = \sigma_1[P(H)] - (\kappa - 1)|P(H)|$, the first term is constant inside $\pi^0_1(H) = \mathscr{S}[\hat{P}_2 H]$, while $-(\kappa - 1)|P(H)|$ is minimized inside $\pi^0_1(H)$ by $\hat{P}_1 H$ alone if $\kappa < 1$.

[32] $\hat{P}_0 = \hat{1}$ also follows from Lemma 4.5(i), which implies $\hat{1}H = \{H\} \in \pi^0_0(H)$. $\hat{P}_1 = \hat{P}'_2$ also follow from Lemma 4.5(iii), which implies $\hat{P}'_2 \geqslant \hat{P}_1$, and Lemma 4.11.

[33] Note that if $P(H)$ is a c-partition, then $\hat{P}_1^{-1}\hat{P}_1 P(H) = \hat{P}_1^{-1}P(H)$, while if $P(H) \notin \pi_c(H)$, then $\hat{P}_1^{-1}\hat{P}_1 P(H) \neq \hat{P}_1^{-1}P(H) = \varnothing$.

[34] Note that a condition like Eq. (5.1) would be self-inconsistent if $\langle 1 \rangle \neq 1$, for the case $f = g \equiv 1$ would give $\langle 1 \rangle = \langle 1 \rangle\,\langle 1 \rangle$.

[35] We call k-(v) *cut* any (v) cut C of size $|C| = k$.

Determination of point group harmonics for arbitrary j by a projection method.[a] I. Cubic group, quantization along an axis of order 4

R. Conte
Service de Physique du Solide et de Résonance Magnétique, Cen-Saclay, 91191 Gif sur Yvette Cedex, France

Jacques Raynal
Service de Physique Théorique, Cen-Saclay, 91191 Gif sur Yvette Cedex, France

E. Soulié
Département de Physico-Chimie, Cen-Saclay, 91191 Gif sur Yvette Cedex, France

(Received 10 May 1983; accepted for publication 2 September 1983)

We give a method for systematically building cubic harmonics, i.e., base vectors of irreducible representations of the group of invariance of a cube. Applications include solid state physics, molecular chemistry, spectroscopy, etc. The only necessary operators are C_4^z and R_y, rotation of $\pi/2$ around $0y$. Projectors on representations are constructed; then their rows are orthogonalized by a Schmidt method and normalized simply by dividing by the square root of a diagonal element.

PACS numbers: 02.20. + b

INTRODUCTION

The knowledge of linear combinations of spherical harmonics [or more generally of the standard vectors of an irreducible representation of SU(2)] that are invariant under the action of elements of a point group is essential to the efficient solution of numerous problems in solid state physics, molecular spectroscopy, electronic structure, and vibrational structure. The determination of these linear combinations derives trivially from the reduction of the successive irreducible representations of the rotation group O(3) or SU(2) (corresponding respectively to integer or half-integer values of the quantum number j) by all point groups but those of the cube (and therefore tetrahedron) and icosahedron.

Since Bethe[1] computed cubic harmonics for crystal field calculations, the need for such harmonics has been encountered in quite different areas (for a list of applications, see, e.g., Ref. 2). In the field of molecular spectroscopy, molecules such as CH_4 and SF_6 in the gas phase exhibit cubic point symmetry and cubic harmonics are required for integer values of j up to 150.[3] In the field of solid state physics, the degeneracy of d^n or f^n configurations is removed by the combination of electrostatic repulsion, spin–orbit coupling, and crystal field. When this crystal field has cubic symmetry, the numerical calculations involved in the determination of eigenvalues and eigenvectors of the Hamiltonian are considerably reduced by the block decomposition of its representing matrix into much smaller blocks. This reduction requires that cubic harmonics be chosen as base vectors, and, moreover, the use of the fictitious angular momentum[4] simplifies the calculation of magnetic properties (see, e.g., Ref. 5). For instance, in the case of ions in f^n configurations, values of j may be required up to 25/2 in order to describe ground states.

Since 1929, when Bethe published his method and tables for integer values of j up to 4, the determination of cubic harmonics has been tackled repeatedly by numerous authors for the cube [6–18,2] and by a few ones for the icosahedron (see references in Ref. 19). The half-integer case, for which there are no associated harmonic polynomials in Cartesian coordinates, has been dealt with very rarely[9,10,20,21] or incompletely,[22–24] as far as tables are concerned. A great variety exists for the presentation of tables: representations can be listed as functions of standard vectors $|jm\rangle$, Cartesian coordinates, spherical coordinates, or monopoles, i.e., partial derivatives of $1/r$, the coefficients being either rational[1,6–8,13–15,17,18,20,21] or real.

Unaware of previous work, mainly that of Jahn[6] (integer $j \leqslant 10$) and Hecht[8] (integer $j = 11,12,13$), a lot of authors[5,25–37] have published already known tables.

Much more interesting than that historical survey are the methods used. Cubic harmonics have been determined in one of three ways, namely:

(1) As orthogonal sets of partial derivatives of $1/r$ having the symmetry of the desired Γ_i representation: this method has been devised at the same period by Elert[38] and Bethe,[1] used by Jahn,[6] well formulated by von der Lage and Bethe,[25] and greatly improved by Moret-Bailly.[27] Its main disadvantage, besides its tediousness, is its built-in inability to handle half-integer representations.

(2) As eigenvectors of the homogeneous polynomial in J_+, J_z, J_- of lowest degree (i.e., 4, apart from 0) which transforms like a Γ_1 representation, i.e., which is completely invariant under action of the group. This operator-equivalent technique, which handles in the same way integer and half-integer representations, has been sketched for by Kynch,[22] nicely formulated by Stevens,[39] and mainly applied by Lea, Leask, and Wolf.[9] One disadvantage is that, in case a representation occurs more than once for a given j, coefficients of eigenvectors are not the square root of rational numbers. Many authors were interested only in the eigenvalues.[22,40–42]

(3) As eigenvectors of projection operators: Wigner[43] defined general projectors and transfer operators using group characters, then Melvin[44] and McIntosh[45] expressed them as a product of two factors, Altmann[46] gave relatively

[a] Preliminary results of this paper were given in N. Boccara, "Symmetries and broken symmetries in condensed matter physics," Idset 1981, Paris (proceedings of Colloque Pierre Curie, Paris, 1980), poster session.

 0022-2488/84/051176-11$02.50

short expressions for these projectors and diagonalized them numerically, Fox and Ozier[47] and Fox and Krohn[2] made various improvements to the numerical work. None of these authors applied the method to half-integer representations, probably because molecular spectroscopy requires only integer representations.

A fourth method must be mentioned which may be very promising: it is that of orthogonal polynomials, generating functions and so on. Indeed, as noted by Betts, Bhatia, and Wyman,[7] in the case of even j, polynomials of x, y, z expressing Γ_1 representations, when divided by z^j, are particular Appell polynomials in the two variables $x^2/z^2, y^2/z^2$. Using generating functions techniques, Puff[15] obtained a closed form formula, with two summations, for Γ_1 representations in spherical coordinates. Dunkl[48] used Krawtchouk polynomials in variable $j-m$ to express the projectors on Γ_1 and Γ_2 representations in the standard basis $|jm\rangle$, but no attempt was made to build orthogonal sets in case of degeneracy. Let us remark that, from the work of Dunkl, one can deduce, by subtraction from the identity operator, compact expressions using Krawtchouk polynomials for projectors on the two components of Γ_3. Since Γ_4 and Γ_5 are very easy to handle, only Γ_6, Γ_7, and Γ_8 need compact expressions for their projectors, from this point of view.

For summarizing, the implementation of Fox and Krohn[2] is up to now the best available algorithm.

This wide need for cubic harmonics prompted us to devise a general, efficient method. Even if it is closely related, our method is essentially different from that of Fox and Krohn.

(1) As any standard base vector of SU(2) can be projected on at most two representations of O_h, the knowledge of the behavior of these representations under one carefully chosen element of O_h is enough to obtain the projectors.

(2) This projection method gives unnormalized vectors the coefficients of which are rotation matrix elements, and it also gives the square of the norm as a constant plus a rotation matrix element.

(3) In case of degeneracy, a simple Schmidt orthogonalization of the unnormalized vectors introduces minors of rotation matrix elements of increasing order and the square of the norm is still one of them.

This method does not require any determination of the set of eigenvectors of a matrix and, therefore, does not involve a heavy equipment. Instead, it takes full advantage of the properties of projectors for both integer and half-integer values of j. This approach has two benefits: it significantly reduces the workload, and it enables an exact determination of coefficients in the form of square roots of rational numbers. Moreover, no complex numbers occur in our method even for half-integers. Two forthcoming papers solve the same kind of problem: one for the cubic group with a ternary axis of quantization[49] and the other for icosahedral harmonics.[19]

I. THE CUBIC GROUP AND ITS REPRESENTATIONS

The cubic group[50] has three fourfold rotations around three mutually orthogonal axes, which we choose as our reference frame $0xyz$. The two rotations whose Euler angles[51] (α,β,γ) are respectively $(\pi/2, 0,0)$ and $(0,\pi/2,0)$ are a system of generators of the group whose 24 elements have the following Euler angles:

$$\alpha = 0, \pi/2, \pi, 3\pi/2, \quad \beta = 0, \pi, \quad \gamma = 0, \tag{1a}$$

$$\alpha = 0, \pi/2, \pi, 3\pi/2, \quad \beta = \pi/2, \quad \gamma = 0, \pi/2, \pi, 3\pi/2. \tag{1b}$$

The double cubic group O^+ is obtained by adding to that list the 24 elements derived by addition of 2π to α.

The action of an arbitrary element on a base vector $|jm\rangle$ of an irreducible representation $\mathscr{D}_j$ of SU(2) is given by the rotation matrix having the corresponding Euler angles. The only nontrivial matrix we will have to deal with is $R(0, \pi/2, 0) = R_y$; this matrix is real; we will note its elements as $d^{(j)}_{mm'}(\pi/2)$.[51]

The eight elements (1a) define what is called the D_4 group.[50] It is suitable to introduce the subspaces of $\mathscr{D}_j$ which are invariant under D_4.

TABLES I and II. Correspondence between irreducible representations of O_h and subspaces adapted to D_4. Each of subspaces adapted to the irreducible representations of D_4, listed in first column, is the direct sum of those subspaces of irreducible representations of O_h which are listed in its row. Every case contains two notations: the one used in the present paper, adapted from Bethe[1] and the one defined by Placzek[53] and more widely known as the Mulliken[54] notation. Our notation is chosen as follows: $|\Gamma_i\, \hat{p}\rangle$ notes a component p of a respresentation Γ_i, $i = 1$ to 8, which behaves under a rotation like its associated base vectors listed in (2) and (3).

TABLE I.

O D_4	Γ_1 A_1	Γ_2 A_2	Γ_3 E	Γ_4 T_1	Γ_5 T_2
$\lvert\hat{0}+\rangle$ A_1	$\lvert\Gamma_1\hat{0}\rangle$ $\lvert A_1 a_1\rangle$		$\lvert\Gamma_3\hat{0}\rangle$ $\lvert E\theta\rangle$		
$\lvert\hat{2}+\rangle$ B_1		$\lvert\Gamma_2\hat{2}\rangle$ $\lvert A_2 a_2\rangle$	$\lvert\Gamma_3\hat{2}\rangle$ $\lvert E\epsilon\rangle$		
$\lvert\hat{1}\rangle$ E				$\lvert\Gamma_4\hat{1}\rangle$ $\lvert T_1 1\rangle$	$\lvert\Gamma_5\hat{1}\rangle$ $\lvert T_2 - 1\rangle$
$\lvert\hat{0}-\rangle$ A_2				$\lvert\Gamma_4\hat{0}\rangle$ $\lvert T_1 0\rangle$	
$\lvert\hat{2}-\rangle$ B_2					$\lvert\Gamma_5\hat{2}\rangle$ $\lvert T_2 0\rangle$
$\lvert-\hat{1}\rangle$ E				$\lvert\Gamma_4-\hat{1}\rangle$ $\lvert T_1 - 1\rangle$	$\lvert\Gamma_5-\hat{1}\rangle$ $\lvert T_2 1\rangle$

TABLE II.

$O+$ D_4^+	Γ_6 E'	Γ_7 E''	Γ_8 U'
$\lvert\frac{\hat{3}}{2}\rangle$ E''		$\lvert\Gamma_7\frac{\hat{3}}{2}\rangle$ $\lvert E''\beta''\rangle$	$\lvert\Gamma_8\frac{\hat{3}}{2}\rangle$ $\lvert U'\kappa\rangle$
$\lvert\frac{\hat{1}}{2}\rangle$ E'	$\lvert\Gamma_6\frac{\hat{1}}{2}\rangle$ $\lvert E'\alpha'\rangle$		$\lvert\Gamma_8\frac{\hat{1}}{2}\rangle$ $\lvert U'\lambda\rangle$
$\lvert-\frac{\hat{1}}{2}\rangle$ E'	$\lvert\Gamma_6-\frac{\hat{1}}{2}\rangle$ $\lvert E'\beta'\rangle$		$\lvert\Gamma_8-\frac{\hat{1}}{2}\rangle$ $\lvert U'\mu\rangle$
$\lvert-\frac{\hat{3}}{2}\rangle$ E''		$\lvert\Gamma_7-\frac{\hat{3}}{2}\rangle$ $\lvert E''\alpha''\rangle$	$\lvert\Gamma_8-\frac{\hat{3}}{2}\rangle$ $\lvert U'\nu\rangle$

For integer values of j, they are six:

$$|\hat{0}\pm\rangle = \{|jm\pm\rangle = \frac{1}{\sqrt{2(1+\delta_{m,0})}}(|jm\rangle \pm(-1)^{j-m}|j-m\rangle),$$
$$m\geqslant 0, \quad m\equiv 0(4)\}, \tag{2}$$

$$|\hat{2}\pm\rangle = \{|jm\pm\rangle = \frac{1}{\sqrt{2}}(|jm\rangle \pm(-1)^{j-m}|j-m\rangle),$$
$$m>0, \quad m\equiv 2\ (4)\},$$
$$|\pm\hat{1}\rangle = \{|jm\rangle, \quad m\equiv\pm 1\ (4)\}$$

(the $\hat{}$ stands for "mod 4").

For half-integer values of j, they are four:

$$|\pm\hat{\tfrac{3}{2}}\rangle = \{|jm\rangle, \quad m\equiv\pm\tfrac{3}{2}(4)\},$$
$$|\pm\hat{\tfrac{1}{2}}\rangle = \{|jm\rangle, \quad m\equiv\pm\tfrac{1}{2}(4)\}. \tag{3}$$

Let us note that most workers in the field, following Bethe, did introduce such subspaces, but without the good sign $(-1)^{j-m}$, the result being a useless distinction between even j and odd j cases.

For integer values of j, there are five irreducible representations of O: Γ_1, Γ_2, Γ_3, Γ_4, Γ_5 in Bethe notation. The behavior of these representations under a rotation of $\pi/2$ around $0z$ and of π around $0y$ allows us to determine the components of every representation of O in the subspaces defined above relatively to D_4 (see Table I). Γ_1 and Γ_2 are one-dimensional, and Γ_3 is two-dimensional. Γ_1, Γ_2, and Γ_3 are invariant under a rotation of π around $0y$. Γ_4 and Γ_5 are three-dimensional and transform under the rotation $R(0,\pi,0)$ of π around $0y$ according to

$$R(0,\pi,0)|\Gamma_{4\text{ or }5}\hat{m}\rangle = (-1)^{\hat{m}+1}|\Gamma_{4\text{ or }5}-\hat{m}\rangle. \tag{4}$$

Under the rotation R_y of $\pi/2$ around $0y$, Γ_1 transforms like an angular momentum $j=0$, Γ_4 like $j=1$, and the laws of transformation of Γ_2, Γ_3, Γ_5 are deduced from the reduced matrix elements for $j=2$ and $j=3$:

$$|\Gamma_1\ \hat{0}\rangle \to |\Gamma_1\ \hat{0}\rangle,$$
$$|\Gamma_2\ \hat{2}\rangle \to -|\Gamma_2\ \hat{2}\rangle,$$

$$\begin{pmatrix}|\Gamma_3\ \hat{0}\rangle\\ |\Gamma_3\ \hat{2}\rangle\end{pmatrix}\to\begin{pmatrix}-\frac{1}{2} & \frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}|\Gamma_3\ \hat{0}\rangle\\ |\Gamma_3\ \hat{2}\rangle'\end{pmatrix}$$

$$\begin{pmatrix}|\Gamma_4 & \hat{1}\rangle\\ |\Gamma_4 & \hat{0}\rangle\\ |\Gamma_4 & -\hat{1}\rangle\end{pmatrix}\to\begin{pmatrix}\frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2}\\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}}\\ \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}|\Gamma_4 & \hat{1}\rangle\\ |\Gamma_4 & \hat{0}\rangle,\\ |\Gamma_4 & -\hat{1}\rangle\end{pmatrix}$$

$$\begin{pmatrix}|\Gamma_5 & \hat{1}\rangle\\ |\Gamma_5 & \hat{2}\rangle\\ |\Gamma_5 & -\hat{1}\rangle\end{pmatrix}\begin{pmatrix}-\frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{2}\\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}}\\ -\frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{2}\end{pmatrix}\begin{pmatrix}|\Gamma_5 & \hat{1}\rangle\\ |\Gamma_5 & \hat{2}\rangle.\\ |\Gamma_5 & -\hat{1}\rangle\end{pmatrix} \tag{5}$$

For half-integer values of j, there are three additional irreducible representations: Γ_6 and Γ_7 of dimension two and Γ_8 of dimension four. Under the rotation of π around $0y$ they transform according to

$$R(0,\pi,0)|\Gamma_6\hat{m}\rangle = (-1)^{1/2-m}|\Gamma_6 - \hat{m}\rangle,$$
$$R(0,\pi,0)|\Gamma_7\hat{m}\rangle = (-1)^{3/2-m}|\Gamma_7 - \hat{m}\rangle, \tag{6}$$
$$R(0,\pi,0)|\Gamma_8\hat{m}\rangle = (-1)^{3/2-m}|\Gamma_8 - \hat{m}\rangle.$$

Therefore, only components with $\hat{m}\geqslant 0$ need to be determined. The action of the rotation of $\pi/2$ around $0z$ leads to Table II.

Under the rotation R_y of $\pi/2$ around $0y$, Γ_6 transforms like $j=\frac{1}{2}$, Γ_8 like $j=\frac{3}{2}$, and the law of transformation of Γ_7 may be determined from the reduced matrix elements of $j=\frac{5}{2}$ tabulated, for instance, in Ref. 52:

$$\begin{pmatrix} |\Gamma_6 \quad \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_6 \quad -\hat{\tfrac{1}{2}}\rangle \end{pmatrix} \to \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} |\Gamma_6 \quad \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_6 \quad -\hat{\tfrac{1}{2}}\rangle \end{pmatrix},$$

$$\begin{pmatrix} |\Gamma_7 \quad \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_7 \quad -\hat{\tfrac{3}{2}}\rangle \end{pmatrix} \to \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} |\Gamma_7 \quad \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_7 \quad -\hat{\tfrac{3}{2}}\rangle \end{pmatrix}, \tag{7}$$

$$\begin{pmatrix} |\Gamma_8 \quad \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_8 \quad \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8 \quad -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8 \quad -\hat{\tfrac{3}{2}}\rangle \end{pmatrix} \to \begin{pmatrix} \frac{1}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ \frac{\sqrt{3}}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} \\ \frac{\sqrt{3}}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} & -\frac{\sqrt{3}}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{\sqrt{3}}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \end{pmatrix} \begin{pmatrix} |\Gamma_8 \quad \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_8 \quad \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8 \quad -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8 \quad -\hat{\tfrac{3}{2}}\rangle \end{pmatrix}.$$

These laws of transformation will be the basic material for determining projectors.

II. PROJECTION OF A VECTOR $|jm\rangle$ ON AN IRREDUCIBLE REPRESENTATION OF THE DOUBLE CUBIC GROUP O^+ FOR HALF-INTEGER VALUES OF j

An arbitrary vector $|jm\rangle$ is the sum of two unnormalized vectors belonging to two of the three spaces $\Gamma_6, \Gamma_7, \Gamma_8$ (see Table II); for instance, if $m \equiv \frac{1}{2}$ (4):

$$|jm\rangle = |\Gamma_6 \hat{\tfrac{1}{2}}\rangle_m + |\Gamma_8 \hat{\tfrac{1}{2}}\rangle_m, \tag{8}$$

where subscript m reminds us that the vector is unnormalized and depends on m.

Our purpose is to determine not only the two terms of the above sum, but also the other components of the involved representations. Due to relations (6), we will give the formulas only for $m > 0$.

The action of R_y upon $|jm\rangle$ yields

$$\begin{aligned} m \equiv \tfrac{1}{2}\ (4): \quad R_y|jm\rangle &= \sum_{m'} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'\rangle \\ &= \frac{1}{\sqrt{2}}|\Gamma_6 \hat{\tfrac{1}{2}}\rangle_m - \frac{1}{\sqrt{2}}|\Gamma_6 - \hat{\tfrac{1}{2}}\rangle_m \\ &\quad + \frac{\sqrt{3}}{2\sqrt{2}}|\Gamma_8 \hat{\tfrac{3}{2}}\rangle_m - \frac{1}{2\sqrt{2}}|\Gamma_8 \hat{\tfrac{1}{2}}\rangle_m - \frac{1}{2\sqrt{2}}|\Gamma_8 - \hat{\tfrac{1}{2}}\rangle_m + \frac{\sqrt{3}}{2\sqrt{2}}|\Gamma_8 - \hat{\tfrac{3}{2}}\rangle_m. \end{aligned} \tag{9}$$

Identifying these two expressions leads to

$$m \equiv \tfrac{1}{2}\ (4): \quad |\Gamma_8 \hat{\tfrac{3}{2}}\rangle_m = \frac{2\sqrt{2}}{\sqrt{3}} \sum_{m' \equiv 3/2\ (4)} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'\rangle, \tag{10}$$

which is one of the desired components.

In order to isolate the component $|\Gamma_6 \hat{\tfrac{1}{2}}\rangle_m$, we perform the combination of (8) and (9) which eliminates $|\Gamma_8 \hat{\tfrac{1}{2}}\rangle_m$ and gives a coefficient of $|\Gamma_6 \hat{\tfrac{1}{2}}\rangle_m$ equal to 1:

$$m \equiv \frac{1}{2}\ (4): |\Gamma_6 \hat{\tfrac{1}{2}}\rangle_m = \frac{1}{3} \sum_{m' \equiv 1/2\ (4)} \left(\delta_{m,m'} + 2\sqrt{2} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'\rangle. \tag{11}$$

Similarly,

$$m \equiv \frac{1}{2}\ (4): \quad |\Gamma_8 \hat{\tfrac{1}{2}}\rangle_m = \frac{2}{3} \sum_{m' \equiv 1/2\ (4)} \left(\delta_{m,m'} - \sqrt{2} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'\rangle. \tag{12}$$

In the same manner, if m is congruent to $\frac{3}{2}$ mod 4, then

$$m\equiv\frac{3}{2}\ (4):\quad |\,jm\rangle = |\Gamma_{7\frac{3}{2}}\rangle_m + |\Gamma_{8\frac{3}{2}}\rangle_m, \tag{13}$$

and the action of R_y upon $|\,jm\rangle$ leads to

$$\begin{aligned} m\equiv\frac{3}{2}\ (4):\quad R_y|\,jm\rangle &= \sum_{m'} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|\,jm'\rangle = -\frac{1}{\sqrt{2}}|\Gamma_{7\frac{3}{2}}\rangle_m + \frac{1}{\sqrt{2}}|\Gamma_{7\,-\frac{3}{2}}\rangle_m \\ &\quad + \frac{1}{2\sqrt{2}}|\Gamma_{8\frac{3}{2}}\rangle_m - \frac{\sqrt{3}}{2\sqrt{2}}|\Gamma_{8\frac{1}{2}}\rangle_m + \frac{\sqrt{3}}{2\sqrt{2}}|\Gamma_{8\,-\frac{1}{2}}\rangle_m - \frac{1}{2\sqrt{2}}|\Gamma_{8\,-\frac{3}{2}}\rangle_m \end{aligned} \tag{14}$$

from which we deduce

$$m\equiv\frac{3}{2}\ (4):\quad |\Gamma_{8\frac{1}{2}}\rangle_m = \frac{-2\sqrt{2}}{\sqrt{3}} \sum_{m'\equiv 1/2(4)} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|\,jm'\rangle. \tag{15}$$

By linear combinations of (13) and (14) we obtain

$$m\equiv\frac{3}{2}\ (4):\quad |\Gamma_{7\frac{3}{2}}\rangle_m = \frac{1}{3}\sum_{m'\equiv 3/2\ (4)}\left(\delta_{m'm} - 2\sqrt{2}d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|\,jm'\rangle \tag{16}$$

and

$$m\equiv\frac{3}{2}\ (4):\quad |\Gamma_{8\frac{3}{2}}\rangle_m = \frac{2}{3}\sum_{m'\equiv 3/2\ (4)}\left(\delta_{m'm} + \sqrt{2}d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|\,jm'\rangle. \tag{17}$$

In conclusion, if we perform these operations on all vectors of subspaces $|\frac{1}{2}\rangle$ and $|\frac{3}{2}\rangle$, the projectors on Γ_6, Γ_7, and Γ_8 are, respectively,

$$P_6 = \begin{pmatrix} 0 & 0 \\ 0 & (1+2\sqrt{2}R_y)/3 \end{pmatrix} \tag{18a}$$

$$P_7 = \begin{pmatrix} (1-2\sqrt{2}R_y)/3 & 0 \\ 0 & 0 \end{pmatrix}, \tag{18b}$$

$$P_8 = \begin{pmatrix} (2+2\sqrt{2}R_y)/3 & -(2\sqrt{2}/\sqrt{3})R_y \\ (2\sqrt{2}/\sqrt{3})R_y & (2-2\sqrt{2}R_y)/3 \end{pmatrix}, \tag{18c}$$

where the first row or column corresponds to space $|\frac{3}{2}\rangle$ and the second one to $|\frac{1}{2}\rangle$. The three projectors are symmetrical (see Sec. IV), including P_8 because the opposite signs appearing in front of $(2\sqrt{2}/\sqrt{3})R_y$ cancel the opposite signs of the two restrictions of R_y.

III. PROJECTION OF A VECTOR $|\,jm\rangle$ ON AN IRREDUCIBLE REPRESENTATION OF THE CUBIC GROUP O FOR INTEGER VALUES OF j

Examining Table I shows that it is advisable, for even values of m, to use the symmetrical and antisymmetrical combinations (2) of vectors $|\,jm\rangle$ which span the irreducible representations of the point group D_4. Moreover, Γ_1, Γ_2, and Γ_3 generate the subspaces $|\hat{0}+\rangle$ and $|\hat{2}+\rangle$ while Γ_4 and Γ_5 generate the four others; this will make us split the study in two parts.

A. Projection of Γ_1,Γ_2,Γ_3

sv1,42.5p

Considering an arbitrary vector $|\,jm+\rangle$ of $|\hat{0}+\rangle$ with $m\geqslant 0$, we write it as the sum

$$|\,jm+\rangle = |\Gamma_1\hat{0}\rangle_m + |\Gamma_3\hat{0}\rangle_m. \tag{19}$$

The action of R_y gives

$$\begin{aligned} m\equiv 0(4),\ m\geqslant 0:\quad R_y|\,jm+\rangle &= \frac{1}{\sqrt{2(1+\delta_{m0})}}\sum_{m'}\left(d^{(j)}_{m'm}\left(\frac{\pi}{2}\right) + (-1)^{j-m}d^{(j)}_{m',-m}\left(\frac{\pi}{2}\right)\right)|\,jm'\rangle \\ &= \sum_{m'\geqslant 0}\frac{2}{\sqrt{(1+\delta_{m,0})(1+\delta_{m',0})}}\,d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|\,jm'+\rangle = |\Gamma_1\hat{0}\rangle_m - \frac{1}{2}|\Gamma_3\hat{0}\rangle_m + \frac{\sqrt{3}}{2}|\Gamma_3\hat{2}\rangle_m, \end{aligned}$$

where we have used the symmetry relations of $d^{(j)}_{mm'}(\pi/2)$ to derive the second line from the first one (see the Appendix).

Therefore

$$m\equiv 0(4),\ m\geqslant 0:\quad |\Gamma_3\hat{2}\rangle_m = \frac{4}{\sqrt{3}\sqrt{1+\delta_{m,0}}}\sum_{\substack{m'\equiv 2\,(4)\\ m'>0}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'+\rangle. \tag{21}$$

By linear combinations of (19) and (20) we get

$$m\equiv 0(4),\ m\geqslant 0:\quad |\Gamma_1\hat{0}\rangle_m = \frac{1}{3}\sum_{\substack{m'\equiv 0\,(4)\\ m'>0}}\left(\delta_{m'm} + \frac{4}{\sqrt{(1+\delta_{m,0})(1+\delta_{m',0})}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'+\rangle, \tag{22}$$

$$m\equiv 0(4),\ m\geqslant 0:\quad |\Gamma_3\hat{0}\rangle_m = \frac{1}{3}\sum_{\substack{m'\equiv 0\,(4)\\ m'>0}}\left(2\delta_{m'm} - \frac{4}{\sqrt{(1+\delta_{m,0})(1+\delta_{m',0})}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'+\rangle. \tag{23}$$

In order to obtain $|\Gamma_2\hat{2}\rangle$, we consider a vector of $|\hat{2}+\rangle$

$$m\equiv 2(4),\ m>0:\quad |jm+\rangle = |\Gamma_2\hat{2}\rangle_m + |\Gamma_3\hat{2}\rangle_m \tag{24}$$

and the same type of computation leads to

$$m\equiv 2\,(4),\ m\geqslant 0:\quad |\Gamma_3\hat{0}\rangle_m = \frac{4}{\sqrt{3}}\sum_{\substack{m'\equiv 0\,(4)\\ m'>0}} \frac{1}{\sqrt{1+\delta_{m',0}}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'+\rangle, \tag{25}$$

$$m\equiv 2\,(4),\ m>0:\quad |\Gamma_2\hat{2}\rangle_m = \frac{1}{3}\sum_{\substack{m'\equiv 2\,(4)\\ m'>0}}\left(\delta_{m'm} - 4d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'+\rangle, \tag{26}$$

$$m\equiv 2\,(4),\ m>0:\quad |\Gamma_3\hat{2}\rangle_m = \frac{1}{3}\sum_{\substack{m'\equiv 2\,(4)\\ m'>0}}\left(2\delta_{m'm} + 4d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'+\rangle. \tag{27}$$

As for half-integer values of j, if we perform these computations for every value of m, we can obtain the projector on every subspace. Using the formula

$$\langle jm'+|R_y|jm+\rangle = \frac{2}{\sqrt{(1+\delta_{m,0})(1+\delta_{m',0})}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right). \tag{28}$$

We get

$$P_1 = \begin{pmatrix} \frac{1}{3}(1+2R_y) & 0 \\ 0 & 0 \end{pmatrix}, \tag{29a}$$

$$P_2 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{3}(1-2R_y) \end{pmatrix}, \tag{29b}$$

$$P_3 = \begin{pmatrix} \frac{2}{3}(1-R_y) & (2/\sqrt{3})R_y \\ (2/\sqrt{3})R_y & \frac{2}{3}(1+R_y) \end{pmatrix}, \tag{29c}$$

where the first row or column corresponds to space $|\hat{0}+\rangle$ and the second one to $|\hat{2}+\rangle$.

We could have derived these results in another way. Γ_1 and Γ_2 are invariant under the rotation R_3 of $2\pi/3$ around the ternary axis (1,1,1) with Euler angles $(\pi/2, \pi/2, 0)$. Therefore, the projector $(1+R_3+R_3^2)/3$ is equal to P_1+P_2; the restriction of R_3 to $|\hat{0}+\rangle$ is equal to R_y while its restriction to $|\hat{2}+\rangle$ is equal to $-R_y$.

B. Projection of Γ_4 and Γ_5

Since the components $|\Gamma_4\hat{0}\rangle$ and $|\Gamma_5\hat{2}\rangle$ are the only ones to generate subspaces $|\hat{0}-\rangle$ and $|\hat{2}-\rangle$, respectively, a simple convention is to choose the basis vectors (2) as the representations

$$|\Gamma_4\hat{0}\rangle_m = |jm-\rangle, \quad |\Gamma_5\hat{2}\rangle_m = |jm-\rangle. \tag{30}$$

We still have to determine $|\Gamma_4\hat{1}\rangle$ and $|\Gamma_5\hat{1}\rangle$. The action of R_y upon $|\Gamma_4\hat{0}\rangle_m$ gives

$$m\equiv 0\,(4),\ m\geqslant 0:\quad R_y|\Gamma_4\hat{0}\rangle_m = \frac{\sqrt{2}}{\sqrt{1+\delta_{m,0}}}\sum_{m' \text{ odd}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'\rangle = \frac{1}{\sqrt{2}}|\Gamma_4\hat{1}\rangle_m - \frac{1}{\sqrt{2}}|\Gamma_4-\hat{1}\rangle_m. \tag{31}$$

Therefore

$$m\equiv 0\,(4),\quad m\geqslant 0:\quad |\Gamma_4\hat{1}\rangle_m = \frac{2}{\sqrt{1+\delta_{m,0}}}\sum_{m'\equiv 1\,(4)} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'\rangle. \tag{32}$$

Similarly,

$$m\equiv 2\,(4),\quad m>0:\quad |\Gamma_5\hat{1}\rangle_m = 2\sum_{m'\equiv 1\,(4)} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'\rangle. \tag{33}$$

Although every component of every representation is at this point determined, it may be of interest to project a vector of $|\hat{1}\rangle$ on Γ_4 and Γ_5 separately:

$$m\equiv 1\,(4):\quad |jm\rangle = |\Gamma_4\hat{1}\rangle_m + |\Gamma_5\hat{1}\rangle_m, \tag{34}$$

$$m\equiv 1\ (4):\quad R_y|jm\rangle = \tfrac{1}{2}|\Gamma_4\hat{1}\rangle_m - (1/\sqrt{2})|\Gamma_4\hat{0}\rangle_m + \tfrac{1}{2}|\Gamma_4-\hat{1}\rangle_m - \tfrac{1}{2}|\Gamma_5\hat{1}\rangle_m - (1/\sqrt{2})|\Gamma_5\hat{0}\rangle_m - \tfrac{1}{2}|\Gamma_5-\hat{1}\rangle_m. \tag{35}$$

Therefore, another choice may be

$$m\equiv 1\ (4):\quad |\Gamma_4\hat{0}\rangle_m = -2\sum_{\substack{m'\equiv 0\,(4)\\ m'>0}} \frac{1}{\sqrt{1+\delta_{m'0}}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'-\rangle,$$

$$|\Gamma_5\hat{2}\rangle_m = -2\sum_{\substack{m'\equiv 2\,(4)\\ m'>0}} d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)|jm'-\rangle,$$

$$|\Gamma_4\hat{1}\rangle_m = \frac{1}{2}\sum_{\substack{m'\equiv 1\,(4)\\ m'>0}}\left(\delta_{m'm} + 2d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'\rangle, \tag{36}$$

$$|\Gamma_5\hat{1}\rangle_m = \frac{1}{2}\sum_{\substack{m'\equiv 1\,(4)\\ m'>0}}\left(\delta_{m'm} - 2d^{(j)}_{m'm}\left(\frac{\pi}{2}\right)\right)|jm'\rangle,$$

but in this case, components $|\Gamma_4\hat{0}\rangle$ and $|\Gamma_5\hat{2}\rangle$ mix all vectors, to the contrary of formula (30). To compute a matrix element, it is easier to use formula (30). For example, the proportionality coefficient of the fictitious spin[4,21] $l=1$ can be obtained from

$$\alpha = \langle\Gamma_4\hat{1}|\frac{1}{\sqrt{2}}J_+|\Gamma_4\hat{0}\rangle = \frac{1}{1+\delta_{m,0}}\left(\sqrt{j(j+1)-m(m+1)}\,d^{(j)}_{m+1,m}\left(\frac{\pi}{2}\right) - \sqrt{j(j+1)-m(m-1)}\,d^{(j)}_{m-1,m}\left(\frac{\pi}{2}\right)\right)$$

(α is by definition equal to 1 for $j=1$).

For any representation other than Γ_4 or Γ_5, for which a fictitious spin can be introduced, i.e., Γ_6, Γ_7, Γ_8, each coefficient α is a sum with many more terms.

IV. ORTHOGONALIZATION AND NORMALIZATION

The projection method yields directly the norm of representations. Indeed, for instance, if m is congruent to $\frac{1}{2}$ mod 4,

$$m\equiv\tfrac{1}{2}\,(4):\quad |jm\rangle = |\Gamma_{6\frac{1}{2}}\rangle_m + |\Gamma_{8\frac{1}{2}}\rangle_m \tag{37}$$

$$P_6|jm\rangle = |\Gamma_{6\frac{1}{2}}\rangle_m; \tag{38}$$

therefore,

$$\langle jm|P_6|jm\rangle = {}_m\langle\Gamma_{6\frac{1}{2}}|\Gamma_{6\frac{1}{2}}\rangle_m \tag{39}$$

and $|\Gamma_{6\frac{1}{2}}\hat{1}\rangle_m$ is normalized simply by dividing it by the square root of the diagonal element of the projector. If this element is zero, the representation does not exist, at least for the projected $|jm\rangle$; as the diagonal element of a projector, it can never be negative. A similar computation gives the norm of $|\Gamma_{8\frac{1}{2}}\hat{1}\rangle_m$ and the same value must also be used for the associated $|\Gamma_{8\frac{1}{2}}\hat{2}\rangle_m$ component as given by Eq. (10).

In the formulas giving the projectors, P_1, P_2, P_6, and P_7 are true projectors in the sense that $P^2=P$ while, for P_3 and P_8, $P^2=2P$ because two components are involved. However, the restrictions of P_3 to $|\hat{0}+\rangle$ or $|\hat{2}+\rangle$ only and of P_8 to $|\hat{\frac{1}{2}}\rangle$ or $|\hat{\frac{3}{2}}\rangle$ only are true projectors.

When the trace of a true projector is equal to 1, after the above described operation of normalization, all the rows of the projector are equal to within an undetermined sign.

When the trace of a true projector is greater than 1, the number of representations is equal to the trace; every row, after having been divided by its diagonal element, is a normalized representation, but these rows are neither identical nor orthogonal. However, the overlap of $|\Gamma_{6\frac{1}{2}}\rangle_m$ and $|\Gamma_{6\frac{1}{2}}\rangle_{m'}$ before normalization, is given by

$${}_m\langle\Gamma_{6\,\frac{1}{2}}\hat{|}\Gamma_{6\,\frac{1}{2}}\hat{\rangle}_{m'} = \langle jm|P_6|jm'\rangle. \tag{40}$$

Instead of projecting the vectors of basis $|jm_i\rangle$ in which the projector is expressed [throughout this section, $|jm_i\rangle$ will mean an arbitrary vector defined by (2) and (3)], we project a new sequence of vectors $|jm_i\rangle$ defined as follows:

$$|jm_i) = |jm_i\rangle + \sum_{l<i}\alpha_{il}|jm_l\rangle, \tag{41}$$

where the α_{il}'s are chosen such that the projected vector $P|jm_i)$ has no components on $|jm_k\rangle$ for k inferior to i:

$$\forall i,\ \forall k\quad k<i \Rightarrow \langle jm_k|P|jm_i) = 0. \tag{42}$$

As long as the elimination of only $|jm_l\rangle$ is involved, the coefficients α_{il} are easy to obtain, and the resulting new ma-

trix elements are

$$\langle jm_k|P|jm_i\rangle = \langle jm_k|P|jm_i\rangle - \frac{\langle jm_l|P|jm_i\rangle\langle jm_k|P|jm_l\rangle}{\langle jm_l|P|jm_l\rangle}. \tag{43}$$

This is identical to a triangulation method performed successively on every line. Due to relation (40) this method is equivalent to the Schmidt orthogonalization method.

The original elements $\langle jm_k|P|jm_i\rangle$ contain the square root of a quotient of product of factorials (see the Appendix). We easily see that exactly the same square root is found in $\langle jm_k|P|jm_i\rangle$, but a large number may have appeared in denominator. As for diagonal elements, the original ones are rational and the new ones also.

All this process can be iterated, and it will stop after r times, r being the trace of the projector: the $d-r$ remaining rows are then all zeros.

If the considered representation has more than one component (e.g., Γ_3,Γ_8), the present elimination process must be performed on all components simultaneously.

At the end of the iterations, the rectangular matrix of the projector has only r nonzero lines. The element at intersection of the pth line and the qth column is 0 for $q<p$; otherwise, it is equal to the quotient of the minor of order p made with the first p elements of columns 1 to $p-1$ and column q by the product of diagonal minors of order 1 to $p-1$. Of course, practically, it is better to apply repeatedly formula (43).

A consequence of the triangulation method is

$$\langle jm_i|P|jm_i\rangle = \langle jm_i|P|jm_i\rangle,$$

expression equal to the square of the norm.

To illustrate the method, let us take the example of $j=\frac{11}{2}$. After reordering base vectors so as to group vectors of subspaces (3), the restriction of R_y to subspaces $|+\hat{\tfrac{3}{2}}\rangle$ and $|+\hat{\tfrac{1}{2}}\rangle$ is

$$R_y = \begin{array}{c} \begin{matrix} |\frac{11}{2}\rangle & |\frac{3}{2}\rangle & |-\frac{5}{2}\rangle & |\frac{9}{2}\rangle & |\frac{1}{2}\rangle & |-\frac{7}{2}\rangle \end{matrix} \\ \begin{pmatrix} \frac{1}{32\sqrt{2}} & \frac{\sqrt{11\times15}}{32} & \frac{\sqrt{15\times11}}{32\sqrt{2}} & \frac{11}{32\sqrt{2}} & \frac{3\sqrt{15}}{32} & -\frac{5\sqrt{15}}{32\sqrt{2}} \\ \frac{\sqrt{11\times15}}{32} & -\frac{3}{16\sqrt{2}} & -\frac{11}{32} & \frac{\sqrt{21\times11}}{32} & \frac{\sqrt{35}}{16\sqrt{2}} & \frac{\sqrt{35}}{32} \\ \frac{\sqrt{11\times15}}{32\sqrt{2}} & -\frac{11}{32} & \frac{5}{32\sqrt{2}} & \frac{55}{32\sqrt{2}} & -\frac{\sqrt{3}}{32} & \frac{7\sqrt{3}}{32\sqrt{2}} \\ \frac{11}{32\sqrt{2}} & \frac{\sqrt{21\times11}}{32} & \frac{\sqrt{55}}{32\sqrt{2}} & -\frac{9}{32\sqrt{2}} & -\frac{\sqrt{21}}{32} & \frac{7\sqrt{5}}{32\sqrt{2}} \\ \frac{3\sqrt{15}}{32} & \frac{\sqrt{35}}{16\sqrt{2}} & -\frac{\sqrt{3}}{32} & -\frac{\sqrt{21}}{32} & -\frac{5}{16\sqrt{2}} & -\frac{\sqrt{105}}{32} \\ -\frac{5\sqrt{15}}{32\sqrt{2}} & \frac{\sqrt{35}}{32} & \frac{7\sqrt{3}}{32\sqrt{2}} & \frac{7\sqrt{5}}{32\sqrt{2}} & -\frac{\sqrt{105}}{32} & \frac{19}{32\sqrt{2}} \end{pmatrix} \end{array} \begin{matrix} \\ |\frac{11}{2}\rangle \\ |\frac{3}{2}\rangle \\ |-\frac{5}{2}\rangle \\ |\frac{9}{2}\rangle \\ |\frac{1}{2}\rangle \\ |-\frac{7}{2}\rangle \end{matrix}.$$

Projectors P_6 on $|\Gamma_6\hat{\tfrac{1}{2}}\rangle$ and P_7 on $|\Gamma_7\hat{\tfrac{3}{2}}\rangle$ are given by formulas (18)

$$P_6 = \begin{pmatrix} \frac{7}{48} & -\frac{\sqrt{7}}{8\sqrt{6}} & \frac{7\sqrt{5}}{48} \\ -\frac{\sqrt{7}}{8\sqrt{6}} & \frac{1}{8} & -\frac{\sqrt{35}}{8\sqrt{6}} \\ \frac{7\sqrt{5}}{48} & -\frac{\sqrt{35}}{8\sqrt{6}} & \frac{35}{48} \end{pmatrix}, \quad P_7 = \begin{pmatrix} \frac{5}{16} & -\frac{\sqrt{55}}{8\sqrt{6}} & -\frac{\sqrt{55}}{16\sqrt{3}} \\ -\frac{\sqrt{55}}{8\sqrt{6}} & \frac{11}{24} & \frac{11}{24\sqrt{2}} \\ -\frac{\sqrt{55}}{16\sqrt{3}} & \frac{11}{24\sqrt{2}} & \frac{11}{48} \end{pmatrix}.$$

For each of them, the trace is 1; therefore, the three rows are proportional and, after dividing each row by its diagonal element, the three rows are equal to within an undetermined sign:

$$P_6 \to \begin{pmatrix} \frac{\sqrt{7}}{4\sqrt{3}} & -\frac{1}{2\sqrt{2}} & \frac{\sqrt{35}}{4\sqrt{3}} \\ -\frac{\sqrt{7}}{4\sqrt{3}} & \frac{1}{2\sqrt{2}} & -\frac{\sqrt{35}}{4\sqrt{3}} \\ \frac{\sqrt{7}}{4\sqrt{3}} & -\frac{1}{2\sqrt{2}} & \frac{\sqrt{35}}{4\sqrt{3}} \end{pmatrix}, \quad P_7 \to \begin{pmatrix} \frac{\sqrt{5}}{4} & -\frac{\sqrt{11}}{2\sqrt{6}} & -\frac{\sqrt{11}}{4\sqrt{3}} \\ -\frac{\sqrt{5}}{4} & \frac{\sqrt{11}}{2\sqrt{6}} & \frac{\sqrt{11}}{4\sqrt{3}} \\ -\frac{\sqrt{5}}{4} & \frac{\sqrt{11}}{2\sqrt{6}} & \frac{\sqrt{11}}{4\sqrt{3}} \end{pmatrix}.$$

As for projector P_8 on $|\Gamma_8^{\frac{3}{2}}\rangle$ and $|\Gamma_8^{\frac{1}{2}}\rangle$, its matrix is, in the same basis,

$$P_8 = \begin{pmatrix} \frac{11}{16} & \frac{\sqrt{55}}{8\sqrt{6}} & \frac{\sqrt{55}}{16\sqrt{3}} & \frac{\sqrt{11}}{16\sqrt{3}} & \frac{\sqrt{77}}{8\sqrt{2}} & \frac{\sqrt{55}}{16\sqrt{3}} \\ \frac{\sqrt{55}}{8\sqrt{6}} & \frac{13}{24} & -\frac{11}{24\sqrt{2}} & \frac{3\sqrt{5}}{8\sqrt{2}} & \frac{\sqrt{35}}{8\sqrt{3}} & -\frac{1}{8\sqrt{2}} \\ \frac{\sqrt{55}}{16\sqrt{3}} & -\frac{11}{24\sqrt{2}} & \frac{37}{48} & -\frac{5\sqrt{5}}{16} & \frac{\sqrt{35}}{8\sqrt{6}} & \frac{7}{16} \\ \frac{\sqrt{11}}{16\sqrt{3}} & \frac{3\sqrt{5}}{8\sqrt{2}} & -\frac{5\sqrt{5}}{16} & \frac{41}{48} & \frac{7}{8\sqrt{6}} & -\frac{7\sqrt{5}}{48} \\ \frac{\sqrt{77}}{8\sqrt{2}} & \frac{\sqrt{35}}{8\sqrt{3}} & \frac{\sqrt{35}}{8\sqrt{6}} & \frac{\sqrt{7}}{8\sqrt{6}} & \frac{7}{8} & \frac{\sqrt{35}}{8\sqrt{6}} \\ \frac{\sqrt{55}}{16\sqrt{3}} & -\frac{1}{8\sqrt{2}} & \frac{7}{16} & -\frac{7\sqrt{5}}{48} & \frac{\sqrt{35}}{8\sqrt{6}} & \frac{13}{48} \end{pmatrix}.$$

The trace of each of the two 3×3 diagonal blocks is 2; therefore, the rank of P_8 is 2. The elimination procedure (43), which is just to be performed on the first three rows or on the last three rows, can be extended for clarity to the whole matrix which becomes, after the first step,

$$P_8 \rightarrow \begin{pmatrix} \frac{11}{16} & \frac{\sqrt{55}}{8\sqrt{6}} & \frac{\sqrt{55}}{16\sqrt{3}} & \frac{11}{16\sqrt{3}} & \frac{\sqrt{77}}{8\sqrt{2}} & \frac{\sqrt{55}}{16\sqrt{3}} \\ 0 & \frac{1}{3} & \frac{\sqrt{2}}{3} & \frac{\sqrt{5}}{3\sqrt{2}} & 0 & -\frac{1}{3\sqrt{2}} \\ 0 & -\frac{\sqrt{2}}{3} & \frac{2}{3} & -\frac{\sqrt{5}}{3} & 0 & \frac{1}{3} \\ 0 & \frac{\sqrt{5}}{3\sqrt{2}} & -\frac{\sqrt{5}}{3} & \frac{5}{6} & 0 & -\frac{\sqrt{5}}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{-1}{3\sqrt{2}} & \frac{1}{3} & -\frac{\sqrt{5}}{6} & 0 & \frac{1}{6} \end{pmatrix}.$$

We notice that in this case the fifth row is accidentally zero; this means that the Γ_8 obtained by projection from $|\frac{11}{2}\rangle$ of the $|\hat{\frac{3}{2}}\rangle$ subspace (i.e., our choice) coincides with the Γ_8 obtained by projection from $|\frac{1}{2}\rangle$ of the $|\hat{\frac{1}{2}}\rangle$ subspace. Rows 2, 3, 4, and 6 are proportional and have no component on $|\frac{11}{2}\rangle$ and $|\frac{1}{2}\rangle$.

Therefore, only the first two rows of R_y have been necessary to determine Γ_8 and Γ_6. Three more elements of the fourth row lead to Γ_7. We could also notice that all these results could have been obtained with only the first and the fourth rows.

In order to obtain the Γ_8 representations for an arbitrary half-integer j, one should project a number of vectors of $|\hat{\frac{1}{2}}\rangle$ equal to the number of Γ_6, plus a number of vectors of $|\hat{\frac{3}{2}}\rangle$ equal to the number of Γ_7; if $j-\frac{3}{2}$ is a multiple of 3, one more vector need be projected. However, there can occur an "unhappy" choice such as the one of $|\frac{11}{2}\rangle$ and $|\frac{1}{2}\rangle$ for our example $j=\frac{11}{2}$.

CONCLUSION

The determination of cubic harmonics has been performed completely using only a small amount of group theory. Moreover, once the reduced rotation matrix elements $d^{(j)}_{mm'}(\pi/2)$ are known, the computation is so simple that it can be carried out without the help of a computer, even for relatively high values of j. Even if we do not write them down explicitly, we have obtained closed expressions for the components of normalized cubic harmonics; for instance, a normalized $|\Gamma_7^{\frac{3}{2}}\rangle$ component is simply given by

$$|\Gamma_7^{\frac{3}{2}}\rangle_m = \frac{1}{\sqrt{3}}\left(1-2\sqrt{2}d^{(j)}_{mm}\left(\frac{\pi}{2}\right)\right)^{-1/2}\sum_{m'\equiv 3/2\,(4)}\left(\delta_{m',m} - 2\sqrt{2}d^{(j)}_{m',m}\left(\frac{\pi}{2}\right)\right)|jm'\rangle.$$

In case there is no degeneracy, these expressions are independent of m, except for a sign; in case of degeneracy, the orthogonalization process described in Sec. IV may be applied.

For every representation, a convention must be chosen in case of degeneracy in order to perform the orthogonalization. For instance, if r is the number of Γ_i representations for a given j, a set of r orthogonal representations Γ_i is uniquely determined, up to a permutation, by the choice of successive values $m_1, m_2, \ldots, m_{r-1}$ in (43).

Let us mention a few possible conventions:

(a) The "low m" convention:
$m_k = 4(k-1) + m_1$, $k = 2,...,r-1$, where m_1 is chosen to be the lowest possible value of m in the associated D_4 subspace.

(b) The "high m" convention:
$m_k = -4(k-1) + m_1$, $k = 2,...,r-1$, with m_1 equal to the highest possible value.

Every author who tabulated orthonormal sets[2,6,8,12,16–18,28,29,31] chose either the low m or the high m convention. If one wants, as we do, to express the coefficients in terms of integers, we can devise a "smallest length" convention: As explained in the text, the new matrix elements may have in their denominator a large prime number coming from the numerator of diagonal element $\langle jm_i|P|jm_i\rangle$, and this large prime number will not cancel after normalization. Hence the convention: at each stage, choose for m_i the value which minimizes the rank of the highest prime number in the numerator of current diagonal elements.

These three conventions, like any use of algorithm (43), create at least $r(r-1)/2$ zeros among the set of coefficients. None of them can avoid the appearance of large prime numbers; the third one only delays it without completely avoiding it, particularly for Γ_8.

We have built simple orthonormal sets of representations for all values of j up to $\frac{41}{2}$ (this is the maximum value of the total angular momentum for fundamental electronic states of all known elements, given by $l(l+1) + \frac{1}{2}$ for $l = 4$, which corresponds to a half-filled g shell), using algorithm (43) followed by orthonormal combinations in order to suppress square roots of high prime numbers. Our plan was to publish these tables in the present article, but we just found quite simple relations between some representations, which makes useless the computation of certain kinds of representations.

If $\Gamma_1(l)$ is a representation for a given value of l, $\sqrt{(2s+1)(2l+1)}\{\Gamma_1(l)|s\hat{\sigma}\}_j/\sqrt{2j+1}$, where $\{\ \}_j$ means coupling by a Clebsch–Gordan coefficient l with s to j, is the $\hat{\sigma}$ component of a Γ_6 representation with $s = \frac{1}{2}$, a Γ_4 representation with $s = 1$ or a Γ_8 representation with $s = 3/2$; these representations are orthonormalized. As the coefficients of these three kinds of representations are the product of Γ_1 coefficients by simple Clebsch–Gordan coefficients, we consider tables for Γ_4, Γ_6, Γ_8, as unnecessary. Similarly, Γ_5 and Γ_7 can be expressed in terms of Γ_2 representations. Such a subject is quite different from the matter of the present article; proofs of these relations and some of their consequences will be the matter of a forthcoming article which will include tables of Γ_1, Γ_2, and Γ_3 only.

APPENDIX: ON ROTATION MATRIX ELEMENTS

The rotation matrix elements related to Euler angles (α, β, γ) are

$$\langle jm|R(\alpha,\beta,\gamma)|jm'\rangle = D^{(j)}_{m,m'}(\alpha,\beta,\gamma),$$

and they can be expressed in terms of the reduced matrix elements $d^{(j)}_{m,m'}(\beta)$ by

$$D^{(j)}_{m,m'}(\alpha,\beta\,\gamma) = e^{-im\alpha} d^{(j)}_{m,m'}(\beta) e^{-im'\gamma}$$

with

$$d^{(j)}_{m,m'}(\beta) = \sum_t (-1)^t \frac{\sqrt{(j+m)!(j-m)!(j+m')!(j-m')!}}{(j+m-t)!t!(j-m'-t)!(t-m+m')!} \times \left(\cos\frac{\beta}{2}\right)^{2j+m-m'-2t} \left(\sin\frac{\beta}{2}\right)^{2t-m+m'}.$$

These functions are related by many recurrence relations (see, for instance, Edmonds[51]). For example,

$$\sqrt{j(j+1)-m'(m'-1)}\, d^{(j)}_{m,m'-1}(\beta) + \sqrt{j(j+1)-m'(m'+1)}\, d^{(j)}_{m,m'+1}(\beta) = 2\frac{m'\cos\beta - m}{\sin\beta} d^{(j)}_{m,m'}(\beta).$$

Unfortunately, for large values of j, this formula is unstable and it is better to use

$$(j+1)[(j^2-m^2)(j^2-m'^2)]^{1/2} d^{(j-1)}_{m,m'}(\beta) + j[((j+1)^2-m^2)((j+1)^2-m'^2)]^{1/2} d^{(j+1)}_{m,m'}(\beta) = (2j+1)(j(j+1)\cos\beta - mm') d^{(j)}_{m,m'}(\beta).$$

The reduced matrix elements have the following symmetry relations:

$$d^{(j)}_{m,m'}(\beta) = d^{(j)}_{-m',-m}(\beta) = (-1)^{m-m'} d^{(j)}_{m',m}(\beta) = (-1)^{m-m'} d^{(j)}_{-m,-m'}(\beta),$$
$$d^{(j)}_{m,m'}(\beta) = (-1)^{j-m'} d^{(j)}_{-m,m'}(\pi-\beta) = (-1)^{j+m} d^{(j)}_{m,-m'}(\pi-\beta).$$

We are specially interested by the particular value $\beta = \pi/2$, in which case $d^{(j)}_{m,m'}(\pi/2)$ can be expressed in the form

$$d^{(j)}_{m,m'}\left(\frac{\pi}{2}\right) = \left(\frac{1}{\sqrt{2}}\right)^{2j} \left(\frac{(j+m')!(j-m')!}{(j+m)!(j-m)!}\right)^{1/2} \sum_t (-1)^t \times \frac{(j+m)!}{(j+m-t)!t!} \frac{(j-m)!}{(j-m'-t)!(t-m+m')!}.$$

In this expression we note the following.

The sum over t is a sum of products of binomial coefficients and the resulting integer can be very large.

The square root evaluates to 1 for diagonal matrix elements; if we take into account the $\sqrt{2}$ accompanying R_y in the formula giving the projectors for half-integer values of j, the power of $\sqrt{2}$ is even in all cases, and the final value of diagonal elements of projectors is an integer divided by a power of 2 and maybe by 3.

Under the square root there is no prime factor greater than $2j$.

[1] H. Bethe "Termaufspaltung in Kristallen," Ann. Phys. 3, 133–208 (1929), translated by A. P. Cracknell, *Applied Group Theory* (Pergamon, New York, 1968).

[2] K. Fox and B. J. Krohn, "Computation of cubic harmonics," J. Comput. Phys. 25, 386–408 (1977).

[3] K. C. Kim, W. B. Person, D. Seitz, and B. J. Krohn, "Analysis of the ν_4(615 cm^{-1}) region of the Fourier transform and diode laser spectra of

SF_6," J. Mol. Spectros. **76**, 322–40 (1979); C. D. Cantrell, A. A. Makarov, and W. H. Louisell, "Photoselective chemistry part I," Advan. Chem. Phys. **47** (1), 583–624 (1981).
[4]A. Abragam and M. H. L. Pryce, "Theory of the nuclear hyperfine structure of paramagnetic resonance spectra in crystals," Proc. Roy. Soc. A **205**, 135–153 (1951).
[5]E. Soulié and G. Goodman, "Niveaux d'énergie électronique et susceptibilité magnétique des ions de configuration f^2 en champ cristallin cubique," Theor. Chimi. Acta. **41**, 17 (1976) [erratum **51**, 259 (1979)].
[6]H. A. Jahn, "A new Coriolis perturbation in the methane spectrum. I. Vibrational-rotational Hamiltonian and wave functions," Proc. Roy. Soc. A **168**, 469–95 (1938).
[7]D. D. Betts, A. B. Bhatia, and M. Wyman "Houston's method and its application to the calculation of characteristic temperatures of cubic crystals," Phys. Rev. **104**, 37–42 (1956).
[8]K. T. Hecht, "The vibration-rotation energies of tetrahedral XY_4 molecules, part I: Theory of spherical top molecules," J. Mol. Spectros. **5**, 355–89 (1960).
[9]K. R. Lea, M. J. M. Leask, and W. P. Wolf, "The raising of angular momentum degeneracy of f-electron terms by cubic crystal fields," J. Phys. Chem. Solids **23**, 1381–1405 (1962).
[10]J. S. Griffith, *The Theory of Transition-Metal Ions* (Cambridge U. P., Cambridge, 1964), 2nd ed.
[11]J. Moret-Bailly, L. Gautier, and J. Montagutelli, "Clebsch–Gordan coefficients adapted to cubic symmetry," J. Mol. Spectros. **15**, 355–77 (1965).
[12]F. M. Mueller and M. G. Priestley, "Inversion of cubic de Haas-van Alphen data, with an application to palladium," Phys. Rev. **148**, 638–43 (1966).
[13]A. J. Hughes, "Wave function expansion in the Brillouin zone: silicon," Phys. Rev. **166**, 776–84 (1968).
[14]M. J. G. Lee and L. M. Falicov, "The de Haas-van Alphen effect and the Fermi surface of potassium," Proc. Roy. Soc. A **304**, 319–34 (1968).
[15]H. Puff, "Contribution to the theory of cubic harmonics," Phys. Status Solidi **41**, 11–22 (1970).
[16]J. Muggli, Cubic harmonics as linear combinations of spherical harmonics, Z. Angew. Math. Phys. **23**, 311–17 (1972).
[17]S. Asano and K. Yonei, "The actual forms of the cubic harmonic functions up to the 16th order. I. Quantization axis: (100) direction," Rep. Res. Lab. Surface Sci., Okayama Univ. Jpn. **4**, 55–76 (1974); S. Asano and Y. Nakao, "II. Quantization axis: (111) direction," *ibid.* **4**, 159–192 (1976).
[18]S. Asano, "Tables of the Clebsch–Gordan coefficients $\langle \Gamma_1 \gamma_1 \Gamma_2 \gamma_2 | \Gamma \gamma \rangle$ and the coefficients $\langle JM_J | \Gamma \gamma \rangle$ $(0 \leqslant J \leqslant 21/2)$ for the doubled-valued O_h group," Res. Notes Dept. Phys., Fac. Sci., Okayama Univ. **16**, 1–35 (1979).
[19]J. Raynal, "Determination of point group harmonics for arbitrary j by a projection method. II. Icosahedral group, quantization along an axis of order 5," J. Math. Phys. **25**, 1187 (1984).
[20]Y. Onodera and M. Okazaki, "Tables of basis functions for double point groups," J. Phys. Soc. Jpn. **21**, 2400–8 (1966).
[21]A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon, Oxford, 1970).
[22]G. J. Kynch, "Multiplet structure in a crystalline electric field of cubic symmetry," Trans. Faraday Soc. **33**, 1402–18 (1937).
[23]E. Teleman and A. Glodeanu, "Basis functions for the double groups O_h and D_{3h}," Rev. Roum. Phys. **12**, 551–60 (1967).
[24]A. P. Cracknell and S. J. Joshua, "Symmetry adapted functions for double point groups. II. Cubic point groups," Proc. Cambridge Philos. Soc. **67**, 647–56 (1970).
[25]F. C. von der Lage and H. A. Bethe, "A method for obtaining electronic eigenfunctions and eigenvalues in solids with an application to sodium," Phys. Rev. **71**, 612–22 (1947).
[26]D. G. Bell, "Group theory and crystal lattices," Rev. Mod. Phys. **26**, 311–20 (1954).
[27]J. Moret-Bailly, "Sur l'interprétation des spectres de vibration-rotation des molécules à symétrie tétraédrique ou octaédrique," Cah. Phys. **15**, 237–314 (1961).
[28]S. L. Altmann and C. J. Bradley, "On the symmetries of spherical harmonics," Phil. Trans. Roy. Soc. London A **255**, 199–215 (1963).
[29]S. L. Altmann and A. P. Cracknell, "Lattice harmonics I, Cubic groups," Rev. Mod. Phys. **37**, 19–32 (1965).
[30]J. Muggli, "Punktdefekte im kubischen elastischen Kontinuum," Phys. Konden. Mater. **12**, 237–61 (1971).
[31]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).
[32]W. R. Fehlner and S. H. Vosko, "A product representation for cubic harmonics and special directions for the determination of the Fermi surface and related properties," Can. J. Phys. **54**, 2159–69 (1976).
[33]T. Kwiatkowski, S. Olszewski, and A. Wierzbicki, "Cubic harmonics in Cartesian coordinates," Int. J. Quantum Chem. **11**, 21–47 (1977).
[34]P. P. Pavinskii and V. G. Kuznetsov, "Calculation of higher harmonics for Coulomb potential of a crystal. I. Spherical functions of the symmetry O_h," Vestn. Leningrad Univ. Fiz. Khim., 23–7 (1980).
[35]V. G. Kuznetsov and P. P. Pavinskii, "Calculation of higher harmonics for Coulomb potential of crystal. II. Expansion of the Ewald function for a fcc lattice in a series of cubic harmonics," Vestn. Leningrad Univ. Fiz. Khim., 11–7 (1980).
[36]F. S. Khan and P. B. Allen, "Orthonormal functions for Fermi surfaces. A test of convergence for Nb and Cu," Phys. Status Solidi B **100**, 227–34 (1980).
[37]J. P. Amoureux, J. L. Sauvajol, and M. Bée, "A symmetry-adapted function analysis of plastic crystals. Application to 1-cyanoadamantane at room temperature," Acta. Crystallogr. A **37**, 97–104 (1981).
[38]W. Elert, "Über das Schwingungs—und Rotationspektrum einer Molekel vom Typus CH_4," Z. Phys. **51**, 6–33 (1928).
[39]K. W. H. Stevens, "Matrix elements and operator equivalents connected with the magnetic properties of rare earth ions," Proc. Phys. Soc. **65**, 209–15 (1952).
[40]B. J. Krohn, "Diagonal $F^{(4)}$ and $F^{(6)}$ coefficients for spherical-top molecules in angular momentum states up to $J = 100$," Los Alamos Report LA-6554-MS, 1976.
[41]K. Fox, H. W. Galbraith, B. J. Krohn, and J. D. Louck, "Theory of level splitting: spectrum of the octahedrally invariant fourth-rank tensor operator," Phys. Rev. A **15**, 1363–81 (1977).
[42]W. G. Harter and C. W. Patterson, "Orbital level splitting in octahedral symmetry and SF_6 rotational spectra. I. Qualitative features of high J levels," J. Chem. Phys. **66**, 4872–85 (1977); C. W. Patterson and W. G. Harter, "II. Quantitative Treatment of High J Levels," J. Chem. Phys. **66**, 4886–92 (1977).
[43]E. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Vieweg, Braunschweig, 1931), translated by J. J. Griffin [*Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959)].
[44]M. A. Melvin, "Simplification in finding symmetry-adapted eigenfunctions," Rev. Mod. Phys. **28**, 18–44 (1956).
[45]H. V. McIntosh, "Towards a theory of the crystallographic point groups," J. Mol. Spectros. **5**, 269–83 (1960).
[46]S. L. Altmann, "On the symmetries of spherical harmonics," Proc. Cambridge Philos. Soc. (Math. Phys. Sci.) **53**, 343–67 (1957).
[47]K. Fox and I. Ozier, "Construction of tetrahedral harmonics," J. Chem. Phys. **52**, 5044–56 (1970).
[48]C. F. Dunkl, "Cube group invariant spherical harmonics and Krawtchouk polynomials," Mat. Z. **177** (4), 561–77 (1981).
[49]J. Raynal and R. Conte, "Determination of point group harmonics for arbitrary j by a projection method. III. Cubic group, quantization along a ternary axis," to be published.
[50]G. F. Koster, J. O. Dimmock, R. G. Wheeler, and H. Statz, *Properties of the Thirty-Two Point Groups* (MIT Press, Cambridge, MA, 1963).
[51]A. E. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, U.P., Princeton, NJ, 1957).
[52]D. A. Varshalovich, A. N. Moskaliev, and V. K. Khersonskii, *Quantum Theory of Angular Momenta* , in Russian (Nauka, Leningrad, 1975).
[53]G. Placzek, in *Handbuch der Radiologie*, edited by M. von Laue (Akademische-Verlagsges., Leipzig, 1925),Vol. 6, Part II, p. 205.
[54]R. S. Mulliken, "Electronic structures of polyatomic molecules and valence. IV. Electronic states, quantum theory of the double bond," Phys. Rev. **43**, 279–302 (1933).

Determination of point group harmonics for arbitrary j by a projection method. II. Icosahedral group, quantization along an axis of order 5

Jacques Raynal
Service de Physique Theorique, C. E. N. Saclay, F-91191 Gif-sur-Yvette, Cedex, France

(Received 10 May 1983; accepted for publication 2 September 1983)

The method described in the previous paper to obtain cubic harmonics is applied to the similar problem for the icosahedral group. The projection operators are expressed in terms of rotation matrix elements of $R(\pi,\varphi,0)$ and $R(0,\pi-\varphi,\pi)$ $\varphi = \arctan 2$, acting on subspaces invariant under D_5.

PACS numbers: 02.20. + b

The method used for cubic harmonics[1] can be applied to the icosahedral group. The icosahedral group is one of the five finite rotation groups and it is the only one for which the problem of point group harmonics is not trivial or is not related to the one of the cubic group. Icosahedral harmonics were studied by Mc Lellan[2] who tabulated them up to $j = 8$. They have some applications in molecular chemistry.[3]

The method leads to explicit formula (5) of Mc Lellan in terms of reduced rotation matrix elements for each representation of the icosahedral group. We discuss the structure of the coefficients but give no tables because we found simple relations between some representations. As these relations are very different from the ones of this work, their explanation is postponed to a further article in which tables will be given.

I. THE ICOSAHEDRAL GROUP AND ITS REPRESENTATIONS

The icosahedral group has six fivefold rotations around six axes with an angle $\varphi = \arctan 2$ between them. We choose one of them as z-axis and another one in the $\mathbf{x}\cdot\mathbf{z}$ plane, with the direction of the unitary vector $(\mathbf{z} - 2\mathbf{x})/\sqrt{5}$. The Euler angles (α, β, γ) describing a rotation of $2\pi/5$ around this second axis are, respectively, $(3\pi/5, \varphi, -2\pi/5)$. This rotation together with the rotation of $2\pi/5$ around z generate the 60 elements of this subgroup of O(3) and the 120 elements of the double icosahedral group as a subgroup of SU(2). The 60 sets of Euler angles are

$$\alpha = 0,\frac{2\pi}{5},\frac{4\pi}{5},\frac{6\pi}{5},\frac{8\pi}{5},\quad \beta = 0,\pi,\quad \gamma = 0, \tag{1a}$$

$$\alpha = \frac{\pi}{5},\frac{3\pi}{5},\pi,\frac{7\pi}{5},\frac{9\pi}{5},\quad \beta = \varphi,\quad \gamma = 0,\frac{2\pi}{5},\frac{4\pi}{5},\frac{6\pi}{5},\frac{8\pi}{5}, \tag{1b}$$

$$\alpha = 0,\frac{2\pi}{5},\frac{4\pi}{5},\frac{6\pi}{5},\frac{8\pi}{5},\quad \beta = \pi - \varphi, \tag{1c}$$

$$\gamma = \frac{\pi}{5},\frac{3\pi}{5},\pi,\frac{7\pi}{5},\frac{9\pi}{5},$$

and the 60 other ones of the double group are obtained by adding 2π to α.

The action of an arbitrary element on a base vector $|jm\rangle$ of an irreducible representation $\mathscr{D}_j$ of SU(2) is given by the rotation matrix having the corresponding Euler angles. The only nontrivial matrices we will have to deal with are $R(\pi,\varphi,0)$ and $R(0,\pi-\varphi,\pi)$: these matrices are real for integer values of j, purely imaginary for half-integer values of j and they are always symmetric. They are related by the rotation of π around $0y$ which is also an element of the group.

The 10 elements (1a) define what is called the D_5 group. It is suitable to introduce the subspaces of $\mathscr{D}_j$ which are invariant under D_5.

For integer values of j, there are 6:

$$\begin{aligned}
&|j\,\hat{0}\pm\rangle = \{|jm\pm\rangle = (1/\sqrt{2(1+\delta_{m,0})})(|jm\rangle \\
&\qquad \pm(-1)^{j-m}|j-m\rangle), m\geqslant 0,\ m\equiv 0(5)\},\\
&|j\pm\hat{1}\rangle = \{|jm\rangle,\ m\equiv\pm 1(5)\},\\
&|j\pm\hat{2}\rangle = \{|jm\rangle,\ m\equiv\pm 2(5)\}.
\end{aligned} \tag{2}$$

For half-integer values of j, there are five:

$$\begin{aligned}
&|j\pm\hat{\tfrac{1}{2}}\rangle = \{|jm\rangle,\ m\equiv\pm\tfrac{1}{2}(5)\},\\
&|j\pm\hat{\tfrac{3}{2}}\rangle = \{|jm\rangle,\ m\equiv\pm\tfrac{3}{2}(5)\},\\
&|j\,\hat{\tfrac{5}{2}}\rangle = \{|jm\rangle,\ m\equiv\tfrac{5}{2}(5)\}.
\end{aligned} \tag{3}$$

For integer values of j there are five irreducible representations of the group: Γ_1 to Γ_5 (Ref. 2) (A, T_1, T_2, U, and V, respectively, by other authors[4]). The behavior of Γ_1, Γ_2, and Γ_5 for all the rotations of the group is the same as the one of the irreducible representations of SU(2) for $j = 0$, 1, and 2, respectively. We identify their components $|\Gamma_i\ \hat{m}\rangle$ with the corresponding vectors $|jm\rangle$. The irreducible representation of SU(2) for $j = 3$ reduces to Γ_3 and Γ_4. We choose

$$|\Gamma_3\ \hat{0}\rangle = |3\ 0\rangle,$$

$$|\Gamma_3\pm\hat{2}\rangle = \sqrt{\tfrac{3}{5}}|3\pm 2\rangle \mp \sqrt{\tfrac{2}{5}}|3\mp 3\rangle \tag{4}$$

and

$$|\Gamma_4\pm\hat{1}\rangle = |3\pm 1\rangle,$$

$$|\Gamma_4\pm\hat{2}\rangle = \sqrt{\tfrac{2}{5}}|3\pm 2\rangle \pm \sqrt{\tfrac{3}{5}}|3\mp 3\rangle. \tag{5}$$

The behavior of these representations under a rotation of $2\pi/5$ around $0z$ and of π around $0y$, allows us to determine each component of these representations in a subspace defined above relatively to D_5 as shown in Table I.

 0022-2488/84/051187-08$02.50

Under the rotation $R_1 = R(\pi,\varphi,0)$ and the rotation $R_2 = R(0,\pi-\varphi,\pi)$ (using $\epsilon = 1$ for R_1 and $\epsilon = -1$ for R_2) the laws of transformation[5] are

$$|\Gamma_1\rangle \to |\Gamma_1\rangle, \tag{6a}$$

$$\begin{pmatrix} |\Gamma_2\ \hat{1}\rangle \\ |\Gamma_2\ \hat{0}\rangle \\ |\Gamma_2 - \hat{1}\rangle \end{pmatrix} \to \begin{pmatrix} -\frac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\frac{2}{5}} & -\frac{\sqrt{5}-\epsilon}{2\sqrt{5}} \\ \sqrt{\frac{2}{5}} & \frac{\epsilon}{\sqrt{5}} & -\epsilon\sqrt{\frac{2}{5}} \\ -\frac{\sqrt{5}-\epsilon}{2\sqrt{5}} & -\epsilon\sqrt{\frac{2}{5}} & -\frac{\sqrt{5}+\epsilon}{2\sqrt{5}} \end{pmatrix} \begin{pmatrix} |\Gamma_2\ \hat{1}\rangle \\ |\Gamma_2\ \hat{0}\rangle \\ |\Gamma_2 - \hat{1}\rangle \end{pmatrix}, \tag{6b}$$

$$\begin{pmatrix} |\Gamma_3\ \hat{2}\rangle \\ |\Gamma_3\ \hat{0}\rangle \\ |\Gamma_3 - \hat{1}\rangle \end{pmatrix} \to \begin{pmatrix} \frac{\sqrt{5}-\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\frac{2}{5}} & \frac{\sqrt{5}+\epsilon}{2\sqrt{5}} \\ \epsilon\sqrt{\frac{2}{5}} & -\frac{\epsilon}{\sqrt{5}} & \epsilon\sqrt{\frac{2}{5}} \\ -\frac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\frac{2}{5}} & -\frac{\sqrt{5}-\epsilon}{2\sqrt{5}} \end{pmatrix} \begin{pmatrix} |\Gamma_3\ \hat{2}\rangle \\ |\Gamma_3\ \hat{0}\rangle \\ |\Gamma_3 - \hat{1}\rangle \end{pmatrix}, \tag{6c}$$

$$\begin{pmatrix} |\Gamma_4\ \hat{2}\rangle \\ |\Gamma_4\ \hat{1}\rangle \\ |\Gamma_4 - \hat{1}\rangle \\ |\Gamma_4 - \hat{2}\rangle \end{pmatrix} \to \begin{pmatrix} -\frac{\epsilon}{\sqrt{5}} & -\frac{\sqrt{5}-\epsilon}{2\sqrt{5}} & -\frac{\sqrt{5}+\epsilon}{2\sqrt{5}} & -\frac{\epsilon}{\sqrt{5}} \\ -\frac{\sqrt{5}-\epsilon}{2\sqrt{5}} & \frac{\epsilon}{\sqrt{5}} & -\frac{\epsilon}{\sqrt{5}} & \frac{\sqrt{5}+\epsilon}{2\sqrt{5}} \\ -\frac{\sqrt{5}+\epsilon}{2\sqrt{5}} & -\frac{\epsilon}{\sqrt{5}} & \frac{\epsilon}{\sqrt{5}} & \frac{\sqrt{5}-\epsilon}{2\sqrt{5}} \\ -\frac{\epsilon}{\sqrt{5}} & \frac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \frac{\sqrt{5}-\epsilon}{2\sqrt{5}} & -\frac{\epsilon}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} |\Gamma_4\ \hat{2}\rangle \\ |\Gamma_4\ \hat{1}\rangle \\ |\Gamma_4 - \hat{1}\rangle \\ |\Gamma_4 - \hat{2}\rangle \end{pmatrix}, \tag{6d}$$

$$\begin{pmatrix} |\Gamma_5\ \hat{2}\rangle \\ |\Gamma_5\ \hat{1}\rangle \\ |\Gamma_5\ \hat{0}\rangle \\ |\Gamma_5 - \hat{1}\rangle \\ |\Gamma_5 - \hat{2}\rangle \end{pmatrix} \to \begin{pmatrix} \frac{3+\epsilon\sqrt{5}}{10} & -\frac{1+\epsilon\sqrt{5}}{5} & \frac{\sqrt{6}}{5} & \frac{1-\epsilon\sqrt{5}}{5} & \frac{3-\epsilon\sqrt{5}}{10} \\ -\frac{1+\epsilon\sqrt{5}}{5} & \frac{3-\epsilon\sqrt{5}}{10} & \frac{\sqrt{6}}{5} & -\frac{3+\epsilon\sqrt{5}}{10} & -\frac{1-\epsilon\sqrt{5}}{5} \\ \frac{\sqrt{6}}{5} & \frac{\sqrt{6}}{5} & -\frac{1}{5} & \frac{\sqrt{6}}{5} & \frac{\sqrt{6}}{5} \\ \frac{1-\epsilon\sqrt{5}}{5} & -\frac{3+\epsilon\sqrt{5}}{10} & -\frac{\sqrt{6}}{5} & \frac{3-\epsilon\sqrt{5}}{10} & \frac{1+\epsilon\sqrt{5}}{5} \\ \frac{3-\epsilon\sqrt{5}}{10} & -\frac{1-\epsilon\sqrt{5}}{5} & \frac{\sqrt{6}}{5} & \frac{1+\epsilon\sqrt{5}}{5} & \frac{3+\epsilon\sqrt{5}}{10} \end{pmatrix} \begin{pmatrix} |\Gamma_5\ \hat{2}\rangle \\ |\Gamma_5\ \hat{1}\rangle \\ |\Gamma_5\ \hat{0}\rangle \\ |\Gamma_5 - \hat{1}\rangle \\ |\Gamma_5 - \hat{2}\rangle \end{pmatrix}. \tag{6e}$$

For half-integer values of j, there are four additional irreducible representations: Γ_6 to Γ_9 (E', E'', U', and V', respectively, for other authors[3]). Among them, Γ_6, Γ_8, and Γ_9 transform like $j = \frac{1}{2}, \frac{3}{2}$ and $\frac{5}{2}$, respectively. The irreducible representation of SU(2) with $j = \frac{7}{2}$ includes a Γ_7 and a Γ_9. We choose

$$|\Gamma_7 \pm \tfrac{3}{2}\rangle = \sqrt{\tfrac{7}{10}}|\tfrac{7}{2} \pm \tfrac{3}{2}\rangle \pm \sqrt{\tfrac{3}{10}}|\tfrac{7}{2} \mp \tfrac{7}{2}\rangle. \tag{7}$$

The correspondence between the components of each representation and the subspaces of D_5 is given in Table II. As the matrix elements of the rotations $(\pi,\varphi,0)$ and $(0,\pi-\varphi,\pi)$ are purely imaginary, we multiply the result by i. The transformations $R_1 = iR(\pi,\varphi,0)$ and $R_2 = iR(0,\pi-\varphi,\pi)$ for these representations are

$$\begin{pmatrix} |\Gamma_6\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_6 - \hat{\tfrac{1}{2}}\rangle \end{pmatrix} \to \alpha \begin{pmatrix} 1 & \frac{1-\epsilon\sqrt{5}}{2} \\ \frac{1-\epsilon\sqrt{5}}{2} & -1 \end{pmatrix} \begin{pmatrix} |\Gamma_6\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_6 - \hat{\tfrac{1}{2}}\rangle' \end{pmatrix} \tag{8a}$$

$$\begin{pmatrix} |\Gamma_7\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_7\ -\hat{\tfrac{3}{2}}\rangle \end{pmatrix} \to \alpha \begin{pmatrix} -\dfrac{1-\epsilon\sqrt{5}}{2} & 1 \\ 1 & \dfrac{1-\epsilon\sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} |\Gamma_7\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_7\ -\hat{\tfrac{3}{2}}\rangle \end{pmatrix}, \tag{8b}$$

$$\begin{pmatrix} |\Gamma_8\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_8\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8\ -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8\ -\hat{\tfrac{3}{2}}\rangle \end{pmatrix} \to \alpha \begin{pmatrix} -\dfrac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\dfrac{3}{5}} & -\dfrac{\sqrt{15}-\epsilon\sqrt{3}}{2\sqrt{5}} & -\dfrac{\sqrt{5}-3\epsilon}{2\sqrt{5}} \\ \epsilon\sqrt{\dfrac{3}{5}} & -\dfrac{\sqrt{5}-3\epsilon}{2\sqrt{5}} & -\dfrac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \dfrac{\sqrt{15}-\epsilon\sqrt{3}}{2\sqrt{5}} \\ -\dfrac{\sqrt{15}-\epsilon\sqrt{3}}{2\sqrt{5}} & -\dfrac{\sqrt{5}+\epsilon}{2\sqrt{5}} & \dfrac{\sqrt{5}-3\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\dfrac{3}{5}} \\ -\dfrac{\sqrt{5}-3\epsilon}{2\sqrt{5}} & \dfrac{\sqrt{15}-\epsilon\sqrt{3}}{2\sqrt{5}} & \epsilon\sqrt{\dfrac{3}{5}} & \dfrac{\sqrt{5}+\epsilon}{2\sqrt{5}} \end{pmatrix} \begin{pmatrix} |\Gamma_8\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_8\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8\ -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_8\ -\hat{\tfrac{3}{2}}\rangle \end{pmatrix}, \tag{8c}$$

$$\begin{pmatrix} |\Gamma_9\ \hat{\tfrac{5}{2}}\rangle \\ |\Gamma_9\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_9\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{3}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{5}{2}}\rangle \end{pmatrix} \to \alpha \begin{pmatrix} \dfrac{3+\epsilon\sqrt{5}}{10} & -\dfrac{1+\sqrt{5}\epsilon}{2\sqrt{5}} & \sqrt{\dfrac{2}{5}} & \dfrac{1-\sqrt{5}\epsilon}{\sqrt{10}} & \dfrac{3-\epsilon\sqrt{5}}{2\sqrt{5}} & \dfrac{2-\epsilon\sqrt{5}}{5} \\ -\dfrac{1+\epsilon\sqrt{5}}{2\sqrt{5}} & \dfrac{\sqrt{5}-\epsilon}{2\sqrt{5}} & \dfrac{\sqrt{5}-\epsilon}{\sqrt{10}} & -\epsilon\sqrt{\dfrac{2}{5}} & \dfrac{\epsilon}{\sqrt{5}} & -\dfrac{3-\epsilon\sqrt{5}}{2\sqrt{5}} \\ \sqrt{\dfrac{2}{5}} & \dfrac{\sqrt{5}-\epsilon}{\sqrt{10}} & -\dfrac{\epsilon}{\sqrt{5}} & -\dfrac{\sqrt{5}-\epsilon}{2\sqrt{5}} & \epsilon\sqrt{\dfrac{2}{5}} & \dfrac{1-\epsilon\sqrt{5}}{\sqrt{10}} \\ \dfrac{1-\epsilon\sqrt{5}}{10} & -\epsilon\sqrt{\dfrac{2}{5}} & -\dfrac{\sqrt{5}-\epsilon}{2\sqrt{5}} & \dfrac{\epsilon}{\sqrt{5}} & \dfrac{\sqrt{5}-\epsilon}{\sqrt{10}} & -\sqrt{\dfrac{2}{3}} \\ \dfrac{3-\epsilon\sqrt{5}}{2\sqrt{5}} & \dfrac{\epsilon}{\sqrt{5}} & \epsilon\sqrt{\dfrac{2}{5}} & \dfrac{\sqrt{5}-\epsilon}{\sqrt{10}} & -\dfrac{\sqrt{5}-\epsilon}{2\sqrt{5}} & -\dfrac{1+\epsilon\sqrt{5}}{2\sqrt{5}} \\ \dfrac{2-\epsilon\sqrt{5}}{5} & -\dfrac{3-\epsilon\sqrt{5}}{2\sqrt{5}} & \dfrac{1-\epsilon\sqrt{5}}{\sqrt{10}} & -\sqrt{\dfrac{2}{5}} & -\dfrac{1+\epsilon\sqrt{5}}{2\sqrt{5}} & -\dfrac{3+\epsilon\sqrt{5}}{10} \end{pmatrix} \begin{pmatrix} |\Gamma_9\ \hat{\tfrac{5}{2}}\rangle \\ |\Gamma_9\ \hat{\tfrac{3}{2}}\rangle \\ |\Gamma_9\ \hat{\tfrac{1}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{1}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{3}{2}}\rangle \\ |\Gamma_9\ -\hat{\tfrac{5}{2}}\rangle \end{pmatrix}, \tag{8d}$$

where $\alpha^2 = (\sqrt{5}+\epsilon)/(2\sqrt{5})$, $\epsilon = 1$ for R_1 and $\epsilon = -1$ for R_2.

The reference frame chosen here is not the one used by Mc Lellan[2] but has the advantage of giving only real coefficients. In our frame, the rotation of π around $0y$ is an element of the group, whereas there was a rotation of π around $0x$ in the Mc Lellan's frame. It is easy to relate these different results by substituting $i^m|jm\rangle$ for $|jm\rangle$ in Mc Lellan's results.

We consider only the components $|\Gamma_i\ \hat{m}\rangle$ for nonnegative values of $\hat{m}$ because the other components are easily obtained by a rotation around $0y$:

$$R(0,\pi,0)|\Gamma_i\ \hat{m}\rangle = (-)^{j_i-\hat{m}}|\Gamma_i\ -\hat{m}\rangle, \tag{9a}$$

$$R(0,\pi,0)|jm\rangle = (-)^{j-m}|j\ -m\rangle. \tag{9b}$$

In the first of these equations, j_i is 0 for Γ_1 and Γ_5; 1 for Γ_2, Γ_3, and Γ_4; $\frac{1}{2}$ for Γ_6 and Γ_9; $\frac{3}{2}$ for Γ_7 and Γ_8.

II. PROJECTION OF A VECTOR $|jm\rangle$ ON AN IRREDUCIBLE REPRESENTATION OF THE DOUBLE ICOSAHEDRAL GROUP FOR HALF INTEGER VALUES OF j

Among the five subspaces (3) of $\mathscr{D}_j$ related to D_5, $|j\ -\hat{\tfrac{1}{2}}\rangle$ and $|j\ -\hat{\tfrac{3}{2}}\rangle$ can be obtained from $|j\ \hat{\tfrac{1}{2}}\rangle$ and $|j\ \hat{\tfrac{3}{2}}\rangle$ by a rotation of π around $0y$. On the contrary $|j\ \hat{\tfrac{5}{2}}\rangle$ is invariant under this transformation. As shown in Table II, any vector of the subspaces

$|j\ \hat{\tfrac{1}{2}}\rangle$ and $|j\ \hat{\tfrac{3}{2}}\rangle$ is the sum of at most three unnormalized vectors belonging to three of the four subspaces $\Gamma_6, \Gamma_7, \Gamma_8, \Gamma_9$ and any vector of the subspace $|j\ \hat{\tfrac{5}{2}}\rangle$ is the sum of two unnormalized vectors belonging to the $|\Gamma_9\ \hat{\tfrac{5}{2}}\rangle$ and the $|\Gamma_9 - \hat{\tfrac{5}{2}}\rangle$ subspaces. None of these two problems were encountered for the double cubic group.[1]

Let us consider a vector $|jm\rangle$ of the subspace $|j\ \hat{\tfrac{1}{2}}\rangle$:

$$m \equiv \tfrac{1}{2}(5) \quad |jm\rangle = |\Gamma_6\ \hat{\tfrac{1}{2}}\rangle_m + |\Gamma_8\ \hat{\tfrac{1}{2}}\rangle_m + |\Gamma_9\ \hat{\tfrac{1}{2}}\rangle_m, \tag{10}$$

where the subscript m reminds us that the vector is unnormalized and depends on m. The action of $R_1 = iR(\pi,\varphi,0)$ on $|jm\rangle$ yields

$$m \equiv \frac{1}{2}(5) \quad R_1|jm\rangle = \sum_{m'} (-)^{1/2-m'} d^{(j)}_{m'm}(\varphi)|jm'\rangle$$

$$= \cos\frac{\varphi}{2}\left\{\left|\Gamma_6\ \frac{\hat{1}}{2}\right\rangle_m + \sqrt{\frac{3}{5}}\left|\Gamma_8\ \frac{\hat{3}}{2}\right\rangle_m - \frac{\sqrt{5}-3}{2\sqrt{5}}\left|\Gamma_8\ \frac{\hat{1}}{2}\right\rangle_m + \cdots\right.$$

$$\left. + \sqrt{\frac{2}{5}}\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m + \frac{\sqrt{5}-1}{\sqrt{10}}\left|\Gamma_9\ \frac{\hat{3}}{2}\right\rangle_m - \frac{1}{\sqrt{5}}\left|\Gamma_9\ \frac{\hat{1}}{2}\right\rangle_m + \cdots + \frac{1-\sqrt{5}}{\sqrt{10}}\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_m\right\}, \tag{11}$$

and the action of $R_2 = iR(0,\pi-\varphi,\pi)$ on $|jm\rangle$ yields

$$m \equiv \frac{1}{2}(5) \quad R_2|jm\rangle = (-)^{1/2-m}\sum_{m'} d^{(j)}_{m'm}(\pi-\varphi)|jm'\rangle$$

$$= \sin\frac{\varphi}{2}\left\{\left|\Gamma_6\ \frac{\hat{1}}{2}\right\rangle_m - \sqrt{\frac{3}{5}}\left|\Gamma_8\ \frac{\hat{3}}{2}\right\rangle_m - \frac{\sqrt{5}+3}{2\sqrt{5}}\left|\Gamma_8\ \frac{\hat{1}}{2}\right\rangle_m + \cdots\right.$$

$$\left. + \sqrt{\frac{2}{5}}\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m + \frac{\sqrt{5}+1}{\sqrt{10}}\left|\Gamma_9\ \frac{\hat{3}}{2}\right\rangle_m + \frac{1}{\sqrt{5}}\left|\Gamma_9\ \frac{\hat{1}}{2}\right\rangle_m + \cdots + \frac{1+\sqrt{5}}{\sqrt{10}}\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_m\right\}. \tag{12}$$

A simple linear combination of (11) and (12) eliminates $|\Gamma_9\ \hat{\tfrac{3}{2}}\rangle_m$ and leads to

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_8\ \frac{\hat{3}}{2}\right\rangle_m = \sqrt{\frac{5}{3}}\sum_{m'\equiv\frac{3}{2}(5)}\left[(-)^{1/2-m'}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi) - (-)^{1/2-m}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right]|jm'\rangle, \tag{13}$$

and another leads to

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_9\ \frac{\hat{3}}{2}\right\rangle_m = \frac{1}{2}\sqrt{\frac{5}{2}}\sum_{m'\equiv\frac{3}{2}(5)}\left[(-)^{1/2-m'}(\sqrt{5}-1)\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi)\right.$$

$$\left. + (-)^{1/2-m}(\sqrt{5}+1)\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right]|jm'\rangle, \tag{14}$$

or

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m \equiv \sqrt{\frac{5}{2}}\sum_{m'\equiv\frac{5}{2}(5)}\left\{(-)^{1/2-m'}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi) + (-)^{1/2-m}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle. \tag{15}$$

Taking into account the three equations (10)–(12), we get

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_6\ \frac{\hat{1}}{2}\right\rangle_m = \frac{1}{6}\sum_{m'\equiv\frac{1}{2}(5)}\left\{\delta_{mm'} + (-)^{1/2-m'}\,5\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi) + (-)^{1/2-m}\,5\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle, \tag{16}$$

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_8\ \frac{\hat{1}}{2}\right\rangle_m = \frac{1}{6}\sum_{m'\equiv\frac{1}{2}(5)}\left\{2\delta_{mm'} + (-)^{1/2-m'}(3\sqrt{5}-5)\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi)\right.$$

$$\left. - (-)^{1/2-m}(3\sqrt{5}+5)\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle, \tag{17}$$

$$m \equiv \frac{1}{2}(5) \quad \left|\Gamma_9\ \frac{\hat{1}}{2}\right\rangle_m = \frac{1}{2}\sum_{m'\equiv\frac{1}{2}(5)}\left\{\delta_{mm'} - (-)^{1/2-m'}\sqrt{5}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi)\right.$$

$$\left. + (-)^{1/2-m}\sqrt{5}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle. \tag{18}$$

In a similar way, projections of Γ_7, Γ_8, and Γ_9 representations can be obtained from any vector of the subspace $|j\ \hat{\tfrac{3}{2}}\rangle$. To project from an arbitrary vector $|jm\rangle$ of the subspace $|j\ \hat{\tfrac{5}{2}}\rangle$, we have to consider together

$$m \equiv \frac{5}{2}(5) \quad |jm\rangle = |\Gamma_9\ \hat{\tfrac{5}{2}}\rangle_m - (-)^{j-m}|\Gamma_9 - \hat{\tfrac{5}{2}}\rangle_{-m},$$

$$|j-m\rangle = (-)^{j-m}|\Gamma_9 - \hat{\tfrac{5}{2}}\rangle_m + |\Gamma_9\ \hat{\tfrac{5}{2}}\rangle_{-m}, \tag{19}$$

i.e., two vectors related by a rotation of π around the $0y$ axis. The representations $|\Gamma_9\rangle_m$ and $|\Gamma_9\rangle_{-m}$ are different: $|\Gamma_9\rangle_m$ is the one for which the vector $|\Gamma_9\ \hat{\frac{5}{2}}\rangle_m$ can be projected from $|jm\rangle$ and $|\Gamma_9\rangle_{-m}$ the one for which the vector $|\Gamma_9 - \hat{\frac{5}{2}}\rangle_{-m}$ can be projected from the same vector $|jm\rangle$.

As far as the subspace $|j\ \hat{\frac{5}{2}}\rangle$ is concerned, only the first equation (19) is needed. In this subspace, the action of R_1 and R_2 yields, respectively,

$$m \equiv \frac{5}{2}(5) \quad R_1|jm\rangle = \sum_{m'} (-)^{1/2-m'} d^{(j)}_{m'm}(\varphi)|jm'\rangle$$

$$= \cos\frac{\varphi}{2}\left\{\frac{3+\sqrt{5}}{10}\left(\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m + (-)^{j-m}\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_{-m}\right)\right.$$

$$\left. + \frac{2-\sqrt{5}}{5}\left(\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_m - (-)^{j-m}\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_{-m}\right) + \cdots\right\}, \tag{20a}$$

$$m \equiv \frac{5}{2}(5) \quad R_2|jm\rangle = (-)^{1/2-m}\sum_{m'} d^{(j)}_{m'm}(\pi-\varphi)|jm'\rangle$$

$$= \sin\frac{\varphi}{2}\left\{\frac{3-\sqrt{5}}{10}\left(\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m + (-)^{j-m}\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_{-m}\right)\right.$$

$$\left. + \frac{2+\sqrt{5}}{5}\left(\left|\Gamma_9 - \frac{\hat{5}}{2}\right\rangle_m - (-)^{j-m}\left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_{-m}\right) + \cdots\right\}. \tag{20b}$$

From these equations, we get

$$m \equiv \frac{5}{2}(5) \quad \left|\Gamma_9\ \frac{\hat{5}}{2}\right\rangle_m = \frac{1}{2}\sum_{m' \equiv \frac{5}{2}(5)}\left\{\delta_{mm'} + (-)^{1/2-m'}\frac{3+\sqrt{5}}{2}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi)\right.$$

$$\left. + (-)^{1/2-m}\frac{3-\sqrt{5}}{2}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle. \tag{21}$$

The two equations (19) and the symmetry properties of reduced matrix elements are needed to obtain the other components of a representation Γ_9 from a vector of the subspace $|j\ \hat{\frac{5}{2}}\rangle$.

Equations (13)–(18) and (21) are simpler when the reduced rotation matrix elements are expressed with the matrices:

$$A^{(j)}_{m'm} = \frac{1}{2}\left\{(-)^{1/2-m'}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi) + (-)^{1/2-m}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}, \tag{22a}$$

$$B^{(j)}_{m'm} = \frac{\sqrt{5}}{2}\left\{(-)^{1/2-m'}\cos\frac{\varphi}{2}d^{(j)}_{m'm}(\varphi) - (-)^{1/2-m}\sin\frac{\varphi}{2}d^{(j)}_{m'm}(\pi-\varphi)\right\}. \tag{22b}$$

With these notations, the projectors on Γ_6, Γ_7, Γ_8, and Γ_9 are, respectively,

$$P_6 = |\tfrac{1}{6}I + \tfrac{5}{3}A| \quad |j\ \hat{\tfrac{1}{2}}\rangle, \tag{23a}$$

$$P_7 = |\tfrac{1}{6}I - \tfrac{5}{6}A + \tfrac{5}{6}B| \quad |j\ \tfrac{3}{2}\rangle, \tag{23b}$$

$$P_8 = \left|\begin{array}{cc} \frac{1}{3}I - \frac{5}{3}A - \frac{1}{3}B & (2/\sqrt{3})B \\ (2/\sqrt{3})B & \frac{1}{3}I - \frac{5}{3}A + B \end{array}\right| \begin{array}{c} |j\ \hat{\frac{3}{2}}\rangle \\ |j\ \hat{\frac{1}{2}}\rangle \end{array}, \tag{23c}$$

$$P_9 = \left|\begin{array}{ccc} \frac{1}{2}I + \frac{3}{2}A + \frac{1}{2}B & -(\sqrt{5}/2)(A+B) & \sqrt{10}A \\ -(\sqrt{5}/2)(A+B) & \frac{1}{2}I + \frac{5}{2}A - \frac{1}{2}B & (1/\sqrt{2})(5A-B) \\ \sqrt{10}A & (1/\sqrt{2})(5A-B) & \frac{1}{2}I - B \end{array}\right| \begin{array}{c} |j\ \hat{\frac{5}{2}}\rangle \\ |j\ \hat{\frac{3}{2}}\rangle \\ |j\ \hat{\frac{1}{2}}\rangle \end{array}. \tag{23d}$$

The first column of P_9 deals with the $|\Gamma_9\ \hat{\frac{5}{2}}\rangle$ vector. The zero matrix elements are not indicated and the subspaces of D_5 are written on the right.

III. PROJECTION OF A VECTOR $|jm\rangle$ ON AN IRREDUCIBLE REPRESENTATION OF THE ICOSAHEDRAL GROUP FOR INTEGER VALUES OF j

As seen in Table I, any vector of the subspaces $|\hat{0}+\rangle$ and $|\hat{0}-\rangle$ can be projected on two irreducible representations of the icosahedral group and any vector of the other subspaces can be projected on three irreducible representations. It seems

that only one of the two transformations $R_1 = R(\pi,\varphi,0)$ or $R_2 = R(0,\pi-\varphi,\pi)$ is needed in the first case and the two of them are needed in the second case. However, the matrix elements of both matrices are involved, due to the symmetry

$$d^{(j)}_{mm'}(\varphi) = (-)^{j-m'} d^{(j)}_{-mm'}(\pi-\varphi). \tag{24}$$

Let us consider a vector $|jm\rangle$ of the subspace $|j\,\hat{0}+\rangle$ as defined by (2). It can be written

$$m \equiv 0(5) \quad |jm+\rangle = |\Gamma_1\ \hat{0}\rangle_m + |\Gamma_5\ \hat{0}\rangle_m. \tag{25}$$

The action of R_1 gives

$$\begin{aligned} m \equiv 0(5) \quad R_1|jm+\rangle &= \frac{1}{\sqrt{2(1+\delta_{m0})}}\left\{\sum_{\substack{m'\equiv 0(5)\\ m'>0}} \frac{2}{\sqrt{2(1+\delta_{m'0})}}\left[(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right]|jm'+\rangle \right. \\ &\quad \left. + \sum_{m'\neq 0(5)} \left[(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right]|jm'\rangle \right\} \\ &= |\Gamma_1\ \hat{0}\rangle_m + \frac{\sqrt{6}}{5}\{|\Gamma_5\ \hat{2}\rangle_m + |\Gamma_5\ \hat{1}\rangle_m + |\Gamma_5 - \hat{1}\rangle_m + |\Gamma_5 - \hat{2}\rangle_m\} - \frac{1}{5}|\Gamma_5\ \hat{0}\rangle_m. \end{aligned} \tag{26}$$

Therefore, for $M \neq 0$,

$$m \equiv 0(5) \quad |\Gamma_5\ \hat{M}\rangle = \frac{5}{\sqrt{12(1+\delta_{m0})}} \sum_{m'\equiv M(5)} \left\{(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right\}|jm'\rangle. \tag{27}$$

By linear combination of (25) and (26), we get

$$m \equiv 0(5) \quad |\Gamma_1\ \hat{0}\rangle_m = \frac{1}{6} \sum_{m'\equiv 0(5)} \left\{\delta_{m'm} + \frac{5}{\sqrt{(1+\delta_{m0})(1+\delta_{m'0})}}\left[(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right]\right\}|jm'+\rangle, \tag{28}$$

and

$$m \equiv 0(5) \quad |\Gamma_5\ \hat{0}\rangle_m = \frac{1}{6} \sum_{m'\equiv 0(5)} \left\{5\delta_{m'm} - \frac{1}{\sqrt{(1+\delta_{m0})(1+\delta_{m'0})}}\left[(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right]\right\}|jm'+\rangle. \tag{29}$$

The projection of a vector of the subspace $|j\ \hat{1}\rangle$ on the representations Γ_2, Γ_4, Γ_5 or of a vector of the subspace $|j\ \hat{2}\rangle$ on the representations Γ_3, Γ_4, Γ_5 is similar to what was done in the subspace $|j\,\frac{1}{2}\rangle$.

In order to simplify the equations, we introduce the matrices

$$C^{(j)}_{m'm} = \frac{1}{2}\frac{1}{\sqrt{(1+\delta_{m0})(1+\delta_{m'0})}}\left\{(-)^{m'} d^{(j)}_{m'm}(\varphi) + (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right\}, \tag{30a}$$

$$D^{(j)}_{m'm} = \frac{\sqrt{5}}{2}\frac{1}{\sqrt{(1+\delta_{m0})(1+\delta_{m'0})}}\left\{(-)^{m'} d^{(j)}_{m'm}(\varphi) - (-)^m d^{(j)}_{m'm}(\pi-\varphi)\right\}. \tag{30b}$$

With them, the projectors are

$$P_1 = |\tfrac{1}{6}I + \tfrac{5}{3}C| \quad |j\ \hat{0}+\rangle, \tag{31a}$$

$$P_2 = \left|\begin{matrix} \frac{1}{4}I - \frac{5}{4}C - \frac{1}{4}D & D \\ D & \frac{1}{2}I + D \end{matrix}\right| \begin{matrix} |j\ \hat{1}\rangle \\ |j\ \hat{0}-\rangle' \end{matrix}, \tag{31b}$$

$$P_3 = \left|\begin{matrix} \frac{1}{4}I - \frac{5}{4}C + D & D \\ D & \frac{1}{2}I - D \end{matrix}\right| \begin{matrix} |j\,\hat{2}\rangle \\ |j\,\hat{0}-\rangle' \end{matrix}, \tag{31c}$$

$$P_4 = \begin{vmatrix} \frac{1}{3}I - \frac{2}{3}D & -\frac{1}{3}(5C - D) \\ -\frac{1}{3}(5C - D) & \frac{1}{3}I + \frac{2}{3}D \end{vmatrix} \begin{matrix} |j\ \hat{2}\rangle \\ |j\ \hat{1}\rangle \end{matrix}, \tag{31d}$$

$$P_5 = \begin{vmatrix} \frac{5}{12}I + \frac{5}{4}C + \frac{5}{12}D & -\frac{5}{6}(C + D) & (5/\sqrt{3})C \\ -\frac{5}{6}(C + D) & \frac{5}{12}I + \frac{5}{4}C - \frac{5}{12}D & (5/\sqrt{3})C \\ (5/\sqrt{3})C & (5/\sqrt{3})C & \frac{5}{6}I - \frac{5}{3}C \end{vmatrix} \begin{matrix} |j\ \hat{2}\rangle \\ |j\ \hat{1}\rangle \\ |j\ \hat{0}+\rangle \end{matrix}. \tag{31e}$$

The components $|\Gamma_i - \hat{1}\rangle$ and $|\Gamma_i - \hat{2}\rangle$ are obtained by the relation (9a).

Orthogonalization and normalization of the representations are discussed in detail in Ref. 1.

IV. STRUCTURE OF COEFFICIENTS

Reduced rotation matrix elements are

$$d^{(j)}_{mm'}(\varphi) = \sqrt{(j+m)!(j-m)!(j+m')!(j-m')!} \times \sum_t \frac{(-)^t}{(j+m-t)!t!(j-m'-t)!(t-m+m')!} \left(\cos\frac{\varphi}{2}\right)^{2j+m-m'-2t} \left(\sin\frac{\varphi}{2}\right)^{2t-m+m'}, \tag{32}$$

and the angle φ is such that

$$\cos^2\frac{\varphi}{2} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{5}}\right), \quad \sin^2\frac{\varphi}{2} = \frac{1}{2}\left(1 - \frac{1}{\sqrt{5}}\right). \tag{33}$$

Therefore,

$$A^{(j)}_{mm'} = (-)^m \frac{1}{2^{j+3/2}} \sqrt{\frac{(j+m)!(j-m)!}{(j+m')!(j-m')!}} \sum_t \frac{(-)^t (j+m')!(j-m')!}{(j+m-t)!t!(j-m'-t)!(t-m+m')!} \times \left\{ \left(1 + \frac{1}{\sqrt{5}}\right)^{(2j+m-m'-2t+1)/2} \left(1 - \frac{1}{\sqrt{5}}\right)^{(2t-m+m')/2} + (-)^{m-m'} \left(1 - \frac{1}{\sqrt{5}}\right)^{(2j+m-m'-2t+1)/2} \left(1 + \frac{1}{\sqrt{5}}\right)^{(2t-m+m')/2} \right\}. \tag{34}$$

When $m - m'$ is even, all the powers in the expression between brackets are integers, the square roots cancel and the result of the sum is a rational number, the denominator of which is some power of 5. When $m - m'$ is odd, $\cos(\varphi/2)\sin(\varphi/2) = 1/\sqrt{5}$ can be factorized out and, with the sign $(-)^{m-m'}$, the result of the bracket is still a rational number.

A similar analysis can be performed on the matrix B given by (22b) as well as on the matrices C and D given by (30). As a result, a matrix element of a projector (23) or (31) is the square root which appears in (34) multiplying the ratio of two integers, the one of the denominator being essentially some power of 5 because the powers of 2 usually cancel. The square root reduces to unity for the diagonal matrix elements; the numerator can be a very large number for a large value of j, quicker than for cubic harmonics.

Any recursion relation for the reduced rotation matrix elements can be cast into a coupled recursion relation for the A and B or the C and D matrix elements.

After orthogonalization and normalization as explained in Ref. 1, each coefficient is the product of a rational number by the square root of a product of prime numbers. If the ratio of the coefficients of some $|jm\rangle$ for two representations of the same type is written in this manner, the square root is independent of the choice of $|jm\rangle$. If the ratio of the coefficients of $|jm_1\rangle$ and $|jm_2\rangle$ for some representation is written in this manner, the prime numbers of the square root are those which appear with an odd power in the product $(j+m_1)!(j-m_1)!(j+m_2)!(j-m_2)!$

The first degeneracy of Γ_1 representations happens for $j = 30$. Up to this value, the coefficients are very simple but we could find no linear combination of the two Γ_1 for $j = 30$

TABLE I. Number of components of each representation of the icosahedral group in each subspace of D_5 for integer values of j.

K / D_5	Γ_1 A	Γ_2 T_1	Γ_3 T_2	Γ_4 U	Γ_5 V
$\|\hat{0}+\rangle$	1				1
$\|\hat{0}-\rangle$		1	1		
$\|\hat{1}\rangle$		1		1	1
$\|\hat{2}\rangle$			1	1	1

TABLE II. Number of components of each representation of the icosahedral group in each subspace of D_5 for half-integer values of j.

K / D_5	Γ_6 E'	Γ_7 E''	Γ_8 U'	Γ_9 V'
$\|\hat{\frac{1}{2}}\rangle$	1		1	1
$\|\hat{\frac{3}{2}}\rangle$		1	1	1
$\|\hat{\frac{5}{2}}\rangle$				2

with simple coefficients.

Some other representations can be obtained from Γ_1. If $\Gamma_1(l)$ is a representation for a given value of l, $\sqrt{(2s+1)(2l+1)}\times\{\Gamma_1(l)|s\hat{\sigma}\rangle\}_j/\sqrt{2j+1}$, where $\{\ \}_j$ means coupling by a Clebsch–Gordan coefficient l with s to j is the $\hat{\sigma}$ component of another representation. In this manner, we can obtain Γ_6 with $s=\frac{1}{2}$, Γ_2 with $s=1$, Γ_8 with $s=\frac{3}{2}$, Γ_5 with $s=2$, and Γ_9 with $s=\frac{5}{2}$. These representations are orthonormalized. As their coefficients can be expressed by multiplying the coefficient of some Γ_1 representation by a Clebsch–Gordan coefficient, we consider that tables of Γ_2, Γ_5, Γ_6, Γ_8, and Γ_9 are not needed. Similarly Γ_4 can be expressed in terms of Γ_7 representations. More details and tables of Γ_1, Γ_7, and Γ_3 representations will be given in a forthcoming article.

ACKNOWLEDGMENT

We thank the scientific secretariat of the Service de Physique Théorique who typed tables of Γ_1 to Γ_9 coefficients up to $j=18$ before the simplification noted above was noticed.

[1] R. Conte, J. Raynal, and E. Soulié, J. Math. Phys. **25**, 1176 (1984).
[2] A. G. Mc Lellan, J. Chem. Phys. **34**, 1350 (1961).
[3] B. R. Judd, Proc. R. Soc. London Ser. A **241**, 122 (1957).
[4] J. S. Griffith, *The Theory of Transition Metal Ions* (Cambridge U. P., Cambridge, 1964).
[5] D. A. Varshalovich, A. N. Moskaliev, and V. K. Khersonskii, *Quantum Theory of Angular Momenta*, in Russian (Nauka, Leningrad, 1975).

On a class of spinor representations of SO(7)

P. De Wilde,[a] H. De Meyer,[b] J. Van der Jeugt,[c] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 7 June 1983; accepted for publication 16 September 1983)

The reduction of the irreducible representations $[v,0,...,0,1]$ of $SO(2n+1)$ with respect to $SU(2) \otimes SU(2) \otimes SO(2n-3)$ is considered. For the $n=3$ case all the reduced matrix elements of the SO(7) generators in the $[SU(2)]^2 \otimes SO(3)$ basis are calculated with the use of recursion relations.

PACS numbers: 02.20. + b

1. INTRODUCTION

In previous papers[1,2] the symmetric representations of $SO(2n+1)$ and of SO(7) in particular have been investigated in detail. A method originally due to Kemmer *et al.*[3] has been applied to SO(7) in order to calculate reduced matrix elements of the generators in a $[SU(2)]^2 \otimes SO(3)$ basis.

There exists a second class of $SO(2n+1)$ irreps, i.e., $[v,0,...,0,1]$ $(v \geqslant 0)$ which decompose without degeneracy into $SO(2n-3) \otimes SU(2) \otimes SU(2)$ irreps. In Sec. 2 the corresponding branching rule is given. A set of recursion relations between the reduced matrix elements of the SO(7) generators in established. In Sec. 3 a complete solution is derived.

2. RECURSION RELATIONS FOR THE REDUCED MATRIX ELEMENTS

Apart from the symmetric representations $[v,0,...,0]$ of $SO(2n+1)$ one more class of representations labeled by $[v,0,...,0,1]$ $(v \geqslant 0)$ reduces without degeneracy into $SU(2) \otimes SU(2) \otimes SO(2n-3)$ irreps. Indeed, it can be easily shown, for instance by using the method outlined in De Meyer *et al.*,[1] that the corresponding branching rule reads with the same notation

$$[v,0,...,0,1] \rightarrow \sum_{s,t,u} (s,t,[u,0,...,0,1])'$$

with $u = v, v-1,...,1,0,$

$$s+t = v-u+\tfrac{1}{2}, v-u-\tfrac{1}{2},...,\tfrac{1}{2},$$

$$|s-t| = \tfrac{1}{2} \quad (n>3). \tag{2.1a}$$

Due to the specific labeling of SO(3) irreps, we have to replace (2.1a) for $n=3$ by

$$[v,0,1] \rightarrow \sum_{s,t,u} (s,t,u)',$$

with

$$u = v+\tfrac{1}{2}, \quad v-\tfrac{1}{2},...,\tfrac{1}{2},$$

$$s+t = v+1-u, v-u,...,\tfrac{1}{2},$$

$$|s-t| = \tfrac{1}{2}. \tag{2.1b}$$

For a verification of the rules (2.1) for particular v-values we refer the reader to branching rule tables such as the one of McKay and Patera.[4] Let us recall that for SO(7) s, t and u are the SO(3) labels which allow us to denote the SO(7) states by $|s\,t\,u\,\lambda\,\mu\,\nu\rangle$, where $s(s+1), t(t+1), u(u+1), \lambda, \mu,$ and ν are the eigenvalues of $s^2, t^2, u^2, s_0, t_0, u_0$, respectively. Clearly the three commuting subgroups are generated by the sets $\{s_{+1}, s_0, s_{-1}\}$, $\{t_{+1}, t_0, t_{-1}\}$, and $\{u_{+1}, u_0, u_{-1}\}$. The reduced matrix elements of these generators in the $[SU(2)]^3$ basis $\{|s\,t\,u\rangle\}$ are standard. The remaining SO(7) generators which are combined into a bispinor vector $T^{[1/2\ 1/2\ 1]}$ are defined in Vanden Berghe *et al.*[2] We denote a set of independent reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$ in the $[SU(2)]^2 \otimes SO(3)$ basis as follows:

$$A(s,u) = \langle s+\tfrac{1}{2}\ s+1\ u+1\ \|T^{[1/2\ 1/2\ 1]}\|\ s\ s+\tfrac{1}{2}\ u\rangle, \tag{2.2}$$

$$B(s,u) = \langle s+\tfrac{1}{2}\ s+1\ u\ \|T^{[1/2\ 1/2\ 1]}\|\ s\ s+\tfrac{1}{2}\ u\rangle, \tag{2.3}$$

$$C(s,u) = \langle s+\tfrac{1}{2}\ s+1\ u-1\ \|T^{[1/2\ 1/2\ 1]}\|\ s\ s+\tfrac{1}{2}\ u\rangle, \tag{2.4}$$

$$D(s,u) = \langle s+\tfrac{1}{2}\ s\ u+1\ \|T^{[1/2\ 1/2\ 1]}\|\ s\ s+\tfrac{1}{2}\ u\rangle, \tag{2.5}$$

$$E(s,u) = \langle s+\tfrac{1}{2}\ s\ u\ \|T^{[1/2\ 1/2\ 1]}\|\ s\ s+\tfrac{1}{2}\ u\rangle, \tag{2.6}$$

where the other reduced matrix elements of $T^{[1/2\ 1/2\ 1]}$ can be related to the preceding ones on account of symmetry operations with respect to the first two labels and of the property

$$\langle s't'u'\|T^{[1/2\ 1/2\ 1]}\|stu\rangle^* = (-1)^{s'-s+t'-t+u'-u}\langle s\,t\,u\|T^{[1/2\ 1/2\ 1]}\|s'\,t'\,u'\rangle. \tag{2.7}$$

[a] Research Assistant I.W.O.N.L. (Belgium).
[b] Research Associate N.F.W.O. (Belgium).
[c] Research Assistant N.F.W.O. (Belgium).

 0022-2488/84/051195-04$02.50

Equations (3.3)–(3.7) of Vanden Berghe *et al.*[2] remain valid and we can apply again Eq. (15.23) of De Shalit and Talmi[5] to each of them. Using the well-known reduced marix elements of $\vec{s}$, $\vec{t}$, and $\vec{u}$ and the explict analytical formulas for the occurring 6*j*-symbols, we obtain the following set of 24 nonlinear recursion relations between the reduced matrix elements (2.2)–(2.6):

$$2s[|A(s,u)|^2+|B(s,u)|^2+|C(s,u)|^2]-(2s+2)[|A(s-\tfrac{1}{2},u-1)|^2+|B(s-\tfrac{1}{2},u)|^2+|C(s-\tfrac{1}{2},u+1)|^2]$$
$$+2s[|D(s,u)|^2+|D(s,u-1)|^2+|E(s,u)|^2]=-\tfrac{3}{2}2s(2s+1)(2s+2)^2(2u+1), \tag{2.8}$$

$$(2s+1)[|A(s,u)|^2+|B(s,u)|^2+|C(s,u)|^2]-(2s+3)[|A(s-\tfrac{1}{2},u-1)|^2+|B(s-\tfrac{1}{2},u)|^2+|C(s-\tfrac{1}{2},u+1)|^2]$$
$$-(2s+3)[|D(s,u)|^2+|D(s,u-1)|^2+|E(s,u)|^2]=-\tfrac{3}{2}(2s+1)^2(2s+2)(2s+3)(2u+1), \tag{2.9}$$

$$u|A(s,u)|^2-|B(s,u)|^2-(u+1)|C(s,u)|^2-(u+1)|A(s-\tfrac{1}{2},u-1)|^2-|B(s-\tfrac{1}{2},u)|^2+u|C(s-\tfrac{1}{2},u+1)|^2$$
$$+u|D(s,u)|^2-(u+1)|D(s,u-1)|^2-|E(s,u)|^2=-2(2s+1)(2s+2)u(u+1)(2u+1), \tag{2.10}$$

$$2s(2s+1)[u|A(s,u)|^2-|B(s,u)|^2-(u+1)|C(s,u)|^2]$$
$$+(2s+2)(2s+3)[-(u+1)|A(s-\tfrac{1}{2},u-1)|^2-|B(s-\tfrac{1}{2},u)|^2+u|C(s-\tfrac{1}{2},u+1)|^2]$$
$$-2s(2s+3)[u|D(s,u)|^2-(u+1)|D(s,u-1)|^2-|E(s,u)|^2]=0, \tag{2.11}$$

$$2s[u(2u-1)|A(s,u)|^2-(2u-1)(2u+3)|B(s,u)|^2+(u+1)(2u+3)|C(s,u)|^2]$$
$$-(2s+2)[(u+1)(2u+3)|A(s-\tfrac{1}{2},u-1)|^2-(2u-1)(2u+3)|B(s-\tfrac{1}{2},u)|^2+u(2u-1)|C(s-\tfrac{1}{2},u+1)|^2]$$
$$+2s[u(2u-1)|D(s,u)|^2+(u+1)(2u+3)|D(s,u-1)|^2-(2u-1)(2u+3)|E(s,u)|^2]=0, \tag{2.12}$$

$$(2s+1)[u(2u-1)|A(s,u)|^2-(2u-1)(2u+3)|B(s,u)|^2+(u+1)(2u+3)|C(s,u)|^2]$$
$$-(2s+3)[(u+1)(2u+3)|A(s-\tfrac{1}{2},u-1)|^2-(2u-1)(2u+3)|B(s-\tfrac{1}{2},u)|^2+u(2u-1)|C(s-\tfrac{1}{2},u+1)|^2]$$
$$-(2s+3)[u(2u-1)|D(s,u)|^2+(u+1)(2u+3)|D(s,u-1)|^2-(2u-1)(2u+3)|E(s,u)|^2]=0, \tag{2.13}$$

$$-D^*(s+\tfrac{1}{2},u)A(s,u)-E(s+\tfrac{1}{2},u)B(s,u)+D(s+\tfrac{1}{2},u-1)C(s,u)$$
$$=C(s,u+1)D(s,u)-B(s,u)E(s,u)-A(s,u-1)D^*(s,u-1), \tag{2.14}$$

$$-u(2u-1)D^*(s+\tfrac{1}{2},u)A(s,u)+(2u-1)(2u+3)E(s+\tfrac{1}{2},u)B(s,u)+(u+1)(2u+3)D(s+\tfrac{1}{2},u-1)C(s,u)$$
$$=u(2u-1)C(s,u+1)D(s,u)+(2u-1)(2u+3)B(s,u)E(s,u)-(u+1)(2u+3)A(s,u-1)D^*(s,u-1), \tag{2.15}$$

$$(2s+1)[-uD^*(s+\tfrac{1}{2},u)A(s,u)+E(s+\tfrac{1}{2},u)B(s,u)-(u+1)D(s+\tfrac{1}{2},u-1)C(s,u)]$$
$$=-(2s+3)[uC(s,u+1)D(s,u)+B(s,u)E(s,u)+(u+1)A(s,u-1)D^*(s,u-1)], \tag{2.16}$$

$$[(u+2)(2u+1)]^{1/2}B^*(s,u+1)A(s,u)+[u(2u+3)]^{1/2}C^*(s,u+1)B(s,u)$$
$$-[(u+2)(2u+1)]^{1/2}C^*(s-\tfrac{1}{2},u+1)B(s-\tfrac{1}{2},u+1)-[u(2u+3)]^{1/2}B^*(s-\tfrac{1}{2},u)A(s-\tfrac{1}{2},u)$$
$$+[(u+2)(2u+1)]^{1/2}E(s,u+1)D(s,u)-[u(2u+3)]^{1/2}D(s,u)E(s,u)=0, \tag{2.17}$$

$$2s(2s+1)\{[(u+2)(2u+1)]^{1/2}B^*(s,u+1)A(s,u)+[u(2u+3)]^{1/2}C^*(s,u+1)B(s,u)\}$$
$$-(2s+2)(2s+3)\{[(u+2)(2u+1)]^{1/2}C^*(s-\tfrac{1}{2},u+1)B(s-\tfrac{1}{2},u+1)+[u(2u+3)]^{1/2}B^*(s-\tfrac{1}{2},u)A(s-\tfrac{1}{2},u)\}$$
$$-2s(2s+3)\{[(u+2)(2u+1)]^{1/2}E(s,u+1)D(s,u)-[u(2u+3)]^{1/2}D(s,u)E(s,u)\}=0, \tag{2.18}$$

$$2s\{[u(2u+1)]^{1/2}B^*(s,u+1)A(s,u)-[(u+2)(2u+3)]^{1/2}C^*(s,u+1)B(s,u)\}$$
$$+(2s+2)\{[u(2u+1)]^{1/2}C^*(s-\tfrac{1}{2},u+1)B(s-\tfrac{1}{2},u+1)-[(u+2)(2u+3)]^{1/2}B^*(s-\tfrac{1}{2},u)A(s-\tfrac{1}{2},u)\}$$
$$+2s\{[u(2u+1)]^{1/2}E(s,u+1)D(s,u)+[(u+2)(2u+3)]^{1/2}D(s,u)E(s,u)\}=0, \tag{2.19}$$

$$(2s+1)\{[u(2u+1)]^{1/2}B^*(s,u+1)A(s,u)-[(u+2)(2u+3)]^{1/2}C^*(s,u+1)B(s,u)\}$$
$$+(2s+3)\{[u(2u+1)]^{1/2}C^*(s-\tfrac{1}{2},u+1)B(s-\tfrac{1}{2},u+1)-[(u+2)(2u+3)]^{1/2}B^*(s-\tfrac{1}{2},u)A(s-\tfrac{1}{2},u)\}$$
$$-(2s+3)\{[u(2u+1)]^{1/2}E(s,u+1)D(s,u)+[(u+2)(2u+3)]^{1/2}D(s,u)E(s,u)\}=0, \tag{2.20}$$

$$[u(2u+1)]^{1/2}E(s+\tfrac{1}{2},u+1)A(s,u)+[(u+2)(2u+3)]^{1/2}D(s+\tfrac{1}{2},u)B(s,u)$$
$$=[u(2u+1)]^{1/2}B(s,u+1)D(s,u)+[(u+2)(2u+3)]^{1/2}A(s,u)E(s,u), \tag{2.21}$$

$$(2s+1)\{[(u+2)(2u+1)]^{1/2}E(s+\tfrac{1}{2},u+1)A(s,u)-[u(2u+3)]^{1/2}D(s+\tfrac{1}{2},u)B(s,u)\}$$
$$=-(2s+3)\{[(u+2)(2u+1)]^{1/2}B(s,u+1)D(s,u)-[u(2u+3)]^{1/2}A(s,u)E(s,u)\}, \tag{2.22}$$

$$-[(u-1)(2u-1)]^{1/2}D^*(s+\tfrac{1}{2},u-1)B(s,u)+[(u+1)(2u+1)]^{1/2}E(s+\tfrac{1}{2},u-1)C(s,u)$$
$$=[(u-1)(2u-1)]^{1/2}C(s,u)E(s,u)-[(u+1)(2u+1)]^{1/2}B(s,u-1)D^*(s,u-1), \tag{2.23}$$

$$(2s+1)\{[(u+1)(2u-1)]^{1/2}D^*(s+\tfrac{1}{2},u-1)B(s,u)+[(u-1)(2u+1)]^{1/2}E(s+\tfrac{1}{2},u-1)C(s,u)\}$$
$$=(2s+3)\{[(u+1)(2u-1)]^{1/2}C(s,u)E(s,u)+[(u-1)(2u+1)]^{1/2}B(s,u-1)D^*(s,u-1)\}, \tag{2.24}$$

$$uA(s,u)C(s+\tfrac{1}{2},u+1)+B(s,u)B(s+\tfrac{1}{2},u)-(u+1)A(s+\tfrac{1}{2},u-1)C(s,u)=0, \tag{2.25}$$

$$[(u+2)(2u+1)]^{1/2}A(s,u)B(s+\tfrac{1}{2},u+1) = [u(2u+3)]^{1/2}A(s+\tfrac{1}{2},u)B(s,u), \tag{2.26}$$

$$[(u+1)(2u-1)]^{1/2}B(s,u)C(s+\tfrac{1}{2},u) = [(u-1)(2u+1)]^{1/2}B(s+\tfrac{1}{2},u-1)C(s,u), \tag{2.27}$$

$$A(s,u)D(s+\tfrac{1}{2},u+1) = D(s,u)A(s,u+1), \tag{2.28}$$

$$C(s,u)D^*(s+\tfrac{1}{2},u-2) = D^*(s,u-1)C(s,u-1), \tag{2.29}$$

$$2sA(s,u)C^*(s,u+2) - (2s+2)A(s-\tfrac{1}{2},u+1)C^*(s-\tfrac{1}{2},u+1) - 2sD(s,u)D^*(s,u+1) = 0, \tag{2.30}$$

$$(2s+1)A(s,u)C^*(s,u+2) - (2s+3)A(s-\tfrac{1}{2},u+1)C^*(s-\tfrac{1}{2},u+1) + (2s+3)D(s,u)D^*(s,u+1) = 0. \tag{2.31}$$

We now consider the second order SO(7) Casimir invariant I_2. From Eq. (5.1) of Vanden Berghe *et al.*[2] it follows that

$$\langle s\,t\,u\,\lambda\,\mu\,\nu|I_2|s\,t\,u\,\lambda\,\mu\,\nu\rangle = -2\sqrt{3}\langle s\,t\,u\,\lambda\,\mu\,\nu|(T^{[1/2\ 1/2\ 1]}T^{[1/2\ 1/2\ 1]})^{\begin{bmatrix}0&0&0\\0&0&0\end{bmatrix}}|s\,t\,u\,\lambda\,\mu\,\nu\rangle$$
$$-s(s+1)-t(t+1)-\tfrac{1}{2}u(u+1). \tag{2.32}$$

For $[v,0,1]$ representations the expectation value of I_2 is[6]

$$\langle s\,t\,u\,\lambda\,\mu\,\nu|I_2|s\,t\,u\,\lambda\,\mu\,\nu\rangle = -\tfrac{1}{8}(4v^2+24v+21). \tag{2.33}$$

We next apply the Wigner–Eckart theorem and Eq. (15.23) of De Shalit and Talmi[5] to the matrix element on the right-hand side of (2.32). After making the choice $t = s + \frac{1}{2}$ and taking into account (2.33), we obtain the following equation that can be added to the set (2.8)–(2.31):

$$|A(s,u)|^2 + |B(s,u)|^2 + |C(s,u)|^2$$
$$+ |A(s-\tfrac{1}{2},u-1)|^2 + |B(s-\tfrac{1}{2},u)|^2 + |C(s-\tfrac{1}{2},u+1)|^2 + |D(s,u)|^2 + |E(s,u)|^2 + |D(s,u-1)|^2$$
$$= (2s+1)(2s+2)(2u+1)[\tfrac{1}{8}(4v^2+24v+21) - 2s^2 - 3s - \tfrac{3}{4} - \tfrac{1}{2}u(u+1)]. \tag{2.34}$$

3. SOLUTION OF THE RECURSION RELATIONS

Equations (2.8)–(2.13) and (2.34) can be combined into

$$[(2s+3)(2u+3) + 2s + 1](2u+1)|A(s,u)|^2 + (2s+2)2u(2u+3)|B(s,u)|^2$$
$$+ (2s+3)(2u+1)(2u+2)|D(s,u)|^2 + (2s+3)(2u+1)(2u+2)|C(s-\tfrac{1}{2},u+1)|^2$$
$$= (2s+1)(2s+2)(2s+3)(2u+1)(u+1)(2u+3)[\tfrac{1}{8}(4v^2+24v+21) - 2s^2 - 4s - \tfrac{5}{4} - \tfrac{1}{2}u(u+5)]. \tag{3.1}$$

This equation can be applied to states with maximum s-value, i.e., $2s = 2s_M = v - u + \frac{1}{2}$. As a consequence of the definitions (2.2), (2.3), and (2.5), $A(s_M,u) = B(s_M,u) = D(s_M,u) = 0$. Hence

$$|C(s_M - \tfrac{1}{2},u+1)|^2 = \tfrac{1}{4}(2u+1)(2u+3)(v-u+\tfrac{1}{2})(v-u+\tfrac{3}{2})(v-u+\tfrac{5}{2}). \tag{3.2}$$

Elimination of $|B(s,u)|^2, |C(s,u)|^2, |B(s-\frac{1}{2},u)|^2, |D(s,u)|^2, |D(s,u-1)|^2$, and $|E(s,u)|^2$ from Eqs. (2.8)–(2.13) and (2.34) gives

$$(2s+1)(2s+3)u(2u+3)|A(s-\tfrac{1}{2},u-1)|^2 - 2s(2s+2)(u+1)(2u+1)|A(s,u)|^2 - (2s+1)(2s+3)(2u+2)|C(s-\tfrac{1}{2},u+1)|^2$$
$$= -s(2s+1)(2s+2)(2s+3)(2u+1)(2u+3)[\tfrac{1}{8}(4v^2+24v+21) - 2s^2 - 2s + \tfrac{3}{4} - 2su - \tfrac{3}{2}u(u+2)]. \tag{3.3}$$

Equations (2.30) and (2.31) can be combined into

$$|C(s-\tfrac{1}{2},u+1)|^2 = \left[\frac{2s(2s+2)}{(2s+1)(2s+3)}\right]^2 \frac{|A(s,u)|^2|C(s,u+2)|^2}{|A(s-\tfrac{1}{2},u+1)|^2}, \tag{3.4}$$

which together with (3.3) and the boundary condition (3.2) allows one to solve $|A(s,u)|^2$ and $|C(s,u)|^2$ for s- and u-values which are subject to the condition that $v - 2s - u + \frac{1}{2}$ be even. Therefore we will mark these solutions by a subscript E. On the other hand, for $v - 2s - u + \frac{1}{2}$ odd, an index O will be added. Omission of a subscript will refer to the independency upon a precise subscript. The solutions are

$$|A_E(s,u)|^2 = \frac{1}{16}(2s+1)(2s+3)\frac{(2u+1)(2u+3)}{u+1}\left(v+2s+u+\frac{13}{2}\right)\left(v-2s-u+\frac{1}{2}\right), \tag{3.5}$$

$$|C_E(s,u)|^2 = \frac{1}{16}(2s+1)(2s+3)\frac{(2u-1)(2u+1)}{u}\left(v+2s-u+\frac{9}{2}\right)\left(v-2s+u+\frac{1}{2}\right). \tag{3.6}$$

Obviously Eqs. (3.3) and (3.4) remain valid when $v - 2s - u + \frac{1}{2}$ is odd, but other boundary conditions on A and C must be invoked before solving them. To that aim we will proceed in an indirect way by first eliminating $|B(s,u)|^2$, $|C(s,u)|^2$, $|A(s-\frac{1}{2},u-1)|^2$, $|D(s,u-1)|^2$, and $|E(s,u)|^2$ from Eqs. (2.8)–(2.13) and (2.34), i.e.,

$$2s(u+1)(2u+1)[|A(s,u)|^2 - |D(s,u)|^2] + (2s+1)u(2u+3)|B(s-\tfrac{1}{2},u)|^2 + [s(2u+3)+s+1]|C(s-\tfrac{1}{2},u+1)|^2$$
$$= s(2s+1)(2s+2)(2u+1)(u+1)(2u+3)[\tfrac{1}{8}(4v^2+24v+21) - 2s^2 - 2s + \tfrac{1}{4} - \tfrac{1}{2}u(u+5)]. \tag{3.7}$$

Since $|A_E(s_M,u)|^2$ and $|C_E(s_M-\frac{1}{2},u+1)|^2$ are known and $|D_E(s_M,u)|^2=0$, Eq. (3.7) produces an expression for $|B_O(s_M-\frac{1}{2},u)|^2$. Furthermore, the elimination of $|A(s-\frac{1}{2},u-1)|^2$, $|B(s-\frac{1}{2},u)|^2$, $|C(s-\frac{1}{2},u+1)|^2$, $|D(s,u)|^2$, $|D(s,u-1)|^2$, and $|E(s,u)|^2$ from Eqs. (2.8), (2.9), and (2.34) leads to

$$2[|A(s,u)|^2+|B(s,u)|^2+|C(s,u)|^2]=(2s+1)(2s+3)(2u+1)[\tfrac{1}{8}(4v^2+24v+21)-2s^2-6s-\tfrac{9}{4}-\tfrac{1}{2}u(u+1)]. \quad (3.8)$$

From Eq. (3.8), an expression for $|C_O(s_M-\frac{1}{2},u)|^2$ follows because $|A_O(s_M-\frac{1}{2},u)|^2=0$, and $|B_O(s_M-\frac{1}{2},u)|^2$ is already known. By the aid of (3.3) and (3.4) it is then straightforward to derive

$$|A_O(s,u)|^2=\frac{1}{16}(2s+1)(2s+3)\frac{(2u+1)(2u+3)}{u+1}\times\left(v+2s+u+\frac{11}{2}\right)\left(v-2s-u-\frac{1}{2}\right), \quad (3.9)$$

$$|C_O(s,u)|^2=\frac{1}{16}(2s+1)(2s+3)\frac{(2u-1)(2u+1)}{u}\times\left(v+2s-u+\frac{11}{2}\right)\left(v-2s+u+\frac{3}{2}\right). \quad (3.10)$$

Having now the complete results for $|A(s,u)|^2$ and $|C(s,u)|^2$ at our disposition we learn from (3.8) that

$$|B_E(s,u)|^2=\frac{1}{16}(2s+1)(2s+3)\frac{2u+1}{u(u+1)}\times\left(v+2s-u+\frac{9}{2}\right)\left(v-2s-u+\frac{1}{2}\right), \quad (3.11)$$

$$|B_O(s,u)|^2=\frac{1}{16}(2s+1)(2s+3)\frac{2u+1}{u(u+1)}\times\left(v+2s+u+\frac{11}{2}\right)\left(v-2s+u+\frac{3}{2}\right). \quad (3.12)$$

Similarly (3.7) allows us to calculate

$$|D_E(s,u)|^2=\frac{1}{16}\frac{(2u+1)(2u+3)}{u+1}\times\left(v-2s+u+\frac{5}{2}\right)\left(v-2s-u+\frac{1}{2}\right), \quad (3.13)$$

$$|D_O(s,u)|^2=\frac{1}{16}\frac{(2u+1)(2u+3)}{u+1}\times\left(v+2s-u+\frac{7}{2}\right)\left(v+2s+u+\frac{11}{2}\right). \quad (3.14)$$

Finally, Eq. (2.8) leads to

$$|E_E(s,u)|^2=\frac{1}{16}\frac{2u+1}{u(u+1)}\left(v+2s+4su+3u+\frac{9}{2}\right)^2, \quad (3.15)$$

$$|E_O(s,u)|^2=\frac{1}{16}\frac{2u+1}{u(u+1)}\left(v-2s-4su-3u+\frac{3}{2}\right)^2. \quad (3.16)$$

Clearly, in order to derive general expressions for the squares of the absolute values of the reduced matrix elements, we have only used a restricted part out of the set of relations established in Sec. 2. The relations (2.14)–(2.31) provide us the means for fixing the relative phase factors of the reduced matrix elements themselves. In particular, if we choose

$$B_E(s,u)=e^{i\beta}|B_E(s,u)|,\quad \beta\in\mathbb{R},$$
$$D_E(s,u)=e^{i\delta}|D_E(s,u)|,\quad \delta\in\mathbb{R},$$

then it follows that

$$A(s,u)=e^{i(\beta+\delta+a\pi)}|A(s,u)|,\quad a\in\{0,1\},$$
$$B_O(s,u)=e^{i(\beta+b\pi)}|B_O(s,u)|,\quad b\in\{0,1\},$$
$$C(s,u)=e^{i[(\beta-\delta)+(a-b+1)\pi]}|C(s,u)|,$$
$$D_O(s,u)=e^{i(\delta+b\pi)}|D_O(s,u)|,$$
$$E(s,u)=e^{i(a-b)\pi}|E(s,u)|.$$

ACKNOWLEDGMENT

We like to thank Professor C. Grosjean for his interest during the course of this work.

[1] H. De Meyer, P. De Wilde, and G. Vanden Berghe, J. Phys. A: Math. Gen. **15**, 2665 (1982).
[2] G. Vanden Berghe, H. De Meyer, and P. De Wilde, J. Phys. A: Math. Gen. **15**, 2677 (1982).
[3] N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968).
[4] W. G. McKay and J. Patera, *Table of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1981).
[5] A. De Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[6] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).

Multispinor basis for representations of SO(N)

R. S. Nikam and C. R. Sarma
Department of Physics, Indian Institute of Technology, Bombay-400 076, India

(Received 20 July 1983; accepted for publication 2 September 1983)

The generators of the rotation groups SO(N) ($N = 2n$, $2n + 1$) have been realized using a restriction of the unitary group U(2^n) defined on the 2^n-dimensional fundamental representation space of spinors. These generators have been used to subduce multispinor representations of SO(N) from those of U(2^n). The procedure has been illustrated for the two-spinor vector representations $\langle 10\rangle$ and $\langle 1000\rangle$ of SO(5) and SO(8), respectively.

PACS numbers: 02.20. + b

1. INTRODUCTION

The tensor (integral) representations of the rotation groups SO(N) ($N = 2n$, $2n + 1$) have been extensively studied over the past few years. Gel'fand and Zetlin[1] were the first to display explicitly the canonical basis states spanning the irreducible representations (IR's) $\langle\lambda\rangle \equiv \langle\lambda_1\lambda_2\cdots\lambda_n\rangle$ of SO(N). These basis states were studied in detail by Pang and Hecht[2] and Wong[3] using skew-Hermitian generators J_{ij} of the Lie algebra of SO(N). An alternative approach based on the unitary group restriction U(N)|SO(N) was considered by Chacon[4] and Moshinsky[5] and was found also to lead to the canonical basis set. Though these procedures are not direct enough to be programmed readily, they are, nonetheless, well understood and lead to the required IR's of SO(N).

The spinor representations of SO(N) are, on the other hand, relatively less well studied, though they have applications in a number of areas of nuclear and particle physics. Examples of such applications are the multispinor representations of SO(5),[6,7] SO(6),[8] SO(8),[9] and SO(10).[10] Spinor representations of SO(N) with $N \leqslant 6$ can be studied readily using isomorphisms with unitary and symplectic groups (cf. Ref. 10, p. 360). The procedures for generating the spinor representations become relatively more complicated for $N > 6$. It is, therefore, worth investigating whether a direct scheme could be worked out for generating the multispinor states of SO(N).

In the present note we have tried to generalize the U(N) $\downarrow$SO(N) restriction used for the tensor representations[4,5] to obtain the spinor representations of SO(N). Since the basic spinor space of both SO($2n + 1$) and SO($2n$) is of 2^n dimensions, this required the restriction U(2^n)$\downarrow$SO(N). As such, more complicated relations are expected between the generators of SO(N) and the shift operators of U(2^n) than in the tensor case. We have guessed at the form of these relationships using as guidelines the Lie algebra of SO(N). These generators have been defined in Sec. 2 and used to obtain the basis states of the $\langle 10\rangle$ representations of SO(7) and SO(8). A brief discussion has been presented in Sec. 3.

2. SPINOR BASIS FOR SO(N)

As pointed out in Sec. 1, the present procedure is a generalization of the U(N)$\downarrow$SO(n) restriction. In view of this, we summarize briefly this technique for obtaining the tensor representations before considering the spinor representations.

Given an ordered orthonormal set of single particle basis states,

$$\{V_N : \phi_i \mid i = 1,\ldots,N\}, \tag{1}$$

spanning the fundamental representation space V_N of the unitary group U(N), it is possible to define a set of rth rank tensors spanning the reducible space $V_N \otimes^r$ of the group. A reduction of this space yields a set of irreducible subspaces $V_{[\lambda]}$ of U(N), each of which is stable under the generators E_{ij} ($i, j = 1,\ldots,N$) of the group. These generators define the Lie algebra of U(N),

$$[E_{ij}, E_{km}] = E_{im}\delta_{jk} - E_{kj}\delta_{im} \tag{2}$$

and

$$E_{ij}^{+} = E_{ji}. \tag{3}$$

The generators Λ_{pq} of SO(N) can be defined using the E_{ij} as [cf. Ref. 5, p. 36, Eq. (5.18)]

$$\Lambda_{pq} = E_{pq} - E_{N+1-q,N+1-p}. \tag{4}$$

Such a realization implies that

$$\Lambda_{pq}^{+} = \Lambda_{qp} \tag{5}$$

and

$$\Lambda_{pq} = -\Lambda_{N+1-q,N+1-p}. \tag{6}$$

Thus there exist only $n(2n + 1)$ independent generators for SO($2n + 1$) and $n(2n - 1)$ for SO($2n$). Using Eq. (2) it can be readily shown that the Λ_{pq} define a Lie algebra of SO(N) as

$$[\Lambda_{pq}, \Lambda_{rs}] = \Lambda_{ps}\delta_{rq} - \Lambda_{rq}\delta_{ps} - \Lambda_{N+1-q\,s}\delta_{N+1-p\,r} + \Lambda_{r\,N+1-p}\delta_{N+1-qs}. \tag{7}$$

The weights of the basis states spanning an irrep $\langle\lambda\rangle \equiv \langle\lambda_1\lambda_2\cdots\lambda_n\rangle$ are defined as the eigenvalues of the commuting set of generators,

$$\Lambda_{pp} = E_{pp} - E_{N+1-p\,N+1-p} \quad (p = 1,2,\ldots,N). \tag{8}$$

The weights are assumed to be ordered such that $(w) \equiv (w_1', w_2',\ldots,w_n') > (w') \equiv (w_1', w_2',\ldots,w_n')$ if the first nonzero difference $w_1 - w_1', w_2 - w_2',\ldots,w_n - w_n'$ is positive. On the basis of this, the generators Λ_{pq} ($p \neq q$) can be classified as weight raising ($p < q$) and weight lowering ($p > q$) ones.

The highest weight state (HWS) of any irrep $\langle\lambda\rangle$ of SO(N) follows readily using these generators. The other low-

 0022-2488/84/051199-06$02.50

er weight states can be generated from this using lowering operator polynomials of the generators [cf. Ref. 4, p. 45, for SO(5) and p. 46 for SO(6)]. Thus a direct procedure for generating the tensor representations results from the defining relation in Eq. (4). This has prompted us to examine whether one or more relations exist between the generators of U(2^n).

Consider the 2^n-dimensional space V_{2^n} spanned by an ordered orthonormal set of spinors

$$V_{2^n}: \quad \{|u_k\rangle \equiv |(m_{nk})\rangle \,|k = 1,2,...,2^n \\ |m_{ik}| = \tfrac{1}{2} \text{ for all } i = 1,2,...,n\}. \tag{9}$$

An ordering is defined over the basis states u_k by assuming that $j < k$ implies that the first nonzero difference $m_{1j} - m_{1k}, m_{2j} - m_{2k}, ..., m_{nj} - m_{nk}$ is positive. This space is the fundamental representation space of U(2^n) on which a set of generators,

$$\{E_{jk} \equiv E_{(m_{1j}m_{2j}\cdots m_{nj})(m_{1k}m_{2k}\cdots m_{nk})} \,|\, j,k = 1,...,2^n\}, \tag{10}$$

satisfying Eqs. (2) and (3) can be defined. For $N = 2n + 1$, this space V_{2^n} is also the carrier space for the fundamental representation $\langle \frac{1}{2}\frac{1}{2}\cdots\frac{1}{2}\rangle$ of SO($2n + 1$). If $N = 2n$, on the other hand, V_{2^n} breaks up into two subspaces defining the nonequivalent IR's $\langle \frac{1}{2}\frac{1}{2}\cdots\frac{1}{2}\rangle$ and $\langle \frac{1}{2}\frac{1}{2}\cdots -\frac{1}{2}\rangle$ of SO($2n$). In view of this, it is convenient to consider the two groups separately.

A. SO($2n$)

Generalizing the SO(5) study by Ichimura,[7] we define the set of n commuting generators Λ_{pp} of both SO($2n$) and SO($2n + 1$) as

$$\Lambda_{pp} = \sum_{k=1}^{2^n} m_{pk} E_{kk}. \tag{11}$$

We can readily establish the correspondence between the choice of Eq. (11) and that of Ref. 7 by using the basis set defined as in Eq. (9):

$$u_1 = |(+\ +)\rangle, \quad u_3 = |(-\ +)\rangle, \\ u_2 = |(+\ -)\rangle, \quad u_4 = |(-\ -)\rangle, \tag{12}$$

where we have replaced the weights $\pm\frac{1}{2}$ by their signs. Using this set in Eq. (11), we obtain

$$\Lambda_{11} = \tfrac{1}{2}E_{11} + \tfrac{1}{2}E_{22} - \tfrac{1}{2}E_{33} - \tfrac{1}{2}E_{44} \\ = \tfrac{1}{2}(H_1 + H_2) \tag{13}$$

and

$$\Lambda_{22} = \tfrac{1}{2}E_{11} - \tfrac{1}{2}E_{22} + \tfrac{1}{2}E_{33} - \tfrac{1}{2}E_{44} \\ = \tfrac{1}{2}(H_1 - H_2), \tag{14}$$

where the H_i ($i = 1, 2$) are in the notation of Eq. (4.3), Ref. 7.

The major problem arises in defining the shift generators Λ_{pq} ($p \neq q$). For SO($2n$) there are $2n(n-1)$ generators of this type, of which half are weight raising and the remaining half weight lowering ones. Representing the weight raising generators as Λ_{pq} ($p < q$), we require that the weight lowering generators are Λ^{+}_{pq} and can be represented as Λ_{qp} ($p < q$). Any definition of the generators Λ_{pq} ($p < q$) should then be consistent with this. Further, among these weight raising generators $n(n-1)/2$ are of the form Λ_{pq} ($1 \leqslant p \leqslant q \leqslant n$) and the remaining $n(n-1)/2$ are Λ_{pq} ($n + 1 \leqslant p < q \leqslant 2n$). Thus essentially we need only these two categories on each of which we impose the condition

$$\Lambda_{pq} = -\Lambda_{N+1-q\,N+1-p}. \tag{15}$$

Subject to these conditions, we can define Λ_{pq} as

$$1 \leqslant p < q \leqslant n: \quad \Lambda_{pq} = \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} \\ \times E_{(k:\, m_{pk} = +,\, m_{qk} = -)(k:\, m_{pk} = -,\, m_{qk} = +)}, \tag{16}$$

where

$$E_{(k:\, m_{pk} = +,\, m_{qk} = -)(k:\, m_{pk} = -,\, m_{qp} = +)} \\ = E_{(m_{1k}m_{2k}\cdots m_{p-1\,k} m_{pk}(=+1/2) m_{p+1\,k}\cdots m_{q-1\,k} m_{qk}(=-1/2) m_{q+1\,k}\cdots m_{nk})(m_{1k}\cdots m_{p-1\,k} m_{pk}(=-1/2) m_{p+1\,k}\cdots m_{q-1\,k} m_{qk}(=+1/2) m_{q+1\,k}\cdots m_{nk})}, \tag{17}$$

and the summation is over all the 2^{n-2} states arising from all possible weights $m_{ik} = \pm\frac{1}{2}(i \neq p, q)$. Using the ordering of the basis states as in Eq. (9), we note that the generator of U(2^n) defined in Eq. (17) is a weight raising one. Thus Λ_{pq} ($p < q$) of Eq. (16), which is a linear combination of such generators, is also a weight raising one. Since, from Eq. (3), we have

$$E^{+}_{(k:\, m_{pk} = +,\, m_{qk} = -)(k:\, m_{pk} = -,\, m_{pk} = +)} = E_{(k:\, m_{pk} = -,\, m_{qk} = +)(k:\, m_{pk} = +,\, m_{qk} = -)}, \tag{18}$$

and the right side of Eq. (18) defines a weight lowering generator of U(2^n), we find that

$$1 \leqslant p < q \leqslant n: \quad \Lambda_{qp} = \Lambda^{+}_{pq} \tag{19}$$

is a weight lowering generator. As pointed out earlier, we need to define one more class of weight raising generators, $\Lambda_{p\,N+1-q}$ ($1 \leqslant p < q \leqslant n$), which we assume to be subject to

$$\Lambda_{q\,N+1-p} = -\Lambda_{p\,N+1-q}. \tag{20}$$

We define these generators as

$$1 \leqslant p < q \leqslant n: \quad \Lambda_{p\,N+1-q} = \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} E_{(k:\, m_{pk} = +,\, m_{qk} = +)(k:\, m_{pk} = -,\, m_{qk} = -)}, \tag{21}$$

where, as in Eq. (17),

$$E_{(k:\, m_{pk} = +,\, m_{qk} = +)(k:\, m_{pk} = -,\, m_{qk} = -)} \\ = E_{(m_{1k}m_{2k}\cdots m_{p-1\,k} m_{pk}(=1/2) m_{p+1\,k}\cdots m_{q-1\,k} m_{qk}(=1/2) m_{q+1\,k}\cdots m_{nk})(m_{1k}m_{2k}\cdots m_{p-1\,k} m_{pk}(=-1/2) m_{p+1\,k}\cdots m_{q-1\,k} m_{qk}(=-1/2) m_{q+1\,k}\cdots m_{nk})}. \tag{22}$$

Again since Eq. (22) defines a weight raising generator of U(2^n), the Λ_{pN+1-q} defined by Eq. (21) is also a weight raising generator. Further,

$$1 \leqslant p < q \leqslant n+1: \quad \Lambda^{+}_{pN+1-q} = \Lambda_{N+1-qp} \tag{23}$$

is as before a weight lowering generator. That the generators defined by Eqs. (16) and (21) define a Lie algebra can be verified using Eq. (2). This has been done in the Appendix, subsection 1. As an illustration of the form of these generators, we consider the group SO(8). The 16 basis states of the fundamental spinor irrep of SO(8) are

$$\begin{aligned}
u_1 &= |(+\ +\ +\ +)\rangle, & u_5 &= |(+\ -\ +\ +)\rangle,\\
u_2 &= |(+\ +\ +\ -)\rangle, & u_6 &= |(+\ -\ +\ -)\rangle,\\
u_3 &= |(+\ +\ -\ +)\rangle, & u_7 &= |(+\ -\ -\ +)\rangle,\\
u_4 &= |(+\ +\ -\ -)\rangle, & u_8 &= |(+\ -\ -\ -)\rangle,\\
u_9 &= |(-\ +\ +\ +)\rangle, & u_{13} &= |(-\ -\ +\ +)\rangle,\\
u_{10} &= |(-\ +\ +\ -)\rangle, & u_{14} &= |(-\ -\ +\ -)\rangle,\\
u_{11} &= |(-\ +\ -\ +)\rangle, & u_{15} &= |(-\ -\ -\ +)\rangle,\\
u_{12} &= |(-\ +\ -\ -)\rangle, & u_{16} &= |(-\ -\ -\ -)\rangle.
\end{aligned} \tag{24}$$

Of these $u_1, u_4, u_6, u_7, u_{10}, u_{11}, u_{13}, u_{16}$ span the IR $\langle \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2} \rangle$ and the rest $\langle \frac{1}{2}\frac{1}{2}\frac{1}{2} -\frac{1}{2} \rangle$. Using Eq. (11), we obtain the four weight generators as

$$\begin{aligned}
\Lambda_{11} = \tfrac{1}{2}[&E_{11}+E_{22}+E_{33}+E_{44}+E_{55}+E_{66}+E_{77}+E_{88}\\
&-E_{99}-E_{1010}-E_{1111}-E_{1212}-E_{1313}-E_{1414}\\
&-E_{1515}-E_{1616}],
\end{aligned} \tag{25a}$$

$$\begin{aligned}
\Lambda_{22} = \tfrac{1}{2}[&E_{11}+E_{22}+E_{33}+E_{44}-E_{55}-E_{66}-E_{77}-E_{88}\\
&+E_{99}+E_{1010}+E_{1111}+E_{1212}-E_{1312}-E_{1414}\\
&-E_{1515}-E_{1616}],
\end{aligned} \tag{25b}$$

$$\begin{aligned}
\Lambda_{33} = \tfrac{1}{2}[&E_{11}+E_{22}-E_{33}-E_{44}+E_{55}+E_{66}-E_{77}-E_{88}\\
&+E_{99}+E_{1010}-E_{1111}-E_{1212}+E_{1313}+E_{1414}\\
&-E_{1515}-E_{1616}],
\end{aligned} \tag{25c}$$

$$\begin{aligned}
\Lambda_{44} = \tfrac{1}{2}[&E_{11}-E_{22}+E_{33}-E_{44}+E_{55}-E_{66}+E_{77}-E_{88}\\
&+E_{99}-E_{1010}+E_{1111}-E_{1212}+E_{1313}-E_{1414}\\
&+E_{1515}-E_{1616}].
\end{aligned} \tag{25d}$$

Using Eq. (16), we now obtain the six generators of this category as

$$\begin{aligned}
\Lambda_{12} &= \sum_{k=1}^{4} E_{(1/2\ -1/2\ m_{3k}m_{4k})(-1/2\ 1/2\ m_{3k}m_{4k})}\\
&= E_{(+\,-\,+\,+)(-\,+\,+\,+)} + E_{(+\,-\,+\,-)(-\,+\,+\,-)}\\
&\quad + E_{(+\,-\,-\,+)(-\,+\,-\,+)} + E_{(+\,-\,-\,-)(-\,+\,-\,-)}\\
&= E_{59} + E_{610} + E_{711} + E_{812},
\end{aligned} \tag{26a}$$

$$\begin{aligned}
\Lambda_{13} &= \sum_{k=1}^{4} (-)^{m_{2k}+1/2} E_{(1/2\ m_{2k}\ -1/2\ m_{4k})(-1/2\ m_{2k}\ 1/2\ m_{4k})}\\
&= -E_{(+\,+\,-\,+)(-\,+\,+\,+)} - E_{(+\,+\,-\,-)(-\,+\,+\,-)}\\
&\quad + E_{(+\,-\,-\,+)(-\,-\,+\,+)} + E_{(+\,-\,-\,-)(-\,-\,+\,-)}\\
&= -E_{39} - E_{410} + E_{713} + E_{814},
\end{aligned} \tag{26b}$$

$$\begin{aligned}
\Lambda_{14} &= \sum_{k=1}^{4} (-)^{m_{2k}+m_{3k}+1} E_{(1/2\ m_{2k}m_{3k}\ -1/2)(-1/2\ m_{2k}m_{3k}1/2)}\\
&= E_{(+\,+\,+\,-)(-\,+\,+\,+)} - E_{(+\,+\,-\,-)(-\,+\,+\,-)}\\
&\quad - E_{(+\,-\,+\,-)(-\,-\,+\,+)} + E_{(+\,-\,-\,-)(-\,-\,-\,+)}\\
&= E_{29} - E_{411} - E_{613} + E_{815},
\end{aligned} \tag{26c}$$

$$\begin{aligned}
\Lambda_{23} &= \sum_{k=1}^{4} E_{(m_{1k}\,1/2\,-\,1/2\,m_{4k})(m_{1k}\,-\,1/2\,-\,1/2\,m_{4k})}\\
&= E_{(+\,+\,-\,+)(+\,-\,+\,+)} + E_{(+\,+\,-\,-)(+\,-\,+\,-)}\\
&\quad + E_{(-\,+\,-\,+)(-\,-\,+\,+)} + E_{(-\,+\,-\,-)(-\,-\,+\,-)}\\
&= E_{35} + E_{46} + E_{1113} + E_{1214},
\end{aligned} \tag{26d}$$

$$\begin{aligned}
\Lambda_{24} &= \sum_{k=1}^{4} (-)^{m_{3k}+1/2} E_{(m_{1k}\,1/2\,m_{3k}\,-1/2)(m_{1k}\,-1/2\,m_{3k}\,1/2)}\\
&= -E_{25} + E_{47} - E_{1013} + E_{1215},
\end{aligned} \tag{26e}$$

$$\begin{aligned}
\Lambda_{34} &= \sum_{k=1}^{4} E_{(m_{1k}\,m_{2k}\,1/2\,-1/2)(m_{1k}m_{2k}\,-1/2\,1/2)}\\
&= E_{23} + E_{67} + E_{1011} + E_{1415}.
\end{aligned} \tag{26f}$$

Similarly, using Eq. (21), we obtain

$$\begin{aligned}
\Lambda_{15} &= \sum_{k=1}^{4} (-)^{m_{2k}+m_{3k}+1} E_{(1/2\ m_{2k}\ m_{3k}1/2)(-1/2\ m_{2k}m_{3k}\ -1/2)}\\
&= E_{110} - E_{312} - E_{514} + E_{716},
\end{aligned} \tag{27a}$$

$$\begin{aligned}
\Lambda_{16} &= \sum_{k=1}^{4} (-)^{m_{2k}+1/2} E_{(1/2\ m_{2k}\ 1/2\ m_{4k})(-1/2\ m_{2k}\ -1/2\ m_{4k})}\\
&= -E_{111} - E_{212} + E_{515} + E_{616},
\end{aligned} \tag{27b}$$

$$\begin{aligned}
\Lambda_{17} &= \sum_{k=1}^{4} E_{(1/2\ 1/2\ m_{3k}\ m_{4k})(-1/2\ -1/2\ m_{3k}m_{4k})}\\
&= E_{113} + E_{214} + E_{315} + E_{416},
\end{aligned} \tag{27c}$$

$$\begin{aligned}
\Lambda_{25} &= \sum_{k}^{4} (-)^{m_{3k}+1/2} E_{(m_{1k}\,1/2\,m_{3k}\,1/2)(m_{1k}\,-1/2\,m_{2k}\,-1/2)}\\
&= -E_{16} + E_{38} - E_{914} + E_{1116},
\end{aligned} \tag{27d}$$

$$\begin{aligned}
\Lambda_{26} &= \sum_{k=1}^{4} E_{(m_{1k}\,1/2\,1/2\,m_{4k})(m_{1k}\,-1/2\ \ -1/2\,m_{4k})}\\
&= E_{17} + E_{28} + E_{915} + E_{1016},
\end{aligned} \tag{27e}$$

$$\begin{aligned}
\Lambda_{35} &= \sum_{k=1}^{4} E_{(m_{1k}\,m_{2k}\,1/2\,1/2)(m_{1k}m_{2k}\,-1/2\ \ -1/2)}\\
&= E_{14} + E_{58} + E_{912} + E_{1316}.
\end{aligned} \tag{27f}$$

In addition to checking the commutation relations of the generators in Appendix subsection 1, we can readily verify that the generators of Eqs. (25), (26), and (27) satisfy Eq. (7) on using Eq. (2).

B. SO(2n + 1)

The weight generators of SO(2n + 1) are the same as those defined by Eq. (11). The generators defined using Eqs. (16)–(22) also hold for this group on replacing N by $2n+1$ and range of values $n+1 \leqslant p < q \leqslant 2n$ by $n+2 \leqslant p < q \leqslant 2n+1$. This just leaves one more set of n weight raising generators of the form $\Lambda_{p\,n+1}$ $(1 \leqslant p \leqslant n)$ to be defined subject to

$$\Lambda_{p\,n+1} = -\Lambda_{n+1\,N+1-p}. \tag{28}$$

These generators are now defined as

$$1\leqslant p\leqslant n:\quad \Lambda_{p\,n+1}=(2)^{-1/2}\sum_{k=1}^{2^{n-1}}(-)^{\sum_{a=p+1}^{n}(m_{ak}+1/2)} \times E_{(k:\,m_{pk}=1/2)(k:\,m_{pk}=-1/2)}, \tag{29}$$

where

$$E_{(k:\,m_{pk}=1/2)(k:\,m_{pk}=-1/2)} = E_{(m_{1k}m_{2k}\cdots m_{p-1\,k}m_{pk}(=1/2)m_{p+1\,k}\cdots m_{nk})(m_{1k}m_{2k}\cdots m_{pk}(=-1/2)\cdots m_{nk})}. \tag{30}$$

We also use

$$\Lambda^{\dagger}_{p\,n+1}=\Lambda_{n+1\,p} \tag{31}$$

as the corresponding weight lowering operator. This result follows as before from Eqs. (29) and (3) for the ordering as in Eq. (9).

Using the basis set defined by Eq. (12), we now obtain as an illustration, the $n(n+1)/2=3$ shift generators of SO(5). Equation (16) leads, as for SO(8), to

$$\Lambda_{12}=E_{(+\,-)(-\,+)}=E_{23}. \tag{32a}$$

Similarly, Eq. (21) leads to

$$\Lambda_{14}=E_{(+\,+)(-\,-)}=E_{14}. \tag{32b}$$

Finally, Eq. (28) leads to $n=2$ generators,

$$\Lambda_{13}=(2)^{-1/2}\sum_{k=1}^{2}(-)^{(m_{2k}+1/2)}E_{(1/2\,m_{2k})(-1/2\,m_{2k})} =(2)^{-1/2}[-E_{(+\,+)(-\,+)}+E_{(+\,-)(-\,-)}] =(2)^{-1/2}(-E_{13}+E_{24}), \tag{32c}$$

$$\Lambda_{23}=(2)^{-1/2}\sum_{k=1}^{2}E_{(m_{1k}\,1/2)(m_{1k}\,-1/2)} =(2)^{-1/2}[E_{(+\,+)(+\,-)}+E_{(-\,+)(-\,-)}] =(2)^{-1/2}(E_{12}+E_{34}). \tag{32d}$$

The results of Eq. (32) again compare with those of earlier studies[7] except for relative sign changes and multiplicative factors which at best affect the overall phases and normalization of the basis states generated.

An illustration of the use of the generators defined by Eqs. (11), (16), (21), and (29) is now provided for the vector (two-spinor) representations $\langle 10\rangle$ and $\langle 1000\rangle$ of SO(5) and SO(8) respectively. The representation $\langle 10\rangle$ of SO(5) is subduced from the one defined by the Weyl shape $\square$ of U(4). The highest weight state of $\langle 10\rangle$ of SO(5) corresponds to the weights $\langle 10\rangle$ of Λ_{11}, Λ_{22} defined in Eqs. (13) and (14). This state is unique,

$$|\langle 10\rangle(10)\rangle=\boxed{\begin{matrix}u_1\\u_2\end{matrix}}\equiv\boxed{\begin{matrix}1\\2\end{matrix}}, \tag{33a}$$

where the basis states are as defined by Eq. (12). Applying $\Lambda_{21}=\Lambda^{\dagger}_{12}$ to the above, we obtain

$$|\langle 10\rangle(01)\rangle=\boxed{\begin{matrix}1\\3\end{matrix}}. \tag{33b}$$

Applying $\Lambda_{31}=\Lambda^{\dagger}_{13}$ to the state in Eq. (31a), we get

$$|\langle 10\rangle(00)\rangle=(2)^{-1/2}\left(-\boxed{\begin{matrix}3\\2\end{matrix}}+\boxed{\begin{matrix}1\\4\end{matrix}}\right) =(2)^{-1/2}\left(\boxed{\begin{matrix}2\\3\end{matrix}}+\boxed{\begin{matrix}1\\4\end{matrix}}\right). \tag{33c}$$

Applying $\Lambda_{41}=\Lambda^{\dagger}_{14}$ to Eq. (31a), we have

$$|\langle 10\rangle(0\,-1)\rangle=\boxed{\begin{matrix}4\\2\end{matrix}}=-\boxed{\begin{matrix}2\\4\end{matrix}}. \tag{33d}$$

Finally, applying $\Lambda_{54}=-\Lambda_{21}=-\Lambda^{\dagger}_{12}$ to Eq. (33d) yields

$$|\langle 10\rangle(-10)\rangle=\Lambda_{54}|\langle 10\rangle(0\,-1)\rangle =\Lambda_{21}\boxed{\begin{matrix}2\\4\end{matrix}}=\boxed{\begin{matrix}3\\4\end{matrix}}. \tag{33e}$$

The basis states defined in Eq. (24) and the generators of Eqs. (25)–(27) can similarly be used to obtain the eight basis states spanning the vector representation $\langle 1000\rangle$ of SO(8) subduced from $\square\square$ of U(16). The state with weights (1000) follows from Eqs. (24) and (25) as

$$|\langle 1000\rangle(1000)\rangle=a\boxed{18}+b\boxed{27}+c\boxed{36}+d\boxed{45},$$

where the unknowns are determined using the condition that this is an HWS. Using the weight raising generators of Eqs. (26) and (27) and equating the results to zero, we readily obtain

$$-d=c=-b=a,$$

where a can only be fixed by normalization using an arbitrary phase $+1$ leading to

$$|\langle 1000\rangle(1000)\rangle=\tfrac{1}{2}[\boxed{18}-\boxed{27}+\boxed{36}-\boxed{45}]. \tag{34a}$$

Applying the weight lowering generators as in the SO(5) case, we readily obtain

$$|\langle 1000\rangle(0100)\rangle=\tfrac{1}{2}[\boxed{112}-\boxed{211}+\boxed{310}-\boxed{49}], \tag{34b}$$

$$|\langle 1000\rangle(0010)\rangle=\tfrac{1}{2}[\boxed{114}-\boxed{213}+\boxed{510}-\boxed{69}], \tag{34c}$$

$$|\langle 1000\rangle(0001)\rangle=\tfrac{1}{2}[\boxed{115}-\boxed{313}+\boxed{511}-\boxed{79}], \tag{34d}$$

$$|\langle 1000\rangle(000\,-1)\rangle=-\tfrac{1}{2}[\boxed{216}-\boxed{414}+\boxed{612}-\boxed{810}], \tag{34e}$$

$$|\langle 1000\rangle(00\,-1\,0)\rangle=\tfrac{1}{2}[\boxed{316}-\boxed{415}+\boxed{712}-\boxed{811}], \tag{34f}$$

$$|\langle 1000\rangle(0\,-1\,0\,0)\rangle=-\tfrac{1}{2}[\boxed{516}-\boxed{611}+\boxed{714}-\boxed{813}], \tag{34g}$$

$$|\langle 1000\rangle(-1\,0\,0\,0)\rangle=\tfrac{1}{2}[\boxed{916}-\boxed{1015}+\boxed{1114}-\boxed{1213}], \tag{34h}$$

where Eqs. (34g) and (34h) follow readily on using $\Lambda_{76}=-\Lambda_{32}=-\Lambda^{\dagger}_{23}$ and $\Lambda_{87}=-\Lambda_{21}=-\Lambda^{\dagger}_{12}$ on $|\langle 1000\rangle(00\,-1\,0)\rangle$ and $|\langle 1000\rangle(0\,-1\,0\,0)\rangle$, respectively, with Eqs. (34b–f) being obtained from the HWS using $\Lambda_{i1}=\Lambda^{\dagger}_{1i}$ $(i=2,3,4,5,6)$. It is to be noted that these basis states occur as a reduction of $\langle \frac{1}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}\rangle\times\langle \frac{1}{2}\frac{1}{2}\frac{1}{2}\,-\frac{1}{2}\rangle$ of SO(8), as expected.

3. DISCUSSION

Basically there are four defining relations [cf. Eqs. (11), (16), (21) and (29)] between the generators of SO(N) and those of U(2^n) defined by us. These relations have been presented in a closed algebraic form so that the required Λ_{pq} of any SO($2n$) or SO($2n+1$) can be immediately written down. The ease with which they generate the multispinor basis, as illustrated for SO(8) and SO(5), means that they can be readily

used for any application of the rotation groups to physical problems. Such applications are further facilitated by the fact that programs exist for determining the matrix elements of generators of unitary groups over the basis spanning any IR.[11] In spite of these obvious advantages there are some drawbacks which need to be emphasized. First, the phase factor associated with the HWS of any IR has to be arbitrarily assigned since the basis states are realized only to within an overall normalization factor. This choice is, however, consistent within the set of states spanning the given representation. Second, there is a proliferation of definitions for the generators unlike the one, Eq. (4), used for the tensor basis. This is related to the much larger restriction $U(2^n) \downarrow SO(N)$ studied in the present note in comparison to $U(N) \downarrow SO(N)$ used for the tensor basis. Third, the basis generated is not the canonical Gel'fand–Zetlin[1] half-integral basis, though it is equivalent to it. This aspect will be considered in a forthcoming note. Finally, though some of the definitions such as that of Eq. (20) could be eliminated by redefining Eq. (16), etc., we have not done so and sacrificed elegance for the sake of clarity and utility.

APPENDIX

In this appendix some typical commutators of operators defined by Eqs. (11), (16), (21), and (28) have been obtained and found to satisfy Eq. (7) for the generators of $SO(N)$.

1. $[\Lambda_{pp}, \Lambda_{rs}]$ for $1 \leqslant p < r < s \leqslant n$

The definitions of Eqs. (11) and (16) lead to

$$
\begin{aligned}
[\Lambda_{pp}, \Lambda_{rs}] &= \sum_{j=1}^{2^n} \sum_{k=1}^{2^{n-2}} m_{pj} (-)^{\sum_{a=r+1}^{s-1} (m_{ak} + 1/2)} \left[E_{jj}, E_{(k:\, m_{rk} = 1/2,\, m_{sk} = -1/2)(k:\, m_{rk} = -1/2,\, m_{sk} = 1/2)} \right] \\
&= \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=r+1}^{s-1} (m_{ak} + 1/2)} \sum_{j=1}^{2^n} m_{pj} \delta(m_{rj}, m_{rk} = \tfrac{1}{2}) \{ \delta(m_{sj}, m_{sk} = -\tfrac{1}{2}) E_{(\cdots m_{rj} \cdots m_{sj} \cdots)(\cdots m_{rk} = -1/2 \cdots m_{sk} = 1/2 \cdots)} \\
&\quad - \delta(m_{rj}, m_{rk} = -\tfrac{1}{2}) \delta(m_{sj}, m_{sk} = \tfrac{1}{2}) E_{(\cdots m_{rk} = 1/2 \cdots m_{sk} = -1/2 \cdots)(\cdots m_{rj} \cdots, m_{sj} \cdots)} \} \prod_{\substack{i=1 \\ (i \neq r,s)}}^{n} \delta(m_{ij}, m_{ik}) \\
&= \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=r+1}^{s-1} (m_{ak} + 1/2)} m_{pk} \{ E_{(\cdots m_{rk} = 1/2 \cdots m_{sk} = -1/2 \cdots)(\cdots m_{rk} = -1/2 \cdots m_{sk} = -1/2 \cdots)} \\
&\quad - E_{(\cdots m_{rk} = 1/2 \cdots m_{sk} = -1/2 \cdots)(\cdots m_{rk} = -1/2 \cdots m_{sk} = -1/2 \cdots)} \} = 0,
\end{aligned}
\tag{A1}
$$

where the second equality follows from the explicit display of the δ functions of Eq. (2) in the present case.

2. $[\Lambda_{pp}, \Lambda_{pq}]$ for $1 \leqslant p < q \leqslant n$

Proceeding as in subsection 1 but putting $p = r$ $s = q$, we obtain the result

$$
\begin{aligned}
[\Lambda_{pp}, \Lambda_{pq}] &= \sum_{j=1}^{2^n} \sum_{k=1}^{2^{n-2}} m_{pj} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} \left[E_{jj}, E_{(k:\, m_{pk} = 1/2,\, m_{qk} = -1/2)(k:\, m_{pk} = -1/2,\, m_{qk} = -1/2)} \right] \\
&= \sum_{j=1}^{2^n} \sum_{k=1}^{2^{n-2}} m_{pj} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} \{ \delta(m_{pj}, m_{pk} = \tfrac{1}{2}) \delta(m_{qj}, m_{qk} = -\tfrac{1}{2}) E_{(\cdots m_{pj} \cdots m_{qj} \cdots)(\cdots m_{pk} = -1/2 \cdots m_{qk} = -1/2 \cdots)} \\
&\quad - \delta(m_{pj}, m_{pk} = -\tfrac{1}{2}) \delta(m_{qj}, m_{qk} = \tfrac{1}{2}) E_{(\cdots m_{pk} = 1/2 \cdots m_{qk} = -1/2 \cdots)(\cdots m_{pj} \cdots m_{qj} \cdots)} \} \prod_{\substack{i=1 \\ (i = \neq p,q)}}^{n} {}'' \delta(m_{ij}, m_{ik}) \\
&= \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} \left\{ \left(\frac{1}{2} \right) E_{(\cdots m_{pk} = 1/2 \cdots m_{qk} = -1/2 \cdots)(\cdots m_{pk} = -1/2 \cdots m_{qk} = 1/2 \cdots)} \right. \\
&\quad \left. - \left(-\frac{1}{2} \right) E_{(\cdots m_{pk} = 1/2 \cdots m_{qk} = -1/2 \cdots)(\cdots m_{pk} = -1/2 \cdots m_{qk} = 1/2 \cdots)} \right\} = \Lambda_{pq}.
\end{aligned}
\tag{A2}
$$

3. $[\Lambda_{pq}, \Lambda_{rs}]$ for $1 \leqslant p < q < r < s \leqslant n$

Before determining the value of this commutator, we express Λ_{pq}, Λ_{rs} of Eq. (16) in a more convenient expanded form as follows:

$$
\begin{aligned}
\Lambda_{pq} &= \sum_{j=1}^{2^{n-2}} (-)^{\sum_{a=p+1}^{q-1} (m_{ak} + 1/2)} \\
&\quad \times E_{(\cdots m_{pj} = 1/2 \cdots m_{qj} = -1/2 \cdots)(\cdots m_{pj} = -1/2 \cdots m_{qj} = 1/2 \cdots)} \\
&= \sum_{j=1}^{2^{n-4}} (-)^{\sum_{a=p+1}^{q-1} (m_{aj} + 1/2)} \\
&\quad \times \{ E_{(\cdots 1/2 \cdots -1/2 \cdots 1/2 \cdots 1/2 \cdots)(\cdots -1/2 \cdots 1/2 \cdots 1/2 \cdots 1/2 \cdots)} \\
&\quad + E_{(\cdots 1/2 \cdots -1/2 \cdots 1/2 \cdots -1/2)(\cdots -1/2 \cdots 1/2 \cdots 1/2 \cdots -1/2 \cdots)}
\end{aligned}
$$

$$+E_{(\cdots 1/2\cdots -1/2\cdots -1/2\cdots 1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots 1/2\cdots)}$$
$$+E_{(\cdots 1/2\cdots -1/2\cdots -1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots -1/2\cdots)}\}. \tag{A3}$$

where we have indicated explicitly the weights at the p, q, r, and s positions, and there is no summation over the indicated weights. Similarly,

$$\begin{aligned}\Lambda_{rs} = \sum_{k=1}^{2^{n-4}} (-)^{\sum_{b=r+1}^{s-1}(m_{bk}+1/2)} &\times\{E_{(\cdots 1/2\cdots 1/2\cdots 1/2\cdots -1/2)(\cdots 1/2\cdots 1/2\cdots -1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots 1/2\cdots -1/2\cdots 1/2\cdots -1/2\cdots)(\cdots 1/2\cdots -1/2\cdots -1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots -1/2\cdots 1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots -1/2\cdots -1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots -1/2\cdots -1/2\cdots 1/2\cdots)}\}.\end{aligned} \tag{A4}$$

Forming the commutator of (A3) and (A4) and using Eq. (2) as in (A1) and (A2), we obtain the result

$$\begin{aligned}[\Lambda_{Pq},\Lambda_{rs}] &= \sum_{k=1}^{2^{n-4}} (-)^{\sum_{a=p+1}^{q-1}\sum_{b=r+1}^{s-1}(m_{aj}+m_{bk}+1)} \\ &\times\{E_{(\cdots 1/2\cdots -1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots 1/2\cdots)} \\ &-E_{(\cdots 1/2\cdots -1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots 1/2\cdots 1/2\cdots)}\} \\ &=0.\end{aligned} \tag{A5}$$

4. $[\Lambda_{pq},\Lambda_{qr}]$ for $1\leqslant p<q<r\leqslant n$

As in obtaining (A5), we first express Λ_{pq}, Λ_{qr} in an expanded form as follows:

$$\begin{aligned}\Lambda_{pq} = \sum_{j=1}^{2^{n-3}} (-)^{\sum_{a=p+1}^{q-1}(m_{aj}+1/2)} &\times\{E_{(\cdots 1/2\cdots -1/2\cdots 1/2\cdots)(\cdots -1/2\cdots 1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots 1/2\cdots -1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots -1/2\cdots)}\},\end{aligned} \tag{A6}$$

$$\begin{aligned}\Lambda_{qr} = \sum_{k=1}^{2^{n-3}} (-)^{\sum_{b=q+1}^{r-1}(m_{bk}+1/2)} &\times\{E_{(\cdots 1/2\cdots 1/2\cdots -1/2\cdots)(\cdots 1/2\cdots -1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots -1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots -1/2\cdots 1/2\cdots)}\}.\end{aligned} \tag{A7}$$

Using Eqs. (A6) and (A7)

$$\begin{aligned}[\Lambda_{pq},\Lambda_{qr}] &= \sum_{k=1}^{2^{n-3}} (-)^{\left(\sum_{a=p+1}^{q-1}+\sum_{b=q+1}^{r-1}\right)(m_{ak}+1/2)} \\ &\times\{-E_{(\cdots 1/2\cdots 1/2\cdots -1/2\cdots)(\cdots -1/2\cdots 1/2\cdots 1/2\cdots)} \\ &+E_{(\cdots 1/2\cdots -1/2\cdots -1/2\cdots)(\cdots -1/2\cdots -1/2\cdots -1/2\cdots)}\} \\ &= \sum_{k=1}^{2^{n-2}} (-)^{\sum_{a=p+1}^{r-1}(m_{ak}+1/2)} \\ &\times E_{(\cdots 1/2\cdots m_{qk}\cdots -1/2\cdots)(\cdots -1/2\cdots m_{qk}\cdots 1/2\cdots)} = \Lambda_{pr}.\end{aligned} \tag{A8}$$

Some of the other nonzero results of Eq. (7) may be verified readily from Eq. (A8) on using the definitions of Eqs. (16)–(28) as follows:

$$[\Lambda_{pq},\Lambda_{rp}] = -[\Lambda_{rp},\Lambda_{pq}] = -\Lambda_{rq}, \tag{A9}$$

$$\begin{aligned}[\Lambda_{pq},\Lambda_{N+1-p\,r}] &= -[\Lambda_{pq},\Lambda_{N+1-r\,p}] \\ &= [\Lambda_{N+1-r\,p},\Lambda_{pq}] = \Lambda_{N+1-r\,q} = -\Lambda_{N+1-k\,r},\end{aligned} \tag{A10}$$

$$\begin{aligned}[\Lambda_{pq},\Lambda_{r\,N+1-q}] &= -[\Lambda_{pq},\Lambda_{q\,N+1-r}] = -\Lambda_{p\,N+1-r} \\ &= \Lambda_{r\,N+1-p}.\end{aligned} \tag{A11}$$

Equations (A8)–(A11) along with Eq. (A5) verify Eq. (7) for commutation relations of SO(N).

[1] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk (USSR) **71**, 1017 (1950).
[2] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[3] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[4] E. Chacon, doctoral thesis, Facultad de Ciencias, UNAM, Mexico, 1968.
[5] M. Moshinsky, *Group Theory and the Many-Body Problem* (Gordon and Breach, New York, 1968).
[6] K. T. Hecht, Nucl. Phys. A **102**, 11 (1967).
[7] M. Ichimura, Prog. Theoret. Phys. **33**, 215 (1965).
[8] K. J. Barnes, P. D. Jarvis, and I. J. Ketley, J. Phys. A **11**, 1025 (1978).
[9] G. Feldman, T. Fulton, and P. T. Mathews, J. Phys. G **8**, 295 (1982).
[10] E. U. Condon and H. Odabasi, *Atomic Structure* (Cambridge U. P., Cambridge, 1980).
[11] C. R. Sarma and S. Rettrup, J. Phys. A **13**, 2267 (1980); Phys. Lett. A **75**, 181 (1980).

Group representations in indefinite metric spaces

P. M. van den Broek
Department of Applied Mathematics, Twente University of Technology, 7500 AE Enschede, The Netherlands

(Received 20 April 1983; accepted for publication 18 July 1983)

A group G of symmetry transformations of the rays of an indefinite metric space V with metric operator η leads to a projective representation U of G in V in terms of η-unitary, η-antiunitary, η-pseudounitary, and η-pseudoantiunitary operators. We investigate the restrictions which are put on the irreducible components of U by the metric, and examine to what extent it is possible to decompose V into a direct sum of indefinite metric spaces, each carrying a projective representation of G. Attention is restricted to the cases where the subgroup of G which is represented by η-unitary operators is of index 1 or 2.

PACS numbers: 02.20. + b

1. INTRODUCTION

Our main motivation for studying group representations in indefinite metric spaces is the fact that twistor space is an indefinite metric space which carries in a natural way a (generalized) representation of the conformal group.[1,2] However, our results may be of interest in any field where indefinite metric spaces are used and symmetry arguments do play a role (Ref. 3 for references).

Let V be a n-dimensional complex vector space, $n \geqslant 3$, with a scalar product denoted by $(\,,\,)$ and let η be a Hermitian nonsingular linear operator on V. The indefinite metric of V is given by

$$\langle \phi,\psi \rangle = (\phi,\eta\psi), \tag{1.1}$$

and η is called the metric operator.

The rays of V are the one-dimensional linear subspaces of V. If $\psi \in V$ and $\psi \neq 0$, then $\boldsymbol{\psi}$ denotes the ray which contains ψ.

The rays $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are said to be orthogonal (denoted by $\langle \boldsymbol{\psi} \cdot \boldsymbol{\phi} \rangle = 0$) if and only if $\langle \psi,\phi \rangle = 0$ for each $\psi \in \boldsymbol{\psi}$ and each $\phi \in \boldsymbol{\phi}$.

The rays of V form the projective space $\mathbf{V}$. Let T be a bijective mapping of $\mathbf{V}$ onto $\mathbf{V}$ which has the property

$$\langle \boldsymbol{\psi} \cdot \boldsymbol{\phi} \rangle = 0 \Leftrightarrow \langle T\boldsymbol{\psi} \cdot T\boldsymbol{\phi} \rangle = 0. \tag{1.2}$$

Then we have the following theorem:

Theorem 1: There exists a operator U on V such that

(i) $\psi \in \boldsymbol{\psi} \Rightarrow U\psi \in T\boldsymbol{\psi}$, (1.3)

(ii) Either U is linear and satisfies

$$\langle U\phi,U\psi \rangle = \langle \phi,\psi \rangle \quad \forall \phi,\psi \in V, \tag{1.4}$$

or U is linear and satisfies

$$\langle U\phi,U\psi \rangle = -\langle \phi,\psi \rangle \quad \forall \phi,\psi \in V, \tag{1.5}$$

or U is antilinear and satisfies

$$\langle U\phi,U\psi \rangle = \langle \psi,\phi \rangle \quad \forall \phi,\psi \in V, \tag{1.6}$$

or U is antilinear and satisfies

$$\langle U\phi,U\psi \rangle = -\langle \psi,\phi \rangle \quad \forall \phi,\psi \in V. \tag{1.7}$$

Moreover, U is determined uniquely by T up to a factor of modulus 1.

This theorem, first given in Ref. 4, is a generalization of a theorem in Ref. 3, which in turn is the generalization of Wigner's theorem to indefinite spaces. The set of operators enumerated under (ii) of the theorem are called η-unitary, η-pseudounitary, η-antiunitary, and η-pseudoantiunitary, respectively.

Let G be a finite group and suppose that for each $g \in G$ there is a bijective mapping $T(g)$ of $\mathbf{V}$ onto $\mathbf{V}$ such that Eq. (1.2) holds. Then the theorem above gives us for each $g \in G$ an operator $U(g)$ on V which is either η-unitary, or η-antiunitary, or η-pseudounitary, or η-pseudoantiunitary. Since $U(g)$ is determined by $T(g)$ up to a factor of modulus 1, we have

$$U(g)U(g') = \sigma(g,g')U(gg') \quad \forall g,g' \in G, \tag{1.8}$$

where the mapping σ of $G \times G$ into the complex numbers of modulus 1 is called a factor system of G.

Let G_0 be the normal subgroup of G consisting of those elements g of G for which $U(g)$ is η-unitary. Let a, b, and c be elements of G (if any exist) such that $U(a)$ is η-antiunitary, $U(b)$ is η-pseudounitary, and $U(c)$ is η-pseudoantiunitary. Then aG_0, bG_0, and cG_0 denote the cosets of G with respect to G_0 which consist of the elements g of G for which $U(g)$ is η-antiunitary, η-pseudounitary, and η-pseudoanitunitary, respectively. For the coset decomposition of G with respect to G_0 there are the following five possibilities:

(I) $G = G_0$, (1.9a)

(II) $G = G_0 + aG_0$, (1.9b)

(III) $G = G_0 + bG_0$, (1.9c)

(IV) $G = G_0 + cG_0$, (1.9d)

(V) $G = G_0 + aG_0 + bG_0 + cG_0$. (1.9e)

From Eqs. (1.3)–(1.6), it follows that

$$\langle U(g)\phi,\, U(g)\psi \rangle = (-)_g \langle \phi,\psi \rangle^g, \tag{1.10}$$

where $(-)_g$ is defined by

$$(-)_g = \begin{cases} +1 & \text{if } g \in G_0 + aG_0 \\ -1 & \text{if } g \notin G_0 + aG_0 \end{cases} \tag{1.11}$$

and, if $\lambda \in C$, λ^g is defined by

$$\lambda^g = \begin{cases} \lambda & \text{if } g \in G_0 + bG_0 \\ \lambda^* & \text{if } g \notin G_0 + bG_0, \end{cases} \tag{1.12}$$

the asterisk denoting complex conjugation.

Now choose a basis $\{e_1,e_2,\ldots,e_n\}$ of V which is orthonormal with respect to the ordinary scalar product and form the matrices of $U(g)$ and η via

$$U(g)e_i = \sum_{j=1}^{n} U(g)_{ji} e_j \tag{1.13}$$

 0022-2488/84/051205-06$02.50

and

$$\eta e_i = \sum_{j=1}^{n} \eta_{ji} e_j. \tag{1.14}$$

These matrices are again denoted by $U(g)$ and η; they satisfy

$$U(g)U(g')^g = \sigma(g,g')U(gg'), \tag{1.15}$$

where the upper index g is defined as in Eq. (1.12), and

$$U^\dagger(g)\eta U(g) = (-)_g \eta^g, \tag{1.16}$$

where $U^\dagger$ denotes the Hermitian conjugate of U.

Matrices $U(g)$ satisfying Eq. (1.15) form a projective linear–antilinear (PLA) representation of G. If all the matrices are unitary, they form a projective unitary–antiunitary (PUA) representation of G. For simplicity we call such representations PLA (resp. PUA) representations even if all operators are linear.

Two PLA representations U and U' of G are equivalent if there exists a nonsingular matrix A such that

$$U'(g) = A^{-1}U(g)A^g \quad \forall g \in G. \tag{1.17}$$

If U satisfies Eq. (1.16), then U', defined by Eq. (1.17) for some nonsingular matrix A, satisfies

$$U'^\dagger(g)\eta' U'(g) = (-)_g \eta'^g, \tag{1.18}$$

where

$$\eta' = A^\dagger \eta A. \tag{1.19}$$

So the equivalence transformation given by A transforms the Hermitian matrix η into the Hermitian matrix η', which has the same signature as η. U is said to be decomposable if there exists an equivalent PLA representation U' such that

$$U'(g) = \begin{pmatrix} U'_1(g) & 0 \\ 0 & U'_2(g) \end{pmatrix} \tag{1.20}$$

and

$$\eta' = \begin{pmatrix} \eta'_1 & 0 \\ 0 & \eta'_2 \end{pmatrix}. \tag{1.21}$$

Equation (1.18) then becomes

$$U'^\dagger_i(g)\eta'_i U'_i(g) = (-)_g \eta'^g_i \quad (i = 1,2). \tag{1.22}$$

The aim of this paper is twofold. Our first objective is to examine to what extent it is possible to decompose a PLA representation of G which satisfies Eq. (1.16). A decomposable PLA representation of G is reducible, but the reverse is not generally true, due to the restriction made in Eq. (1.21). Our second objective is to examine which restrictions are put on the irreducible components of U by Eq. (1.16). We will restrict ourselves to the case where G_0 is a subgroup of index 1 or 2 of G; i.e., we consider the cases I, II, III, and IV of Eq. (1.9). The study of case V is left to a later paper.

2. GENERAL REDUCTION

Due to a result of Murthy,[5] U is equivalent to a PLA representation whose matrices are all unitary, i.e., there exists a nonsingular matrix A such that the matrices $U'(g)$ given by

$$U'(g) = A^{-1}U(g)A^g \tag{2.1}$$

form a PUA representation of G. Equation (1.16) then becomes

$$\eta' U'(g) = (-)_g U'(g)\eta'^g, \tag{2.2}$$

where

$$\eta' = A^\dagger \eta A. \tag{2.3}$$

There exists a unitary matrix X such that η'', defined by

$$\eta'' = X^\dagger \eta' X, \tag{2.4}$$

is a real diagonal matrix. If

$$U''(g) = X^{-1}U'(g)X^g, \tag{2.5}$$

then U'' is a PUA representation of G which satisfies

$$\eta'' U''(g) = (-)_g U''(g)\eta''. \tag{2.6}$$

Writing out matrix elements gives

$$\eta''_{ii} U''(g)_{ij} = (-)_g U''(g)_{ij}\eta''_{jj}. \tag{2.7}$$

Now consider the cases I and II [Eqs. (1.9a) and (1.9b)]. In these cases $(-)_g = +1$ for each $g \in G$. Then $U''(g)_{ij} = 0$ whenever $\eta''_{ii} \neq \eta''_{jj}$. So by a suitable renumbering of rows and columns we have

$$U''(g) = \sum_i \oplus U''_i(g) \tag{2.8}$$

and

$$\eta'' = \sum_i \oplus \eta''_{ii}, \tag{2.9}$$

where each η''_i is a multiple of the unit matrix.

So we have decomposed U'' as a direct sum of PUA representations U''_i of G and η'' in a corresponding direct sum of real multiples of the unit matrix. Each U''_i can now be reduced to a direct sum of irreducible PUA representations of G, and this equivalence transformation does not change η''_i. So in the cases I and II our goal has been achieved; we have proved the following theorem.

Theorem 2: In the cases I and II there are no restrictions put by the metric on the irreducible components of a PLA representation U of G, and U is decomposable into irreducible PUA representations.

Now consider the cases where $g \in G$ exist with $(-)_g = -1$. Then we have from Eq. (2.7) that $U''(g)_{ij} = 0$ whenever $|\eta''_{ii}| \neq |\eta''_{jj}|$. So by a suitable renumbering of rows and columns we have the decomposition of Eqs. (2.8) and (2.9); but here each η''_i is a real diagonal matrix with all diagonal elements equal up to a sign.

Each U''_i can now be studied separately; it satisfies

$$\eta''_i U''_i(g) = (-)_g U''_i(g)\eta''_i. \tag{2.10}$$

Let the diagonal elements of η''_i be a_i and $-a_i$ with $a_i > 0$. Since $g \in G$ exists with $(-)_g = -1$, it follows from Eq. (2.10) that $\mathrm{Tr}\,\eta''_i = 0$. So a_i and $-a_i$ have equal multiplicity.

The equivalence transformation on U''_i and η''_i by the matrix A_i which is equal to $a_i^{-1/2}$ times the unit matrix leaves U''_i unchanged, but turns η''_i into a diagonal matrix with diagonal elements $+1$ and -1 with equal multiplicity. So, without loss of generality, we may assume that η''_i has eigenvalues $+1$ and -1. In the next section we study case III and in Sec. 4 we treat case IV.

3. η-UNITARY AND η-PSEUDOUNITARY OPERATORS

In this section we consider case III [Eq. (1.9c)], i.e., the case where G is represented by η-unitary and η-pseudounitary operators. G_0, the subgroup of G which is represented by η-unitary operators, has index 2. According to the theory of induced PUA representations, which in its most general form is given in Ref. 6, each irreducible PUA representation of G is either of type A or of type B.

A PUA representation of type A is equivalent to a PUA representation D, which can be written as

$$D(g) = \begin{pmatrix} \Delta(g) & 0 \\ 0 & \tilde{\Delta}(g) \end{pmatrix} \quad \forall g \in G_0, \tag{3.1}$$

$$D(b) = \begin{pmatrix} 0 & \sigma(b,b)\Delta(b^2) \\ \mathbb{1}_d & 0 \end{pmatrix}. \tag{3.2}$$

Here Δ is an irreducible PUA representation of G_0; $\tilde{\Delta}$ is an irreducible PUA representation of G_0 which is related to Δ by

$$\tilde{\Delta}(g) = \sigma(g,b)\sigma^*(b,b^{-1}gb)\Delta(b^{-1}gb) \tag{3.3}$$

and which is not equivalent to Δ. The dimension of Δ is d. $\mathbb{1}_d$ denotes the unit matrix of dimension d. D is determined up to equivalence by the equivalence classes of PUA representations of G_0 containing Δ and $\tilde{\Delta}$. A PUA representation D of type B has the property that its restriction $\Delta = D{\downarrow}G_0$ to G_0 is irreducible. In this case Δ and $\tilde{\Delta}$ are equivalent. D is not uniquely determined by Δ: The PUA representation D' of G, given by

$$D'(g) = \begin{cases} D(g) & \text{if } g \in G_0 \\ -D(g) & \text{if } g \notin G_0, \end{cases} \tag{3.4}$$

has the same restriction to G_0, but is not equivalent to D. However, the pair (D,D') is determined up to equivalence by the equivalence class of Δ. A pair of PUA representations of G of type B which are not equivalent, but whose restrictions to G_0 are equivalent, are said to be related. Let U be a PUA representation of G with

$$\eta U(g) = (-)_g U(g)\eta. \tag{3.5}$$

Due to the results of the previous section, we may assume that the Hermitian matrix η has eigenvalues $+1$ and -1 only, with equal multiplicity. We may now perform a unitary equivalence transformation

$$U'(g) = W^{-1}U(g)W, \tag{3.6}$$

$$\eta' = W^{\dagger}\eta W \tag{3.7}$$

such that the PUA representation U' of G has the following properties:

(i) U' is a direct sum of irreducible PUA representations D_i.

(ii) The components D_i of this direct sum are pairwise either equal or inequivalent.

(iii) Components of type A have the form of Eqs. (3.1) and (3.2).

(iv) Components of type B which are related satisfy Eq. (3.4).

(v) The components are arranged into blocks U'_i:

$$U'(g) = \sum_i \oplus U'_i(g) \tag{3.8}$$

according to the following rules:

(1) irreducible components of type A are in the same block if and only if they are equivalent (and thus equal);

(2) irreducible components of type B are in the same block if and only if they are either equal or related.

Note that η can be chosen to be a diagonal matrix due to the results of the previous section, but this property is not inherited by η'. If the Hermitian matrix η' is divided into blocks in the same way as U',

$$\eta' = \begin{pmatrix} \eta'_{[1,1]} \cdots \eta'_{[1,p]} \\ \eta'_{[p,1]} \cdots \eta'_{[p,p]} \end{pmatrix}, \tag{3.9}$$

we find

$$\eta'_{[i,k]} U'_k(g) = (-)_g U'_i(g)\eta'_{[i,k]}. \tag{3.10}$$

If $i \neq k$, the blocks U'_k and U'_i, when restricted to G_0, are PUA representations of G_0 which have no irreducible components in common. Due to Schur's lemma, we thus have

$$\eta'_{[i,k]} = 0 \tag{3.11}$$

if $i \neq k$. This means that we have found a decomposition of U' into blocks U'_i; each block U'_i can now be studied further separately. Let V be some block U'_i, and let ζ be the Hermitian matrix $\eta'_{[i,i]}$. Then

$$\zeta V(g) = (-)_g V(g)\zeta. \tag{3.12}$$

ζ is Hermitian and has eigenvalues $+1$ and -1 with equal multiplicity. Consider first the case that V consists of irreducible PUA representations D of G type A. We may then write

$$V(g) = \mathbb{1}_n \otimes \begin{pmatrix} \Delta(g) & 0 \\ 0 & \tilde{\Delta}(g) \end{pmatrix} \quad \forall g \in G_0, \tag{3.13}$$

$$V(b) = \mathbb{1}_n \otimes \begin{pmatrix} 0 & \sigma(b,b)\Delta(b^2) \\ \mathbb{1}_d & 0 \end{pmatrix}. \tag{3.14}$$

Here n is the multiplicity of D in V. Divide ζ into blocks in the same way as V:

$$\zeta = \begin{pmatrix} \zeta_{[1,1]}\zeta_{[1,2]} \cdots \zeta_{[1,n]} \\ \zeta_{[n,1]} \cdots \zeta_{[n,n]} \end{pmatrix}. \tag{3.15}$$

From Eqs. (3.12)–(3.14) and Schur's lemma it follows that each $\zeta_{[ij]}$ satisfies

$$\zeta_{[ij]} = \begin{pmatrix} a_{ij}\mathbb{1}_d & 0 \\ 0 & -a_{ij}\mathbb{1}_d \end{pmatrix} \tag{3.16}$$

for some $a_{ij} \in C$. So we can write

$$\zeta = A \otimes \begin{pmatrix} \mathbb{1}_d & 0 \\ 0 & -\mathbb{1}_d \end{pmatrix}, \tag{3.17}$$

where A is a Hermitian $n \times n$ matrix. There exists a unitary matrix X such that A', defined by

$$A' = X^{\dagger}AX \tag{3.18}$$

is a diagonal matrix; its diagonal elements are equal to $+1$ or -1. Define the unitary matrix Y by

$$Y = X \otimes \begin{pmatrix} \mathbb{1}_d & 0 \\ 0 & \mathbb{1}_d \end{pmatrix} \tag{3.19}$$

and perform the equivalence transformation on V and ζ by

Y. Then V is left unchanged, but ζ turns into ζ', where

$$\zeta' = A' \otimes \begin{pmatrix} \mathbf{1}_d & 0 \\ 0 & -\mathbf{1}_d \end{pmatrix}. \tag{3.20}$$

So ζ' is a diagonal matrix with diagonal elements $+1$ and -1 with equal multiplicity. This means that V has been decomposed to a direct sum of irreducible PUA representations of G.

Now consider the case that V consists of irreducible PUA representations of G of type B; suppose the multiplicity of D in V is n and the multiplicity of D' in V is m. With a suitable numbering of rows and columns we have

$$V(g) = \begin{pmatrix} \mathbf{1}_n \otimes D(g) & 0 \\ 0 & \mathbf{1}_m \otimes D(g) \end{pmatrix} \quad \forall g \in G_0 \tag{3.21}$$

and

$$V(b) = \begin{pmatrix} \mathbf{1}_n \otimes D(b) & 0 \\ 0 & -\mathbf{1}_m \otimes D(b) \end{pmatrix}. \tag{3.22}$$

Let ζ be divided accordingly:

$$\zeta = \begin{pmatrix} \zeta_{11} & \zeta_{12} \\ \zeta_{21} & \zeta_{22} \end{pmatrix}. \tag{3.23}$$

From Eqs. (3.12) and (3.21) and Schur's lemma it follows that $\zeta_{11} = 0$ and $\zeta_{22} = 0$. Since $\det \zeta \neq 0$, the dimension of ζ_{11} and ζ_{22} must be equal. Thus $n = m$, and we arrive at the following restriction put on the irreducible components of U by the metric:

Theorem 3: In a PLA representation of G belonging to case III related irreducible PUA representations have the same multiplicity.

From Eq. (3.12), Schur's lemma, and the Hermiticity of ζ, it follows that

$$\zeta_{12} = \chi \otimes \mathbf{1}_d \tag{3.24}$$

and

$$\zeta_{21} = \chi^\dagger \otimes \mathbf{1}_d \tag{3.25}$$

for some $n \times n$ matrix χ. Since ζ is Hermitian and has eigenvalues $+1$ and -1 only, the matrix ζ^2 equals $\mathbf{1}_n$. This implies that χ is unitary. Define the unitary matrix Y by

$$Y = \begin{pmatrix} \mathbf{1}_{nd} & 0 \\ 0 & \chi^\dagger \otimes \mathbf{1}_d \end{pmatrix} \tag{3.26}$$

and perform the equivalence transformation on V and ζ by Y. Then V is left unchanged, but ζ turns into ζ', where

$$\zeta' = \begin{pmatrix} 0 & \mathbf{1}_{nd} \\ \mathbf{1}_{nd} & 0 \end{pmatrix}. \tag{3.27}$$

By a suitable reordering of rows and columns, we may now write V and ζ' as follows:

$$V(g) = \mathbf{1}_n \otimes \begin{pmatrix} D(g) & 0 \\ 0 & D(g) \end{pmatrix} \quad \forall g \in G_0, \tag{3.28}$$

$$V(b) = \mathbf{1}_n \otimes \begin{pmatrix} D(b) & 0 \\ 0 & -D(b) \end{pmatrix}, \tag{3.29}$$

$$\zeta' = \mathbf{1}_n \otimes \begin{pmatrix} 0 & \mathbf{1}_d \\ \mathbf{1}_d & 0 \end{pmatrix}. \tag{3.30}$$

So V has been decomposed into components which are pairs of an irreducible PUA representation and its related PUA representation. A further decomposition of these components is not possible. This follows from the fact that if V is an irreducible PUA representation of G of type B, there exists no Hermitian matrix ζ which satisfies Eq. (3.12). So we have derived the following theorem:

Theorem 4: Any PLA representation of G belonging to case III is decomposable into irreducible PUA representations of type A and pairs of related irreducible PUA representations of type B.

4. η-UNITARY AND η-PSEUDOANTIUNITARY OPERATORS

In this section we consider case IV [Eq. (1.9d)], i.e., the case where G is represented by η-unitary and η-pseudoantiunitary operators. G_0, the subgroup of G which is represented by η-unitary operators had index 2. Each irreducible PUA representation of G is either of type A, or of type B or of type C.

A PUA representation D of type A has the property that its restriction $\Delta = D{\downarrow}G_0$ to G_0 is irreducible. D is determined up to equivalence by the equivalence class of Δ. A PUA representation of type B is equivalent to a PUA representation D, which can be written as

$$D(g) = \begin{pmatrix} \Delta(g) & 0 \\ 0 & \Delta(g) \end{pmatrix} \quad \forall g \in G_0, \tag{4.1}$$

$$D(c) = \begin{pmatrix} 0 & \mathscr{D} \\ -\mathscr{D} & 0 \end{pmatrix}. \tag{4.2}$$

D is determined up to equivalence by the equivalence class of Δ. A PUA representation of type C is equivalent to a PUA representation D, which can be written as

$$D(g) = \begin{pmatrix} \Delta(g) & 0 \\ 0 & \tilde{\Delta}(g) \end{pmatrix} \quad \forall g \in G_0, \tag{4.3}$$

$$D(c) = \begin{pmatrix} 0 & \sigma(c,c)\Delta(c^2) \\ \mathbf{1}_d & 0 \end{pmatrix}. \tag{4.4}$$

Δ and $\tilde{\Delta}$ are irreducible d-dimensional PUA representations of G_0, which are not equivalent to each other and are related by

$$\tilde{\Delta}(g) = \sigma(g,c)\sigma^*(c,c^{-1}gc)\Delta^*(c^{-1}gc). \tag{4.5}$$

D is determined up to equivalence by the equivalence classes of Δ and $\tilde{\Delta}$. Let U be a PUA representation of G with

$$\eta U(g) = (-)_g U(g)\eta^g. \tag{4.6}$$

Due to the results of Sec. 2, we may assume that the Hermitian matrix η has eigenvalues $+1$ and -1 only, with equal multiplicity. We may now perform a unitary equivalence transformation

$$U'(g) = W^{-1}U(g)W^g, \tag{4.7}$$

$$\eta' = W^\dagger \eta W \tag{4.8}$$

such that the PUA representation U' of G has the following properties:

(i) U' is a direct sum of irreducible PUA representations D_i.

(ii) The components D_i of this direct sum are pairwise either equal or inequivalent.

(iii) Components of type B have the form of Eqs. (4.1)

and (4.2).

(iv) Components of type C have the form of Eqs. (4.3) and (4.4).

(v) The components are arranged into blocks U_i':

$$U'(g) = \sum_i \oplus U_i'(g) \tag{4.9}$$

such that irreducible components of U' are in the same block if and only if they are equivalent (and thus equal).

Equation (4.9) is a decomposition of U'. This is proved in the same way as the corresponding decomposition in the previous section. So each block U_i' can now be studied separately.

Let V be some block U_i', and let ζ be the corresponding block of η'. Then

$$\zeta V(g) = (-)_g V(g)\zeta^g. \tag{4.10}$$

ζ is Hermitian and has eigenvalues $+1$ and -1 with equal multiplicity. Consider first the case that V consists of irreducible PUA representations D of G of type A. Then we may write

$$V(g) = 1_n \otimes \Delta(g) \quad \forall g \in G_0 \tag{4.11}$$

and

$$V(c) = 1_n \otimes D(c). \tag{4.12}$$

Here n is the multiplicity of D in V. Divide ζ into blocks in the same way as V [Eq. (3.15)]. Then it follows from Eqs. (4.10) and (4.11) and Schur's lemma that

$$\zeta_{[i,j]} = a_{ij} 1_d \tag{4.13}$$

for some $a_{ij} \in C$. We thus have

$$\zeta = A \otimes 1_d \tag{4.14}$$

for some Hermitian $n \times n$ matrix A. From Eqs. (4.10) and (4.12) it now follows that

$$A = -A^*. \tag{4.15}$$

Thus the dimension of A is even, and thus we have arrived at the following restriction put by the metric on the irreducible components of U:

Theorem 5: In a PLA representation of G belonging to case IV the irreducible PUA representations of type A have even multiplicity.

Now there exists an orthogonal matrix B such that

$$B^\dagger A B = 1_{n/2} \otimes \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}. \tag{4.16}$$

This follows from §13 of Chap. IX of Ref. 7, since iA is real, skew-symmetric, and has eigenvalues $+i$ and $-i$ with equal multiplicity. Define

$$C = B \otimes 1_d \tag{4.17}$$

and perform the equivalence transformation on V and ζ by C. Then V is left unchanged, and ζ is transformed into

$$\zeta' = 1_{q/2} \otimes \begin{pmatrix} 0 & i1_d \\ -i1_d & 0 \end{pmatrix}. \tag{4.18}$$

It is seen that we have decomposed V into blocks containing two equal irreducible PUA representations of type A. Further decomposition is not possible, due to Theorem 5.

Next consider the case where V consists of irreducible PUA representations D of type B. With a suitable numbering of rows and columns, we may write

$$V(g) = \begin{pmatrix} 1_n \otimes \Delta(g) & 0 \\ 0 & 1_n \otimes \Delta(g) \end{pmatrix} \quad \forall g \in G_0 \tag{4.19}$$

and

$$V(c) = \begin{pmatrix} 0 & 1_n \otimes \mathscr{D} \\ -1_n \otimes \mathscr{D} & 0 \end{pmatrix}. \tag{4.20}$$

From Eqs. (4.10), (4.19), and (4.20) and Schur's lemma it follows that

$$\zeta = \chi \otimes 1_d, \tag{4.21}$$

where χ is a Hermitian $2n \times 2n$ matrix which is partitioned into square blocks as follows:

$$\chi = \begin{pmatrix} \chi_1 & \chi_2 \\ \chi_2^* & -\chi_1^* \end{pmatrix}. \tag{4.22}$$

Due to a lemma, which we prove in the Appendix, there exists a unitary $2n \times 2n$ matrix U which is partitioned into square blocks as follows:

$$U = \begin{pmatrix} U_1 & U_2 \\ -U_2^* & U_1^* \end{pmatrix} \tag{4.23}$$

and which satisfies

$$U^\dagger \chi U = \begin{pmatrix} 1_n & 0 \\ 0 & -1_n \end{pmatrix}. \tag{4.24}$$

Define the unitary matrix Y by

$$Y = U \otimes 1_d \tag{4.25}$$

and perform the equivalence transformation on V and ζ by Y. Then V is left unchanged and ζ is transformed into

$$\zeta' = U^\dagger \chi U \otimes 1_d = \begin{pmatrix} 1_{nd} & 0 \\ 0 & -1_{nd} \end{pmatrix}. \tag{4.26}$$

This means that V has been decomposed to a direct sum of irreducible PUA representations of G.

Finally consider the case that V consists of irreducible PUA representations D of type C. With a suitable numbering of rows and columns, we may write

$$V(g) = \begin{pmatrix} 1_n \otimes \Delta(g) & 0 \\ 0 & 1_n \otimes \tilde{\Delta}(g) \end{pmatrix} \quad \forall g \in G_0 \tag{4.27}$$

and

$$V(c) = \begin{pmatrix} 0 & \sigma(c,c) 1_n \otimes \Delta(c^2) \\ 1_{nd} & 0 \end{pmatrix}. \tag{4.28}$$

From Eqs. (4.10), (4.27), and (4.28) and Schur's lemma it follows that

$$\zeta = \chi \otimes 1_d, \tag{4.29}$$

where χ is a Hermitian $2n \times 2n$ matrix which is partitioned into square blocks as follows:

$$\chi = \begin{pmatrix} \chi_1 & 0 \\ 0 & -\chi_1^* \end{pmatrix}. \tag{4.30}$$

Let U be a unitary $n \times n$ matrix such that $U^\dagger \chi_1 U$ is on diagonal form. Define the unitary matrix Y by

$$Y = \begin{pmatrix} U & 0 \\ 0 & U^* \end{pmatrix} \otimes 1_d \tag{4.31}$$

and perform the equivalence transformation on V and ζ by Y. Then V is left unchanged and ζ is transformed into

$$\zeta' = \begin{pmatrix} U^\dagger \chi_1 U & 0 \\ 0 & -(U^\dagger \chi_1 U)^* \end{pmatrix} \otimes \mathbb{1}_d, \tag{4.32}$$

which is on diagonal form. This means that V has been decomposed into a direct sum of irreducible PUA representations of G. So we have derived the following theorem:

Theorem 6: Any PLA representation of G belonging to case IV is decomposable into a direct sum of irreducible PUA representations of type B, irreducible PUA representations of type C, and pairs of irreducible PUA representations of type A.

APPENDIX

This appendix is devoted to the proof of the following lemma:

Lemma: Let H be a $2n \times 2n$ Hermitian matrix which is partitioned into square blocks as follows:

$$H = \begin{pmatrix} H_1 & H_2 \\ H_2^* & -H_1^* \end{pmatrix} \tag{A1}$$

and which has only eigenvalues $+1$ and -1. Let I be defined by

$$I = \begin{pmatrix} \mathbb{1}_n & 0 \\ 0 & -\mathbb{1}_n \end{pmatrix}. \tag{A2}$$

There exists a unitary matrix U which is partitioned into square blocks as follows:

$$U = \begin{pmatrix} U_1 & -U_2^* \\ U_2 & U_1^* \end{pmatrix}. \tag{A3}$$

and which has the property that

$$U^\dagger H U = I. \tag{A4}$$

Proof: Since $\mathrm{Tr}\, H = 0$ the eigenvalues $+1$ and -1 have equal multiplicity. Thus there exists a unitary matrix V such that

$$V^\dagger H V = I. \tag{A5}$$

This implies that

$$HV = VI. \tag{A6}$$

If V is partitioned into square blocks as follows,

$$V = \begin{pmatrix} V_1 & V_3 \\ V_2 & V_4 \end{pmatrix}, \tag{A7}$$

then Eq. (A6) can be written as

$$H_1 V_1 + H_2 V_2 = V_1, \tag{A8}$$

$$H_1 V_3 + H_2 V_4 = -V_3, \tag{A9}$$

$$H_2^* V_1 - H_1^* V_2 = V_2, \tag{A10}$$

$$H_2 V_3^* - H_1^* V_4 = -V_4. \tag{A11}$$

Now define the matrix U by

$$U = \begin{pmatrix} U_1 & U_3 \\ U_2 & U_4 \end{pmatrix} = \begin{pmatrix} V_1 & -V_2^* \\ V_2 & V_1^* \end{pmatrix}. \tag{A12}$$

Then U satisfies equations (A8)–(A11) if V_i in these equations is replaced by U_i. Consequently,

$$HU = UI. \tag{A13}$$

Since U has the form prescribed by Eq. (A3), the lemma is proved if we show that U is unitary. Since V is unitary, we have

$$V_1^\dagger V_1 + V_2^\dagger V_2 = \mathbb{1}. \tag{A14}$$

Expressing $U^\dagger U$ in terms of V_1 and V_2 gives, with Eq. (A14),

$$U^\dagger U = \begin{pmatrix} \mathbb{1} & -V_1^\dagger V_2^* + V_2^\dagger V_1^* \\ -V_2^{*\dagger} V_1 + V_1^{*\dagger} V_2 & \mathbb{1} \end{pmatrix}. \tag{A15}$$

Equation (A13) implies

$$U^\dagger H U = U^\dagger U I. \tag{A16}$$

Taking the Hermitian adjoint of equation (A13) and multiplying with U from the right gives

$$U^\dagger H U = I U^\dagger U. \tag{A17}$$

From Eqs. (A16) and (A17), it follows that $U^\dagger U$ commutes with I. This fact, together with Eq. (A15), implies that $U^\dagger U$ is the unit matrix. Thus U is unitary, which proves the lemma.

[1] R. Penrose, "Twistor theory, its aims and achievements," in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1975).

[2] P. M. van den Broek, "Twistor space and the conformal group," Memorandum No. 361, Department of Applied Mathematics, Twente University of Technology, The Netherlands, 1981.

[3] L. Bracci, G. Morchio, and F. Strocchi, Commun. Math. Phys. **41**, 289–99 (1975).

[4] P. M. van den Broek, "Twistor geometry," Memorandum No. 410, Department of Applied Mathematics, Twente University of Technology, The Netherlands, 1982.

[5] V. Murthy, J. Math. Phys. **7**, 853–7 (1966).

[6] R. Shaw and J. Lever, Commun. Math. Phys. **38**, 257–77 (1974).

[7] F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1959), Vol. I.

Analytic expressions for the matrix elements of generators of Sp(6) in an Sp(6)⊃U(3) basis

O. Castaños
Centro de Estudios Nucleares, UNAM Apdo. Postal 20-364, Villa Obregón 01000 México, D.F., Mexico

E. Chacón and M. Moshinsky[a)]
Instituto de Física, UNAM Apdo. Postal 20-364, Villa Obregón 01000 México, D.F., Mexico

(Received 26 July 1983; accepted for publication 9 December 1983)

Work done by many authors indicates that important tools for calculations in microscopic collective models are the matrix elements of the generators of the symplectic group Sp(6) in an Sp(6)⊃U(3) basis. Rosensteel has derived recursion relations for these matrix elements while Filippov has determined them using generating function technique, but it would also be convenient to have explicit and analytic formulas for them. This is what we do in this paper for the case of closed shells, i.e., when the irreducible representation (irrep) of Sp(6) is characterized by equal values for the three weight generators in the lowest weight state. We also indicate how our results can be extended to the case of arbitrary irreps of Sp(6), i.e., when we have open shells.

PACS numbers: 02.20. + b, 03.65.Fd, 21.60.Ev, 21.60.Fw

1. INTRODUCTION

Work done in the last few years by Rosensteel and Rowe[1] and Filippov[2] and others, on the microscopic origin of nuclear collective models, has indicated that an important tool for calculations is the determination of the matrix elements of the generators of the symplectic group Sp(6) in an Sp(6)⊃U(3) basis.

Rosensteel[3] has derived recursion relations for these matrix elements, while Filippov has used generating function techniques,[2] but it would also be convenient to have explicit and simple analytic formulas for them. This is what we shall proceed to do here for the case of closed shells, i.e., when the irreducible representation (irrep) of Sp(6) is characterized by equal values for the three weight generators in the lowest weight state. The general case when the three weights have different values, corresponding to open shells, will be discussed in a future publication.

It is convenient to express the generators of Sp(6) in terms of coordinates x_{is} and momenta p_{is} for a system of particles in three-dimensional space, i.e., $i = 1, 2, 3, s = 1, 2, \ldots, n$. Taking units in which $\hbar$, the mass of the particles, and frequency of an oscillator are 1, we get for the creation and annihilation operators the expressions

$$\eta_{is} = (1/\sqrt{2})(x_{is} - i p_{is}), \quad \xi_{is} = (1/\sqrt{2})(x_{is} + i p_{is}). \tag{1.1}$$

The 21 generators of the Sp(6) group can then be written as[1,3,4]

$$B_{ij}^{+} = \sum_{s=1}^{n} \eta_{is}\eta_{js}, \tag{1.2a}$$

$$C_{ij} = \frac{1}{2}\sum_{s=1}^{n} (\eta_{is}\xi_{js} + \xi_{js}\eta_{is}), \tag{1.2b}$$

$$B_{ij} = \sum_{s=1}^{n} \xi_{is}\xi_{js}, \tag{1.2c}$$

which, from $[\xi_{jt}, \eta_{is}] = \delta_{ij}\delta_{st}$, satisfy the commutation relations

$$[B_{ij}, B_{i'j'}] = [B_{ij}^{+}, B_{i'j'}^{+}] = 0, \tag{1.3a,b}$$

$$[C_{ij}, B_{i'j'}^{+}] = B_{ij'}^{+}\delta_{ji'} + B_{ii'}^{+}\delta_{jj'}, \tag{1.3c}$$

$$[C_{ij}, B_{i'j'}] = -B_{jj'}\delta_{ii'} - B_{ji'}\delta_{ij'}, \tag{1.3d}$$

$$[B_{ij}, B_{i'j'}^{+}] = C_{j'j}\delta_{ii'} + C_{i'j}\delta_{ij'} + C_{j'i}\delta_{ji'} + C_{ii'}\delta_{jj'}, \tag{1.3e}$$

$$[C_{ij}, C_{i'j'}] = C_{ij'}\delta_{i'j} - C_{i'j}\delta_{ij'}, \tag{1.3f}$$

corresponding to the Lie algebra of Sp(6). From (1.3f) the C_{ij} are the generators of the U(3) subgroup of Sp(6) that was introduced by Elliott.[5] From a physical standpoint we will also be interested in classifying our basis states by an O(3) subgroup of U(3) whose generators are

$$L_i = -i\sum_{j,k=1}^{3} \epsilon_{ijk}C_{jk}, \quad i = 1, 2, 3, \tag{1.4}$$

where the ϵ_{ijk} is the antisymmetric tensor in the three indices.

We need now to discuss the states that are basis for the irreps Sp(6)⊃U(3), with respect to which we would like to get the matrix elements of the generators (1.2). It is known[1,2] that these states can be constructed by applying polynomial functions of the B_{ij}^{+} of (1.2) to Slater determinants for the many-body systems in which the levels of an oscillator potential are filled compactly by nucleons subject to the restrictions of the Pauli principle. Both the polynomials and the Slater determinants are characterized by definite irreps of U(3), which are coupled with a Wigner coefficient of this group,[1,3] to give for the full state also a definite irrep of U(3). In the case of closed shells the Slater determinant corresponds to the irrep $(\lambda, \mu) = (0, 0)$ of SU(3) so that the irrep of U(3), characterized by the eigenvalues of C_{11}, C_{22}, C_{33} of (1.2), can be denoted by $(n/2 + \omega, n/2 + \omega, n/2 + \omega)$, where 3ω is the total number of quanta for the particles in an harmonic oscillator potential. Denoting by

$$|\omega\rangle \equiv |n/2 + \omega, n/2 + \omega, n/2 + \omega\rangle, \tag{1.5}$$

the Slater determinant for closed shells, we see from its construction that it satisfies

$$\begin{aligned} &B_{ij}|\omega\rangle = 0, \\ &C_{ij}|\omega\rangle = 0 \quad \text{if } i \neq j, \\ &C_{ii}|\omega\rangle = (n/2 + \omega)|\omega\rangle. \end{aligned} \tag{1.6}$$

[a)] Member of the Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

 0022-2488/84/051211-08$02.50

The first relation stems from the fact that as B_{ij} is symmetric under permutation of the particles, $B_{ij}|\omega\rangle \neq 0$ would imply that we can construct a state of two quanta less that still satisfies the Pauli principle. This is in contradiction with our assumption that all levels of the oscillator are compactly filled. The second relation stems from the fact that $C_{ij}|\omega\rangle$, $i \neq j$, implies diminishing by 1 the number of quanta in the direction j for some single particle states in the Slater determinant and increasing it by 1 in the direction i; but as all states have been filled this leads to $C_{ij}|\omega\rangle = 0$ if $i \neq j$. Finally the last relation in (1.6) was already derived above.

As B_{ij}^{+}, C_{12}, C_{13}, C_{23} are raising, C_{11}, C_{22}, C_{33} are weight, and B_{ij}, C_{21}, C_{32}, C_{31} are lowering generators of Sp(6), we see from (1.6) that $|\omega\rangle$ is the lowest weight state of Sp(6) and thus

$$\{\omega\} \equiv (n/2 + \omega, n/2 + \omega, n/2 + \omega) \tag{1.7}$$

gives also the irrep of this group. Furthermore, as we indicated above, $|\omega\rangle$ belongs to the irrep (0,0) of SU(3). Thus we can get all the states characterized by the irreps of the chain $\mathrm{Sp}(6) \supset \mathrm{U}(3)$ if we consider

$$P(B_{ij}^{+})|\omega\rangle \tag{1.8}$$

and the polynomial P is characterized by a definite irrep of U(3) and its subgroups. In the next section we proceed to discuss this polynomial.

2. THE POLYNOMIAL IN B_{ij}^{+} AND THE STATES OF THE PROBLEM

From the commutation relations (1.3c) the six raising operators B_{ij}^{+} correspond to the irrep $(\lambda\mu) = (2,0)$ of SU(3). If we take an homogeneous polynomial of degree N in the B_{ij}^{+}, it will be part of a reducible representation of U(3) associated with the direct product of N irreps (20). To get a polynomial of highest weight corresponding to a definite irrep of U(3), the polynomial P in the B_{ij}^{+} must satisfy the commutation relations[6]

$$[C_{ij}, P] = 0 \quad \text{if } i<j, \quad i,j = 1, 2, 3, \tag{2.1a}$$

$$[C_{ii} - n/2, P] = 2h_i P, \quad i = 1, 2, 3, \tag{2.1b}$$

where it is clear that eigenvalues of $C_{ii} - n/2$ are even as the B_{ij}^{+} are second-order polynomials in the creation operators η_{is}. Furthermore, the operator

$$2\hat{N} \equiv \sum_{i=1}^{3} C_{ii} - \frac{3n}{2}, \tag{2.2}$$

is associated with the quantum number $2N$ given by

$$h_1 + h_2 + h_3 = N, \quad h_1 \geqslant h_2 \geqslant h_3 \geqslant 0. \tag{2.3}$$

It can be immediately seen that the polynomials satisfying (2.1), which we denote by $P_{h_1h_2h_3}$, have the form

$$P_{h_1,h_2,h_3} = (B_{11}^{+})^{h_1 - h_2}(B_{11}^{+}B_{22}^{+} - B_{12}^{+}B_{21}^{+})^{h_2 - h_3}(\Delta^{+})^{h_3}, \tag{2.4}$$

where

$$\Delta^{+} = \det \begin{vmatrix} B_{11}^{+} & B_{12}^{+} & B_{13}^{+} \\ B_{21}^{+} & B_{22}^{+} & B_{23}^{+} \\ B_{31}^{+} & B_{32}^{+} & B_{33}^{+} \end{vmatrix}. \tag{2.5}$$

The commutation relations (2.1a) are satisfied as from (1.3c) we see that for $i<j$ we get

$$[C_{ij}, B_{11}^{+}] = [C_{ij}, B_{11}^{+}B_{22}^{+} - B_{12}^{+}B_{21}^{+}] = [C_{ij}, \Delta^{+}] = 0. \tag{2.6}$$

From the commutation relations $[\xi_{is}, \eta_{jt}] = \delta_{ij}\delta_{st}$ we notice that we can interpret $\xi_{is} = \partial/\partial\eta_{is}$ so that application of $C_{ii} - n/2$ to $P_{h_1h_2h_3}$ provides immediately the eigenvalues on the right-hand side of (2.1b).

The fact that $P_{h_1h_2h_3}(B_{ij}^{+})$ satisfies the commutation rules (2.1) implies that it can be represented by a Gel'fand pattern[7]

$$P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ & 2h_1 & 2h_2 \\ & & 2h_1 \end{pmatrix} = N(\omega; h_1h_2h_3)P_{h_1h_2h_3}(B_{ij}^{+}), \tag{2.7}$$

corresponding to a highest weight state,[6] where the weights are given by the eigenvalues of $C_{ii} - n/2$. We have introduced in the definition of this Gel'fand pattern polynomial a normalization constant $N(\omega; h_1h_2h_3)$ determined by the requirement that

$$N^2(\omega; h_1h_2h_3)\langle\omega|P_{h_1h_2h_3}(B_{ij}) \times P_{h_1h_2h_3}(B_{ij}^{+})|\omega\rangle = 1, \tag{2.8}$$

where $|\omega\rangle$ is defined by (1.5). This normalization constant is obtained in Appendix A using the commutation relations (1.3), and with its help, and of well-known Wigner coefficients of SU(3) and transformation brackets, we shall determine the matrix elements of the generators of Sp(6) for the irrep $\{\omega\}$ in the $\mathrm{Sp}(6) \supset \mathrm{U}(3)$ basis.

The most general polynomial associated with the group chain $\mathrm{U}(3) \supset \mathrm{U}(2) \supset \mathrm{U}(1)$ can then be represented by

$$P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ & q_1 & q_2 \\ & & r_1 \end{pmatrix}, \tag{2.9}$$

where the irreps of the three groups in the chain are given by

$$[2h_1, 2h_2, 2h_3], \quad [q_1, q_2], \quad [r_1], \tag{2.10}$$

which satisfy the inequalities

$$2h_1 \geqslant q_1 \geqslant 2h_2 \geqslant q_2 \geqslant 2h_3, \quad q_1 \geqslant r_1 \geqslant q_2. \tag{2.11}$$

As indicated on p. 24 of Ref. 6, the polynomial (2.9) can be obtained from (2.7), (2.4), when we apply lowering operators that are functions of C_{ii} and C_{ij}, $i>j$, $i,j = 1, 2, 3$. We do not give here the explicit procedure as we shall only need the highest weight polynomial (2.7) for the calculation, in the next section, of the matrix elements of B_{ij}^{+}. We note, though, that, from a physical standpoint, we require polynomials characterized by irreps of the O(3) subgroup of U(3) whose generators L_i are given by (1.4). This implies the need of the transformation brackets relating polynomials in the $\mathrm{U}(3) \supset \mathrm{U}(2) \supset \mathrm{U}(1)$ chain with those in $\mathrm{U}(3) \supset \mathrm{O}(3) \supset \mathrm{O}(2)$, i.e.,

$$P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ \Omega & LM & \end{pmatrix} = \sum_{q_1,q_2,r_1} P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ & q_1 & q_2 \\ & & r_1 \end{pmatrix} \times \left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & q_1 & q_2 \\ & & r_1 \end{matrix} \middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & \Omega & LM \end{matrix} \right\rangle, \tag{2.12}$$

where L and M characterize the irrep of O(3) and O(2) (i.e.,

the total angular momentum of the state and its projection) and Ω is a multiplicity index that distinguishes between repeated irreps L of O(3) in a given irrep $[2h_1, 2h_2, 2h_3]$ of U(3). The transformation brackets in (2.12) have been explicitly determined in other publications.[8,9]

The above discussion indicates that we have at our disposal the states that we require, i.e.,

$$|\omega; N(\lambda\mu)\Omega\, LM\rangle = P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ & \Omega\, LM & \end{pmatrix}|\omega\rangle, \tag{2.13}$$

where N is given by (2.3) and (λ, μ) are

$$\lambda = 2h_1 - 2h_2, \quad \mu = 2h_2 - 2h_3. \tag{2.14}$$

Our objective then is to find the matrix elements of the generators (1.2) of Sp(6) with respect to the states (2.13), which we shall implement in the next section.

3. THE MATRIX ELEMENTS OF THE GENERATORS OF Sp(6)

We start with the generators C_{ij} of the subgroup U(3) of Sp(6), which we separate in $\hat{N}$ given by (2.2) and a traceless part $\overline{C}_{ij}$ defined by

$$\overline{C}_{ij} = C_{ij} - \tfrac{2}{3}\hat{N}\delta_{ij}. \tag{3.1}$$

The matrix element of $\hat{N}$ with respect to the states (2.13) is, of course, of value N and diagonal in all the quantum numbers appearing there if Ω is chosen as the eigenvalue of an appropriate Hermitian operator.[10] For the $\overline{C}_{ij}$ we get the matrix elements

$$\begin{aligned}\langle \omega; N(\lambda\mu)\Omega' L' M'|\overline{C}_{ij}|\omega; N(\lambda\mu)\Omega LM\rangle &= \sum_{q_1 q_2 r_1}\sum_{q_1' q_2' r_1'}\left\{\left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{matrix}\middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & \Omega\, LM & \end{matrix}\right\rangle \right. \\ &\times\left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1' & q_2' & \\ & r_1' & \end{matrix}\middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & \Omega\, LM & \end{matrix}\right\rangle \\ &\times\left.\left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1' & q_2' & \\ & r_1' & \end{matrix}\middle| \overline{C}_{ij} \middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{matrix}\right\rangle\right\},\end{aligned} \tag{3.2}$$

where we made use of the fact that the transformation brackets in (2.12) are real and so we can invert the order of bra and ket.

The matrix element of $\overline{C}_{ij}$ for kets in the basis U(3)⊃U(2)⊃U(1) are given explicitly, and in a very simple form, on p. 26 of Ref. 6. The transformation brackets from the U(3)⊃U(2)⊃U(1) to U(3)⊃O(3)⊃O(2) chain are, as was mentioned above, also available explicitly. Thus we can elaborate tables for all the matrices of $\overline{C}_{ij}$ in the basis (2.13), which we may require in explicit calculations.

Turning now our attention to the generators B_{ij}^{+} and B_{ij} of Sp(6) in (1.2), we notice that B_{ij} is the Hermitian conjugate of B_{ij}^{+}. Thus we need only the matrix elements of B_{ij}^{+} with respect to the states (2.13). We can make then the same development as in (3.2), keeping in mind that now in the bra we have $N+1(\lambda'\mu')$ so that what remains to be determined are the matrix elements

$$\left\langle \begin{matrix} 2h_1' & 2h_2' & 2h_3' \\ q_1' & q_2' & \\ & r_1' & \end{matrix}\middle| B_{ij}^{+} \middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{matrix}\right\rangle, \tag{3.3}$$

where the ket is defined by

$$\left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{matrix}\right\rangle = P\begin{pmatrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{pmatrix}|\omega\rangle, \tag{3.4}$$

and similarly for the bra.

As was mentioned before, B_{ij}^{+} is associated with the irrep [200] of U(3) or, equivalently, $(\lambda\mu) = (2, 0)$ of SU(3). The bra and ket are associated with the irreps $[2h_1' 2h_2' 2h_3']$ and $[2h_1, 2h_2, 2h_3]$ of U(3) in the U(3)⊃U(2)⊃U(1) chain, and thus we see that in the bra we have only the irreps[6]

$$\begin{aligned}&[2h_1', 2h_2', 2h_3'] = [2h_1 + 2, 2h_2, 2h_3], \\ &[2h_1, 2h_2 + 2, 2h_3], [2h_1, 2h_2, 2h_3 + 2].\end{aligned} \tag{3.5}$$

Furthermore, with the help of the Wigner–Eckart theorem, the matrix element (3.3) can be written as

$$\begin{aligned}&\left\langle \begin{matrix} 2h_1' & 2h_2' & 2h_3' \\ q_1' & q_2' & \\ & r_1' & \end{matrix}\middle| \beta_{st}^{+} \middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ q_1 & q_2 & \\ & r_1 & \end{matrix}\right\rangle \\ &= \langle 2h_1' 2h_2' 2h_3'\|B^{+}\|2h_1 2h_2 2h_3\rangle \\ &\times\left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 & 2 & 0 & 0 \\ q_1 & q_2 & ; & s & 0 & \\ & r_1 & & & t & \end{matrix}\middle| \begin{matrix} 2h_1' & 2h_2' & 2h_3' \\ q_1' & q_2' & \\ & r_1' & \end{matrix}\right\rangle,\end{aligned} \tag{3.6}$$

where the last term of the right-hand side is a well-known Wigner coefficient[11,12] of U(3) in Gel'fand–Zetlin notation, which is discussed in Appendix B. In (3.6) the notation β_{st}^{+} implies that we are using a Gel'fand–Zetlin pattern, associated with the irrep [200] of U(3), for the B_{ij}^{+}, i.e.,

$$\beta_{st}^{+} = P\begin{pmatrix} 2 & 0 & 0 \\ & s & 0 \\ & & t \end{pmatrix}, \tag{3.7}$$

where $s, t = 2, 1, 0$ with $t \leqslant s$. The explicit relation between B_{ij}^{+} and β_{st}^{+} is the following[6]:

$$\begin{aligned}&\beta_{22}^{+} = B_{11}^{+}, \quad \beta_{21}^{+} = \sqrt{2}B_{21}^{+}, \quad \beta_{20}^{+} = B_{22}^{+}, \\ &\beta_{11}^{+} = \sqrt{2}B_{31}^{+}, \quad \beta_{10}^{+} = \sqrt{2}B_{32}^{+}, \quad \beta_{00}^{+} = B_{33}^{+}.\end{aligned} \tag{3.8}$$

Thus for determining the matrix element (3.3) we need only to have the reduced matrix element of B^{+} indicated on the right-hand side of (3.6).

To obtain the reduced matrix element of B_{ij}^{+}, we can take for the ket on the right-hand side of Eq. (3.6) the highest weight state, i.e., $q_1 = r_1 = 2h_1, q_2 = 2h_2$. If we then consider $\beta_{22}^{+} = B_{11}^{+}$ and take for the bra the highest weight state corresponding to the irrep $[2h_1', 2h_2', 2h_3']$ with $h_1' = h_1 + 1$, $h_2' = h_2, h_3' = h_3$ we get the relation

$$\left\langle \begin{matrix} 2h_1 + 2 & 2h_2 & 2h_3 \\ 2h_1 + 2 & 2h_2 & \\ & 2h_1 + 2 & \end{matrix}\middle| B_{11}^{+} \middle| \begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ 2h_1 & 2h_2 & \\ & 2h_1 & \end{matrix}\right\rangle$$

$$= \langle 2h_1+2\ 2h_2\ 2h_3 \| B^+ \| 2h_1\ 2h_2\ 2h_3 \rangle \times \left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 & 2 & 0 & 0 \\ 2h_1 & 2h_2 & ; & 2 & 0 & \end{matrix} \middle| \begin{matrix} 2h_1+2 & 2h_2 & 2h_3 \\ 2h_1+2 & 2h_2 & \\ 2h_1+2 & & \end{matrix} \right\rangle. \quad (3.9)$$

But from the expressions (3.4), (2.7), (2.4) for bra and ket on the left-hand side of (3.9), we see that the matrix element indicated there takes the value

$$N(\omega; h_1h_2h_3)/N(\omega; h_1+1\ h_2\ h_3), \quad (3.10)$$

where the normalization coefficients are given explicitly in Appendix A. As the Wigner coefficient of U(3) on the right-hand side of (3.9), discussed in Appendix B, is a stretched one of value 1, we obtain that

$$\langle 2h_1+2\ 2h_2\ 2h_3 \| B^+ \| 2h_1\ 2h_2\ 2h_3 \rangle = \left[\frac{(2h_1+4)(2h_1-2h_3+3)(2h_1-2h_2+2)(2h_1+2\omega+n)}{(2h_1-2h_3+4)(2h_1-2h_2+3)} \right]^{1/2}. \quad (3.11a)$$

A similar analysis, but now with B^+_{22}, B^+_{33}, which is carried out explicitly in Appendix C, gives us then

$$\langle 2h_1\ 2h_2+2\ 2h_3 \| B^+ \| 2h_1\ 2h_2\ 2h_3 \rangle = \left[\frac{(2h_2+3)(2h_2-2h_3+2)(2h_1-2h_2)(2h_2+2\omega+n-1)}{(2h_2-2h_3+3)(2h_1-2h_2-1)} \right]^{1/2}, \quad (3.11b)$$

$$\langle 2h_1\ 2h_2\ 2h_3+2 \| B^+ \| 2h_1\ 2h_2\ 2h_3 \rangle = \left[\frac{(2h_3+2)(2h_2-2h_3)(2h_1-2h_3+1)(2h_3+2\omega+n-2)}{(2h_1-2h_3)(2h_2-2h_3-1)} \right]^{1/2}. \quad (3.11c)$$

Having the explicit expressions (3.11) of the reduced matrix elements, the Wigner coefficients of U(3) discussed in Appendix B, and the transformation brackets from the $U(3) \supset U(2) \supset U(1)$ to the $U(3) \supset O(3) \supset O(2)$ chain of groups discussed in Refs. 8 and 9, we can then write the matrix elements of the generators of Sp(6) with respect to the states (2.13). More practically, we plan to use these results to elaborate computer programs for the matrix elements of all the generators of Sp(6) for states in an $Sp(6) \supset U(3) \supset O(3)$ basis.

We note that so far we have only given an analytic procedure for calculating the matrix elements of the generators B^+_{ij} of Sp(6) in the $Sp(6) \supset U(3)$ basis when the irrep of Sp(6) is the $\{\omega\}$ of (1.7). We would like to obtain these matrix elements for the general irrep of Sp(6)

$$(n/2+\omega_3,\ n/2+\omega_2,\ n/2+\omega_1), \quad (3.12)$$

which, as we indicated in Sec. 1, are obtained from polynomials of the type (2.9) coupled with U(3) Wigner coefficients to Slater determinant associated with open shells, i.e., characterized by an irrep $[\omega_1, \omega_2, \omega_3]$ of U(3). The problem of normalizing these states [i.e., the equivalent of the $N(\omega, h_1h_2h_3)$ for closed shells discussed in Appendix A] becomes then much more difficult. If it could be solved analytically, then the matrix elements of B^+_{ij} in the $Sp(6) \supset U(3)$ basis can be obtained with the help of SU(3) Racah coefficients as discussed by Rosensteel.[3] We plan to look into this problem of normalization in a future publication.

We would like to indicate that Klimyk[13] has also obtained the matrix elements of the generators of Sp(6) but in representations that are not the ones required in the nuclear structure problems discussed in this note[14]. An alternative derivation of our results can also be obtained through the analysis in which Deenen and Quesne[15] consider the boson representation of the dynamical group of microscopic collective states. Finally we stress that our discussion concerns the noncompact symplectic group Sp(6), which is frequently denoted in the literature by Sp(6, R) or Sp(3, R), and not the compact one for which sometimes the notation $Sp(6, C) \cap U(6)$ is used.

APPENDIX A: THE NORMALIZATION COEFFICIENTS $N(\omega; h_1h_2h_3)$

In this appendix we shall determine the value of the normalization coefficient $N(\omega; h_1h_2h_3)$ appearing in Eq. (2.7).

First of all we use two facts:

(i) The determinant of the B_{ij} can be written as

$$\Delta = \sum_{k=1}^{3} \Delta_{k3} B_{k3} \quad (A1a)$$

with

$$\Delta_{k3} \equiv \sum_{ij} \epsilon_{ijk} B_{i1} B_{j2}, \quad (A1b)$$

where ϵ_{ijk} is the completely antisymmetric tensor.

(ii) The relation

$$B_{k3} P_{h_1h_2h_3}(B^+_{ij})|\omega\rangle = \alpha(h_3) \Delta^+_{k3} P_{h_1-1,h_2-1,h_3-1}(B^+_{ij})|\omega\rangle, \quad (A2a)$$

where

$$\alpha(h_3) = 2h_3(2h_3+2\omega+n-4) \quad (A2b)$$

and Δ^+_{k3} is the Hermitian conjugate of the Δ_{k3} of Eq. (A1b). Relation (A2a) can be obtained straightforwardly first for $k=3$ using the commutation relations (1.3), and then for $k=1, 2$ using $B_{k3} = -\frac{1}{2}[C_{k3}, B_{33}]$.

With these two results Eq. (2.8) can be written in the form

$$N^{-2}(\omega; h_1h_2h_3) = \sum_{k=1}^{3} \alpha(h_3) \langle \omega | P_{h_1-1,h_2-1,h_3-1}(B_{ij}) \times \Delta_{k3} \Delta^+_{k3} P_{h_1-1,h_2-1,h_3-1}(B^+_{ij})|\omega\rangle. \quad (A3)$$

Using (1.3), we find the following relations:

$$[C_{3k}, \Delta^+_{33}] = -2\Delta^+_{k3}, \quad k=1,2, \quad (A4a)$$

$$[C_{k3}, \Delta^+_{k3}] = -\Delta^+_{33}, \quad k=1,2. \quad (A4b)$$

It follows then from (A4a) that Δ^+_{k3} acting on the state

$P_{h_1-1,h_2-1,h_3-1}(B_{ij}^{+})|\omega\rangle$ transforms it into the combination

$$\Delta_{k3}^{+}P_{h_1-1,h_2-1,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$=\delta_{k3}P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$-\delta_{k2}\frac{C_{32}}{2(h_2-h_3+1)}P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$-\delta_{k1}\frac{C_{31}}{2(h_2-h_3+1)}P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$+\delta_{k1}\frac{(h_1-h_2)}{(h_2-h_3+1)}B_{13}^{+}P_{h_1-1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle, \tag{A5}$$

as is obvious if we take separately the cases $k=1, 2, 3$.

Substituting the Hermitian conjugate of Eq. (A5) in (A3), we get

$$N^{-2}(\omega; h_1h_2h_3)=\alpha(h_3)\langle\omega|P_{h_1,h_2,h_3-1}(B_{ij})$$
$$\times\left\{\Delta_{33}^{+}-\frac{C_{23}\Delta_{23}^{+}}{2(h_2-h_3+1)}-\frac{C_{13}\Delta_{13}^{+}}{2(h_2-h_3+1)}\right\}P_{h_1-1,h_2-1,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$+\alpha(h_3)\frac{h_1-h_2}{h_2-h_3+1}\langle\omega|P_{h_1-1,h_2,h_3-1}(B_{ij})B_{13}\Delta_{13}^{+}P_{h_1-1,h_2-1,h_3-1}(B_{ij}^{+})|\omega\rangle. \tag{A6}$$

Now on the right-hand side of (A6) we can replace $C_{k3}\Delta_{k3}^{+}$, $k=1, 2$ by the commutator $[C_{k3}, \Delta_{k3}^{+}]$ given in (A4b), thus obtaining

$$N^{-2}(\omega; h_1h_2h_3)=\alpha(h_3)\left[1+\frac{1}{(h_2-h_3+1)}\right]N^{-2}(\omega; h_1h_2, h_3-1)$$
$$+\alpha(h_3)\langle\omega|P_{h_1-1,h_2,h_3-1}(B_{ij})B_{13}\Delta_{13}^{+}P_{h_1-1,h_2-1,h_3-1}(B_{ij}^{+})|\omega\rangle. \tag{A7}$$

In the last term of (A7) we use now Eq. (A5) with $k=1$ and the Hermitian conjugate of the expression

$$C_{13}B_{13}^{+}P_{h_1-1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle=P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle. \tag{A8}$$

In this way we obtain

$$N^{-2}(\omega; h_1h_2h_3)=\alpha(h_3)\left[1-\frac{h_1-h_2}{2(h_2-h_3+1)^2}+\frac{1}{h_2-h_3+1}\right]N^{-2}(\omega; h_1,h_2, h_3-1)$$
$$+\alpha(h_3)\left(\frac{h_1-h_2}{h_2-h_3+1}\right)^2\langle\omega|P_{h_1-1,h_2,h_3-1}(B_{ij})B_{13}B_{13}^{+}P_{h_1-1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle. \tag{A9}$$

In the last term of (A9) the action of B_{13}^{+} on the state $P_{h_1-1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$ can be related to the action of the lowering operator of the unitary group U(3), $L_{31}\equiv C_{31}(C_{11}-C_{22}+1)+C_{21}C_{32}$ on the state $P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$. Thus we have the result

$$\alpha(h_3)\left(\frac{h_1-h_2}{h_2-h_3+1}\right)^2\langle\omega|P_{h_1-1,h_2,h_3-1}(B_{ij})B_{13}B_{13}^{+}P_{h_1-1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle$$
$$=\frac{\alpha(h_3)}{4(2h_1-2h_3+3)^2(h_2-h_3+1)^2}\langle\omega|[L_{31}P_{h_1,h_2,h_3-1}(B_{ij}^{+})]^{+}L_{31}P_{h_1,h_2,h_3-1}(B_{ij}^{+})|\omega\rangle. \tag{A10}$$

The normalization coefficient for the raising and lowering operators of U(3) were obtained in Ref. 6, and for L_{31} appearing in (A10) it is given by

$$[(2h_1-2h_2+1)(2h_1-2h_2+3)(2h_1-2h_2)]^{-1/2}. \tag{A11}$$

Substituting (A10) and (A11) in (A9), we finally arrive at the recurrence relation

$$N^{-2}(\omega; h_1h_2h_3)=\alpha(h_3)\frac{(h_1-h_3+2)(2h_2-2h_3+3)}{(2h_1-2h_3+3)(h_2-h_3+1)}N^{-2}(\omega; h_1h_2, h_3-1). \tag{A12}$$

This is satisfied by

$$N^{-2}(\omega; h_1h_2h_3)=(2h_3)!!\frac{(2h_3+2\omega+n-4)!!}{(2\omega+n-4)!!}\frac{(2h_2+1)!!}{(2h_2-2h_3+1)!!}\frac{(h_1+1)!}{(h_1-h_3+1)!}\frac{(2h_1-2h_3+1)!!}{(2h_1+1)!!}\frac{(h_2-h_3)!}{h_2!}N^{-2}(\omega; h_1h_20). \tag{A13}$$

Then it remains to find the normalization coefficient $N(\omega; h_1h_20)$ for the polynomial

$$P_{h_1h_20}(B_{ij}^{+})|\omega\rangle=(B_{11}^{+})^{h_1-h_2}(\Delta_{33}^{+})^{h_2}|\omega\rangle, \tag{A14}$$

which is the highest weight state of the irrep $[h_1, h_2]$ of the unitary group U(2). To this end, we begin considering the relation analogous to (A2), i.e.,

$$B_{22}P_{h_1h_20}(B_{ij}^{+})|\omega\rangle=\beta(h_2)B_{11}^{+}P_{h_1-1,h_2-1,0}(B_{ij}^{+})|\omega\rangle, \tag{A15a}$$

$$B_{12}P_{h_1h_20}(B_{ij}^{+})|\omega\rangle=-\beta(h_2)B_{12}^{+}P_{h_1-1,h_2-1,0}(B_{ij}^{+})|\omega\rangle, \tag{A15b}$$

where

$$\beta(h_2) = 2h_2(2h_2 + 2\omega + n - 3). \tag{A15c}$$

These results can be established by a method similar to that followed to obtain (A2).

In order to obtain $N^{-2}(\omega; h_1 h_2 0)$, we start by writing $P_{h_1 h_2 0}(B_{ij})$ as

$$\Delta_{33} P_{h_1-1,h_2-1,0}(B_{ij}) \equiv (B_{11}B_{22} - B_{12}B_{21}) P_{h_1-1,h_2-1,0}(B_{ij})$$

and apply B_{22} and one of the B_{12} on the state $P_{h_1 h_2 0}(B^{+}_{ij})|\omega\rangle$ using Eqs. (A15a) and (A15b). In this way we arrive at

$$N^{-2}(\omega; h_1 h_2 0) = \beta(h_2)\{N^{-2}(\omega; h_1, h_2 - 1, 0) + \langle\omega| B_{12} P_{h_1-1,h_2-1,0}(B_{ij}) B^{+}_{12} P_{h_1-1,h_2-1,0}(B^{+}_{ij})|\omega\rangle\}. \tag{A16}$$

In the last term, we again relate the action of B^{+}_{12} on the state $P_{h_1-1,h_2-1,0}(B^{+}_{ij})|\omega\rangle$ with the action of the lowering operator C_{21} of SU(2) on the state $P_{h_1,h_2-1,0}(B^{+}_{ij})|\omega\rangle$.

Thus Eq. (A16) can be written as follows:

$$N^{-2}(\omega; h_1 h_2 0) = \beta(h_2)\left\{N^{-2}(\omega; h_1, h_2 - 1, 0) + \frac{1}{4(h_1 - h_2 + 1)^2}\langle\omega| P_{h_1,h_2-1,0}(B_{ij}) C_{12} C_{21} P_{h_1,h_2-1,0}(B^{+}_{ij})|\omega\rangle\right\}, \tag{A17}$$

from which we can immediately deduce that

$$N^{-2}(\omega; h_1 h_2 0) = \beta(h_2)\frac{2h_1 - 2h_2 + 3}{2(h_1 - h_2 + 1)} N^{-2}(\omega; h_1, h_2 - 1, 0). \tag{A18}$$

This recurrence relation implies that

$$N^{-2}(\omega; h_1 h_2 0) = \frac{(2h_2)!!(2h_2 + 2\omega + n - 3)!!}{(2h_1)!!(2\omega + n - 3)!!}\frac{(2h_1 + 1)!!}{(2h_1 - 2h_2 + 1)!!}(2h_1 - 2h_2)!! N^{-2}(\omega; h_1, 00). \tag{A19}$$

Now we have to find the constant $N(\omega; h_1 00)$ which corresponds to the normalization coefficient of a highest weight state of the unitary group U(1), i.e.,

$$N^{-2}(\omega; h_1 00) = \langle\omega|(B_{11})^{h_1}(B^{+}_{11})^{h_1}|\omega\rangle. \tag{A20}$$

It is straightforward to see that

$$N^{-2}(\omega; h_1 00) = (2h_1)!!\frac{(2h_1 + 2\omega + n - 2)!!}{(2\omega + n - 2)!!}, \tag{A21}$$

where we define $N^{-2}(\omega; 000) = \langle\omega|\omega\rangle = 1$.

Putting together the expressions (A13), (A19), and (A21), we obtain the normalization coefficient of the state (2.7) with highest weight in the unitary group U(3), namely,

$$N^{-2}(\omega; h_1 h_2 h_3) = \frac{(2h_3 + 2\omega + n - 4)!!(2h_2 + 2\omega + n - 3)!!}{(2\omega + n - 4)!!(2\omega + n - 3)!!} \times \frac{(2h_1 + 2\omega + n - 2)!!(h_1 + 1)!(2h_1 - 2h_3 + 1)!!(2h_1 - 2h_2)!!(2h_3)!!}{(2\omega + n - 2)!!(h_1 - h_3 + 1)!(2h_2 - 2h_3 + 1)!!(2h_1 - 2h_2 + 1)!!}\frac{(2h_2 + 1)!(h_2 - h_3)!}{h_2!}. \tag{A22}$$

APPENDIX B: WIGNER COEFFICIENTS FOR U(3) IN THE GEL'FAND–ZETLIN NOTATION

It is well known that the U(3) Wigner coefficients needed in our analysis can be factorized as follows:

$$\left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 & 2 & 0 & 0 \\ q_1 & q_2 & & ; & q & 0 \\ r_1 & & & & r & \end{matrix} \middle| \begin{matrix} 2h'_1 & 2h'_2 & 2h'_3 \\ q'_1 & q'_2 & \\ r' & & \end{matrix} \right\rangle \times \left\langle \begin{matrix} 2h_1 & 2h_2 & 2h_3 & 2 & 0 & 0 \\ q_1 & q_2 & & ; & q & 0 \end{matrix} \middle| \begin{matrix} 2h'_1 & 2h'_2 & 2h'_3 \\ q'_1 & q'_2 & \end{matrix} \right\rangle$$

$$= \langle \tfrac{1}{2}(q_1 - q_2), r_1 - \tfrac{1}{2}(q_1 + q_2); q/2, r - q/2 | \tfrac{1}{2}(q'_1 - q'_2), r'_1 - \tfrac{1}{2}(q'_1 + q'_2)\rangle, \tag{B1}$$

where the first factor on the right-hand side is a U(3) reduced Wigner coefficient or isoscalar factor and the second one is a standard SU(2) Wigner coefficient.

The coefficient $\langle \| \rangle$ can be obtained from the work of Hecht,[11] taking into account that he uses basis states classified by the quantum numbers associated with the chain of groups

$$\underset{(\lambda,\mu)}{SU(3)} \supset \underset{\epsilon}{U(1)} \times \underset{\Lambda}{SU(2)} \supset \underset{\frac{1}{2}\nu}{O(2)}, \tag{B2}$$

where underneath each group we write down the quantum numbers characterizing its irreducible representation (irreps). For U(3) we have the partition $[2h_1, 2h_2, 2h_3]$, where $2h_1 + 2h_2 + 2h_3 = 2N$ is the number of quanta; $\lambda = 2h_1 - 2h_2, \mu = 2h_2 - 2h_3$ gives the irrep of SU(3), ϵ the irrep of U(1), and Λ, $\frac{1}{2}\nu$ the irreps of SU(2) and O(2), respectively.

Fixing the irrep (λ, μ) of SU(3) the possible values of ϵ and Λ are given by

$$\epsilon = 2\lambda + \mu - 3(t + s), \tag{B3a}$$

$$\Lambda = \tfrac{1}{2}\mu + \tfrac{1}{2}(t-s), \tag{B3b}$$

where the integers t and s range over the values $t = 0, 1, ..., \lambda$ and $s = 0, 1, ..., \mu$.

We need the relations between the quantum numbers in the Elliott $|N(\lambda,\mu)\epsilon\Lambda\nu\rangle$ and Gel'fand

$$\left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & q_1 \quad q_2 & \\ & r & \end{matrix}\right\rangle$$

basis states. These are

$$\epsilon = 4h_1 + 4h_2 - 4h_3 - 3(q_1+q_2), \tag{B4a}$$

$$\Lambda = \tfrac{1}{2}(q_1 - q_2), \tag{B4b}$$

$$t = q_1 - 2h_2, \tag{B4c}$$

$$s = q_2 - 2h_3. \tag{B4d}$$

From Eq. (3.5) we know that the possible values of the partition $[2h'_1, 2h'_2, 2h'_3]$ are

$$\begin{aligned} &[2h_1+2,\ 2h_2,\ 2h_3], \\ &[2h_1,\ 2h_2+2,\ 2h_3], \\ &[2h_1,\ 2h_2,\ 2h_3+2], \end{aligned} \tag{B5}$$

and from Table 3 of Ref. 11, using the notation (B4), we obtain the U(3) reduced Wigner coefficients in the Gel'fand–Zetlin notation.

We want to remark that according to Ref. 12 the reduced Wigner coefficients tabulated by Hecht were established with a phase convention for the U(3) basis states which is different from the one adopted for the standard Gel'fand states. As a consequence, the values of the SU(3) reduced Wigner coefficients tabulated by Hecht should be multiplied by $(-1)^{q_2+q'_2}$ before using them in Eq. (3.6).

APPENDIX C: REDUCED MATRIX ELEMENTS OF B^+_{ij}

In this appendix we calculate explicitly the reduced matrix elements

$$\langle 2h_1,\ 2h_2+2,\ 2h_3 \| B^+ \| 2h_1\ 2h_2\ 2h_3\rangle \tag{C1a}$$

and

$$\langle 2h_1,\ 2h_2,\ 2h_3+2 \| B^+ \| 2h_1\ 2h_2\ 2h_3\rangle. \tag{C1b}$$

To evaluate (C1a), we apply B^+_{22} to the highest weight state (3.4), so that

$$B^+_{22}\left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right\rangle = B^+_{22} N(\omega; h_1h_2h_3) P_{h_1h_2h_3}(B^+_{ij})|\omega\rangle, \tag{C2}$$

where $|\omega\rangle$ is defined by (1.5), $P_{h_1h_2h_3}$ is given by the Eq. (2.4), and $N(\omega; h_1h_2h_3)$ is their normalization constant determined in Appendix A, Eq. (A22).

From the explicit form of the polynomial $P_{h_1h_2h_3}$, Eq. (2.4), it is easy to get

$$\begin{aligned} B^+_{22} N(\omega; h_1h_2h_3) P_{h_1h_2h_3}(B^+_{ij})|\omega\rangle &= N(\omega; h_1h_2h_3) \\ &\times [P_{h_1,h_2+1,h_3}(B^+_{ij})|\omega\rangle + (B^+_{12})^2 P_{h_1-1,h_2,h_3}(B^+_{ij})|\omega\rangle]. \end{aligned} \tag{C3}$$

In the following we use the relation

$$\begin{aligned} &(B^+_{12})^2 P_{h_1-1,h_2,h_3}(B^+_{ij})|\omega\rangle \\ &\quad = \frac{(C_{21})^2}{4(h_1-h_2)(h_1-h_2+1)} P_{h_1+1,h_2,h_3}(B^+_{ij})|\omega\rangle \\ &\qquad - \frac{1}{2(h_1-h_2)} B^+_{22} P_{h_1h_2h_3}(B^+_{ij})|\omega\rangle, \end{aligned} \tag{C4}$$

which is obtained through the commutation relations of C_{21} with each factor of the polynomial (2.4).

Substituting (C4) in (C3) and taking the scalar product with the Gel'fand state $\left|\begin{matrix} 2h_1 & 2h_2+2 & 2h_3 \\ & 2h_1 \quad 2h_2+2 & \\ & 2h_1 & \end{matrix}\right\rangle$, we obtain

$$\begin{aligned} &\left\langle\begin{matrix} 2h_1 & 2h_2+2 & 2h_3 \\ & 2h_1 \quad 2h_2+2 & \\ & 2h_1 & \end{matrix}\right| B^+_{22} \left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right\rangle \\ &\quad = \frac{2(h_1-h_2)}{2h_1-2h_2+1} \frac{N(\omega; h_1h_2h_3)}{N(\omega; h_1, h_2+1, h_3)}, \end{aligned} \tag{C5}$$

where it is clear the term containing $(C_{21})^2$ does not contribute to the result.

Furthermore, with the help of the Wigner–Eckart theorem, we get the reduced matrix element (C1a), namely,

$$\begin{aligned} &\langle 2h_1\ 2h_2+2\ 2h_3 \| B^+ \| 2h_1\ 2h_2\ 2h_3\rangle \\ &\quad = \frac{2(h_1-h_2)}{2h_1-2h_2+1} \frac{N(\omega; h_1h_2h_3)}{N(\omega; h_1\, h_2+1, h_3)} \\ &\qquad \times \left[\left\langle\begin{matrix} 2h_1 & 2h_2 & 2h_3 & 2\ 0\ 0 \\ & 2h_1 \quad 2h_2 & ;\ 2\ 0 & \\ & 2h_1 & 0 & \end{matrix}\right| \left.\begin{matrix} 2h_1 & 2h_2+2 & 2h_3 \\ & 2h_1 \quad 2h_2+2 & \\ & 2h_1 & \end{matrix}\right\rangle\right]^{-1}. \end{aligned} \tag{C6}$$

Finally the relation (3.11b) is obtained substituting explicitly the values of the U(3) Wigner coefficients and the normalization constants.

To evaluate (C1b), we consider the matrix element

$$\begin{aligned} &\left\langle\begin{matrix} 2h_1 & 2h_2 & 2h_3+2 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right| B^+_{33} \left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right\rangle \\ &\quad = N(\omega; h_1h_2, h_3+1) N(\omega; h_1h_2h_3) \\ &\qquad \times \langle\omega| P_{h_1h_2h_3}(B_{ij}) B_{33} P_{h_1h_2h_3+1}(B^+_{ij})|\omega\rangle. \end{aligned} \tag{C7}$$

Using the commutation relation between B_{33} and Δ^+, the operator given in (2.5), it is straightforward to prove that

$$\begin{aligned} &B_{33} P_{h_1h_2h_3+1}(B^+_{ij})|\omega\rangle \\ &\quad = 2h_3(2h_3+2\omega+n-4) P_{h_1h_2h_3}(B^+_{ij})|\omega\rangle. \end{aligned} \tag{C8}$$

Introducing (C8) in the Eq. (C7), we obtain

$$\begin{aligned} &\left\langle\begin{matrix} 2h_1 & 2h_2 & 2h_3+2 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right| B^+_{33} \left|\begin{matrix} 2h_1 & 2h_2 & 2h_3 \\ & 2h_1 \quad 2h_2 & \\ & 2h_1 & \end{matrix}\right\rangle \\ &\quad = \frac{N(\omega; h_1, h_2, h_3+1)}{N(\omega; h_1, h_2, h_3)} 2h_3(2h_3+2\omega+n-4), \end{aligned} \tag{C9}$$

where the N's can be taken from Eq. (A22).

As in the previous case, using the Wigner–Eckart theorem in (C9), the reduced matrix element can be written as

$$\begin{aligned} &\langle 2h_1\ 2h_2\ 2h_3+2 \| B^+ \| 2h_1\ 2h_2\ 2h_3\rangle \\ &\quad = \frac{2h_3(2h_3+2\omega+n-4) N(\omega; h_1\, h_2\, h_3+1)}{N(\omega; h_1h_2h_3)} \end{aligned}$$

$$\times\left[\left\langle\begin{matrix}2h_1 & 2h_2 & 2h_3 & 2 & 0 & 0\\ & 2h_1 & 2h_2 & ; & 0 & 0\\ & & 2h_1 & & 0 & \end{matrix}\middle|\begin{matrix}2h_1 & 2h_2 & 2h_3+2\\ & 2h_1 & 2h_2\\ & & 2h_1\end{matrix}\right\rangle\right]^{-1}. \tag{C10}$$

At last, the expression (3.11c) is obtained if we put the explicit values of the U(3) Wigner coefficients and the normalization constants.

[1]G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **126**, 198, 343 (1980).
[2]G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979); V. C. Vasilevsky, Ya. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980).
[3]G. Rosensteel, J. Math. Phys. **21**, 924 (1980).
[4]O. Castaños, E. Chacón, A. Frank, P. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[5]J. P. Elliott, Proc. Roy. Soc. London A **245**, 128 (1958).
[6]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1969), p. 14.
[7]I. M. Gel'fand and M. Zetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).
[8]M. Moshinsky, Rev. Mod. Phys. **38**, 813 (1962); M. Moshinsky and V. Syamala Devi, J. Math. Phys. **10**, 455 (1968).
[9]Appendix B of Ref. 4.
[10]V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961).
[11]K. T. Hecht, Nucl. Phys. **62**, 1 (1965).
[12]C. Quesne, J. Math. Phys. **22**, 1482 (1981).
[13]A. V. Klimyk, J. Math. Phys. **24**, 224 (1983).
[14]G. Rosensteel and D. J. Rowe, J. Math. Phys. **24**, 2461 (1983).
[15]J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982); also C. Quesne, private communication.

A pair of commuting missing label operators for $G \supset [SU(2)]^n$

Joris Van der Jeugt[a)]
Seminarie voor Wiskundige Natuurkunde, R.U.G., Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 4 April 1983; accepted for publication 26 August 1983)

A general procedure, which can lead to a pair of commuting subgroup scalars in the enveloping algebra of a Lie group G, decomposed in $[SU(2)]^n$, is discussed. The technique is illustrated by means of three explicit examples.

PACS numbers: 02.20.Qs

1. INTRODUCTION

The complete classification or labeling of the basis states of an irreducible representation (irrep) of a classical Lie group G, decomposed into irreps of some subgroup H, very often leads to a missing label problem. One of the various ways of solving this problem consists of obtaining, within the enveloping algebra of G, a complete set of commuting Hermitian operators by adding to the Casimir operators of the group G, of the subgroup H, and of appropriate subgroups of H, scalar operators with respect to H.[1,2] The common eigenstates of this complete set are then chosen as the basis states of an irrep of G, and the eigenvalues of the additional operators provide the missing labels.

The study of the subgroup scalars leads to the construction of an integrity basis.[2,3] In cases where there is one missing label, any polynomial in the elements of the integrity basis may be chosen as missing label operator. When the number of missing labels is greater than one, there is a supplementary problem: the missing label operators, constructed out of the elements of the integrity basis, have to commute mutually. The only group-subgroup chain for which the latter problem has been solved explicitly, is the case $SU(4) \supset SU(2) \otimes SU(2)$: Moshinsky and Nagel have constructed the pair Ω and Φ,[4] and Quesne,[5] Partensky, and Maguin[6] the pair $C^{(20;2)}$ and $C^{(02;2)}$.

In the present paper we exhibit a possible procedure for finding a pair of mutually commuting subgroup scalars in the case $G \supset [SU(2)]^n = SU(2) \otimes \cdots \otimes SU(2)$ (n products) if $n \geqslant 2$. Section 2 contains some sufficient conditions which must be fulfilled in order to apply our technique. In Sec. 3, a few examples are listed.

2. GENERAL TECHNIQUE

If a Lie group G contains $[SU(2)]^n$, its generators may be chosen as the generators of the subgroup, namely ${}^i s_{0,\pm 1}$ ($i = 1,2,...,n$), and the components $Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}$ of some irreducible tensor operators with respect to $[SU(2)]^n$. The following commutation relations are valid:

$$[{}^i s_\mu, {}^j s_\nu] = 0 \quad \text{if} \quad i \neq j, \tag{2.1}$$

$$[{}^i s_0, {}^i s_{\pm 1}] = \pm {}^i s_{\pm 1}; \quad [{}^i s_{-1}, {}^i s_{+1}] = {}^i s_0, \tag{2.2}$$

$$\left[{}^i s_0, Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}\right] = \lambda_i Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n},$$

$$\left[{}^i s_{\pm 1}, Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}\right] = \mp (1/\sqrt{2})[(l_i \mp \lambda_i)(l_i \pm \lambda_i + 1)]^{1/2} \times Q^{[l_1,...,l_i,...,l_n]}_{\lambda_1,...,\lambda_i \pm 1,...,\lambda_n}, \tag{2.3}$$

where $\mu, \nu = -1, 0, +1$ and $\lambda_i = -l_i, -l_i + 1, ..., l_i$. Suppose that the generator basis contains a tensor $Q^{[l_1,...,l_n]}$ such that the commutators between its components produce only ${}^i s_\mu$-generators

$$\left[Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}, Q^{[l_1,...,l_n]}_{\lambda'_1,...,\lambda'_n}\right] \in (\text{algebra of } [SU(2)]^n). \tag{2.4}$$

This is the first condition we require. Note that in the case of (2.4), the $[SU(2)]^n$ generators, together with the $Q^{[l_1,...,l_n]}$ components generate a Lie algebra themselves; therefore we assume that in the following the Lie algebra of G decomposes into the $[SU(2)]^n$ subalgebra and *one* tensor $Q^{[l_1,...,l_n]}$. By means of the tensor $Q^{[l_1,...,l_n]}$, a series of coupled-tensor operators $(Q \times Q)^{[k_1,...,k_n]}$ may be defined as follows:

$$(Q \times Q)^{[k_1,...,k_n]}_{\kappa_1,...,\kappa_n} = \sum_{\substack{\lambda_1,...,\lambda_n \\ \lambda'_1,...,\lambda'_n}} \langle l_1 \lambda_1 l_1 \lambda'_1 | k_1 \kappa_1 \rangle \times \cdots \times \langle l_n \lambda_n l_n \lambda'_n | k_n \kappa_n \rangle Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n} Q^{[l_1,...,l_n]}_{\lambda'_1,...,\lambda'_n}. \tag{2.5}$$

The symbol $\langle \cdots | \cdots \rangle$, denotes an SU(2) Clebsch–Gordan coefficient. Symmetry relations between Clebsch–Gordan coefficients show that expression (2.5) is of second degree in the Q-components if and only if

$$\sum_{i=1}^{n} (2l_i + k_i) \tag{2.6}$$

is even.

Define, for $\kappa = -k, -k+1, ..., k$,

$${}^i S^{[k]}_\kappa = 1 \quad \text{if} \quad k = 0,$$

$${}^i S^{[k]}_\kappa = {}^i s_\kappa \quad \text{if} \quad k = 1,$$

$${}^i S^{[k]}_\kappa = \sum_{\substack{\kappa_1,...,\kappa_k \\ \kappa'_2,...,\kappa'_{k-1}}} \langle 1\kappa_1 1\kappa_2 | 2\kappa'_2 \rangle \langle 2\kappa'_2 1\kappa_3 | 3\kappa'_3 \rangle \cdots \times \langle k-1\kappa'_{k-1} 1\kappa_k | k\kappa \rangle {}^i s_{\kappa_1} {}^i s_{\kappa_2} \cdots {}^i s_{\kappa_k} \quad \text{if} \quad k \geqslant 2. \tag{2.7}$$

[a)] Research Assistant N.F.W.O., Belgium.

 0022-2488/84/051219-03$02.50

Then, for $(k_1,...,k_n)$ satisfying (2.6), there exists a subgroup scalar $C^{(k_1,...,k_n;2)}$ of second degree in the Q-tensor,

$$C^{(k_1,...,k_n;2)} = \sum_{\kappa_1,...,\kappa_n} (-1)^{\kappa_1+\cdots+\kappa_n} (Q \times Q)^{[k_1,...,k_n]}_{\kappa_1,...,\kappa_n} \times {}^1S^{[k_1]}_{-\kappa_1} \cdots {}^nS^{[k_n]}_{-\kappa_n}. \tag{2.8}$$

The operator (2.8) is homogeneous of degree $k_1, k_2, ..., k_n$ and 2 in the generators ${}^1s_{\mu_1}, {}^1s_{\mu_2}, ..., {}^ns_{\mu_n}$ and $Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}$, respectively. Obviously, every homogeneous subgroup scalar of second degree in the Q-components must be equal to one of the operators (2.8), to a product of an operator $C^{(k_1,...,k_n;2)}$ with $SU(2)_i$ Casimir operators (which are quadratic in the ${}^is_\mu$ generators), or to a linear combination of such operators.

Let us now consider two operators $C^{(k_1,...,k_n;2)}$ $(\Sigma_{i=1}^n k_i \neq 0)$ and $C^{(k'_1,...,k'_n;2)}$ $(\Sigma_{i=1}^n k'_1 \neq 0)$ such that

$$\sum_{i=1}^n k_i k'_1 = 0. \tag{2.9}$$

Then the commutator of these operators reduces to

$$\left[C^{(k_1,...,k_n;2)}, C^{(k'_1,...,k'_n;2)}\right] = \sum_{\substack{\kappa_1,...,\kappa_n \\ \kappa'_1,...,\kappa'_n}} (-1)^{\kappa_1+\cdots+\kappa_n+\kappa'_1+\cdots+\kappa'_n} \times \left[(Q \times Q)^{[k_1,...,k_n]}_{\kappa_1,...,\kappa_n}, (Q \times Q)^{[k'_1,...,k'_n]}_{\kappa'_1,...,\kappa'_n}\right] \times {}^1S^{[k'_1]}_{-\kappa'_1} {}^1S^{[k_1]}_{-\kappa_1} \cdots {}^nS^{[k'_n]}_{-\kappa'_n} {}^nS^{[k_n]}_{-\kappa_n}. \tag{2.10}$$

Because of (2.4) and (2.5), the commutator appearing in the rhs of (2.10) consists of terms of second degree in the Q-components and of first degree in one of the ${}^is_\mu$-generators. Hence (2.10) decomposes into homogeneous operators of degree

$$\begin{aligned} &(k'_1 + k_1 + 1,\ k'_2 + k_2,...,k'_n + k_n;2),\\ &(k'_1 + k_1,\ k'_2 + k_2 + 1,...,k'_n + k_n;2),\\ &\vdots\\ &(k'_1 + k_1,\ k'_2 + k_2,...,k'_n + k_n + 1;2), \end{aligned} \tag{2.11}$$

in the generators ${}^1s_{\mu_1}, {}^2s_{\mu_2}, ..., {}^ns_{\mu_n}$ and $Q^{[l_1,...,l_n]}_{\lambda_1,...,\lambda_n}$, respectively. But since $C^{(k_1,...,k_n;2)}$ and $C^{(k'_1,...,k'_n;2)}$ are subgroup scalars, the commutator (2.10) is a subgroup scalar too. Because of the homogeneity property,[2,7] every homogeneous operator appearing in the rhs of (2.10) must be a subgroup scalar itself. Consequently, if there doesn't correspond a subgroup scalar with any of the degrees (2.11), in other words, if none of the rows (2.11) satisfies property (2.6), we may conclude

$$\left[C^{(k_1,...,k_n;2)}, C^{(k'_1,...,k'_n;2)}\right] = 0.$$

Note that $C^{(0,...,0;2)}$ is proportional to the G second-order Casimir operator (up to an additional invariant in the $[SU(2)]^n$ subalgebra) and commutes with any other subgroup invariant. Of course, this Casimir operator cannot serve as missing label operator, and is therefore excluded in the following examples.

3. EXAMPLES

3.1. SU(4) ⊃ SU(2) ⊗ SU(2) (2 missing label problem)

The SU(4) basis generators are $s_{0,\pm1}$, $t_{0,\pm1}$ and the tensor of rank (1,1) with respect to SU(2) ⊗ SU(2): $Q^{[1,1]}_{\mu,\nu}$. The tensor components satisfy (2.4).[7] According to (2.6), the subgroup scalars of second degree in the Q-components are

$$C^{(20;2)}, C^{(02;2)}, C^{(11;2)}, \quad \text{and} \quad C^{(22;2)}. \tag{3.1}$$

On account of (2.9) we consider the commutator $[C^{(20;2)}, C^{(02;2)}]$. The discussion in Sec. 2 implies that this commutator equals a linear combination of two scalars of degree (3,2;2) and (2,3;2) in s_μ, t_ν, and $Q^{[1,1]}_{\mu,\nu}$, respectively. But the only scalars of second degree in the $Q^{[1,1]}_{\mu,\nu}$-generators are linear combinations of operators (3.1) or of products of an operator (3.1) with the SU(2) ⊗ SU(2) Casimir operators s^2 and t^2. Consequently, no scalars of degree (3,2;2) and (2,3;2) exist, and

$$[C^{(20;2)}, C^{(02;2)}] = 0. \tag{3.2}$$

This affirms the result of Quesne[5] and Partensky and Maguin.[6]

3.2. SO(7) ⊃ SU(2) ⊗ SU(2) ⊗ SO(3) (3 missing label problem)

The basis generators of SO(7) are $s_{0,\pm1}$, $t_{0,\pm1}$, $u_{0,\pm1}$, and a tensor $Q^{[1/2\ 1/2\ 1]}_{\lambda,\mu,\nu}$ of rank $(\frac{1}{2},\frac{1}{2},1)$. The commutators of the tensor components satisfy (2.4).[8] The subgroup scalars of second degree in the tensor components are

$$C^{(110;2)}, C^{(101;2)}, C^{(011;2)}, C^{(002;2)}, C^{(112;2)}. \tag{3.3}$$

The only pair for which (2.9) is valid, is $(C^{(110;2)}, C^{(002;2)})$. Formula (2.10) would lead to scalars of degree (2,1,2;2), (1,2,2;2) and (1,1,3;2), which do not exist. Consequently,

$$[C^{(110;2)}, C^{(002;2)}] = 0. \tag{3.4}$$

3.3. SO(8) ⊃ [SU(2)]⁴ (4 missing label problem)

According to the decomposition of the adjoint representation of D_4,[9] the SO(8) generator basis consists of $s_{0,\pm1}$, $t_{0,\pm1}$, $u_{0,\pm1}$, $v_{0,\pm1}$ and of the components of the $[SU(2)]^4$ tensor $Q^{[1/2\ 1/2\ 1/2\ 1/2]}$. Because of the half-odd integers in the rank of the tensor, the commutator $[Q^{[1/2\ 1/2\ 1/2\ 1/2]}_{\lambda\ \mu\ \nu\ \sigma}, Q^{[1/2\ 1/2\ 1/2\ 1/2]}_{\lambda'\ \mu'\ \nu'\ \sigma'}]$ cannot contain a tensor component, and certainly (2.4) is satisfied. The $[SU(2)]^4$ scalars of second degree in the Q-components, in the enveloping algebra of SO(8), are

$$\begin{aligned} &C^{(1100;2)}, C^{(1010;2)}, C^{(1001;2)}, C^{(0110;2)},\\ &C^{(0101;2)}, C^{(0011;2)}, \quad \text{and} \quad C^{(1111;2)}. \end{aligned} \tag{3.5}$$

An analogous discussion as in the previous cases leads to

$$\begin{aligned} &[C^{(1100;2)}, C^{(0011;2)}] = 0,\\ &[C^{(1010;2)}, C^{(0101;2)}] = 0,\\ &[C^{(1001;2)}, C^{(0110;2)}] = 0. \end{aligned} \tag{3.6}$$

Note that the technique does not always provide a pair of commuting operators. For instance, in the case of $G_2 \supset SU(2) \otimes SU(2)$, the generator basis is given by $s_{0\pm1}$,

$t_{0,\pm 1}$, and $Q^{[3/2,1/2]}_{\mu,\nu}$. Condition (2.4) is satisfied, and the scalars of second degree in the tensor components are

$$C^{(2,0;2)}, C^{(1,1;2)}, \quad \text{and} \quad C^{(3,1;2)}. \tag{3.7}$$

Obviously, this set does not contain a pair of operators such that (2.9) is satisfied.

4. CONCLUSION

A certain amount of progress is made towards the solution of the complete labeling of states transforming irreducibly under a Lie group G and a subgroup $[SU(2)]^n \subset G$. However, it must be emphasized that some serious restrictions had to be imposed in order to make our construction applicable. There are three conditions required: (1) the validity of (2.4) (which implies that $[SU(2)]^n$ generators, together with the tensor components generate a Lie algebra themselves; the assumption that the Lie algebra of G decomposes into the $[SU(2)]^n$ subalgebra and one tensor is then only a matter of convenience); (2) the condition (2.9), which is a nontrivial restriction; (3) the condition that none of the rows (2.11) satisfy (2.6).

The proposed technique, when applicable, solves the missing label problem, if exactly two labels are missing. If more are missing, the construction leads to a pair of commuting-labeling operators and solves the problem only partially. However, it is not excluded that, if $n \geqslant 3$, three (or more) operators can be found such that each pair of this set satisfies the conditions (2.9) and (2.11). Then this would lead to a set of mutually commuting scalars.

We are fully aware of the fact that the results achieved in this paper are only a step in the right direction. The construction of a complete set of missing label operators for a group-subgroup chain is, and remains a very difficult problem. Obviously, still a great amount of research can be done in this domain. For instance, in Sec. 3 we have only given a few examples of our technique. It would be very interesting (but difficult) if a classification were made of the groups satisfying the imposed conditions. It is also clear that the paper only deals with operators of second degree in the tensor components. Perhaps progress could be made if a similar (but more complicated) technique were constructed for operators of higher degree in the tensor components.

ACKNOWLEDGMENTS

I would like to thank Professor G. Vanden Berghe and Dr. H. De Meyer (Rijksuniversiteit Gent-Belgium) for valuable discussions and helpful criticisms.

[1] G. Racah, Ergeb. Exakten Naturwiss. **37**, 28 (1965).
[2] B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974).
[3] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).
[4] M. Moshinsky and J. G. Nagel, Phys. Lett. **5**, 173 (1963).
[5] C. Quesne, J. Math. Phys. **18**, 1210 (1977).
[6] A. Partensky and C. Maguin, J. Math. Phys. **19**, 511 (1978).
[7] C. Quesne, J. Math. Phys. **17**, 1452 (1976).
[8] G. Vanden Berghe, H. De Meyer, and P. De Wilde, J. Phys. A: Math. Gen. **15**, 2677 (1982).
[9] W. G. McKay and J. Patera, *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1981).

Structure constants for Lie algebras[a)]

G. Feldman and T. Fulton
Physics Department, The Johns Hopkins University, Baltimore, Maryland 21218
P. T. Matthews
University of Bath, Bath, United Kingdom

(Received 28 December 1982; accepted for publication 10 June 1983)

An orthogonal basis in root space, related to the weights of the smallest representation, is used to provide a list of the algebraic conditions which the structure constants $N_{\alpha\beta}$ must satisfy for all simple Lie algebras. A particular explicit set of solutions for all the $N_{\alpha\beta}$ is given.

PACS numbers: 02.20.Sv

I. INTRODUCTION

The present paper is concerned with exhibiting particular consistent sets of structure constants, $N_{\alpha\beta}$, which appear in the subset of commutation relations

$$[E_\alpha, E_\beta] = N_{\alpha\beta} E_{\alpha+\beta},$$

where E_α, etc. are elements of a Lie algebra, α, β are roots, and $N_{\alpha\beta}$ does not vanish if, and only if, $\alpha + \beta$ is a root.

The structure constants play important roles in the analysis[1] of Lie groups and in the application of these groups to physical problems.[2] General conditions satisfied by $N_{\alpha\beta}$, which are based on Jacobi identities involving generators associated with suitably chosen roots,[3] are well known. Simple expressions for the number of possible independently determinable structure constants, and the particular structure constants which can be independently chosen, have also been given.[4] In addition, particular choices of structure constants for specific groups have been derived.[5] More recently,[2] specific simple choices of structure constants were determined in a unified way for classical groups of arbitrary rank and for the exceptional group $G(2)$.

The purpose of the present work is to simplify further the method of Ref. 2 for choosing a consistent set of structure constants, and to extend the method to the remaining exceptional groups. Indeed, the focus of our paper is principally on the latter. We wish to present as compact a scheme as possible, in the sense that the rules for choosing the $N_{\alpha\beta}$ are essentially in the form of mnemonics. There is no difficulty in the choice of the magnitude of the $N_{\alpha\beta}$, and we have nothing new to contribute on this point. Our analysis is devoted entirely to the proper choice of a consistent set of phases for $N_{\alpha\beta}$. The choice of such phases is, of course, not unique.

Our approach is based on the specification of roots and weights in an orthogonal basis related to the weights of the representation of the smallest dimension (quarks), an approach which we have exploited previously.[2] We find that our approach, as contrasted with the use by Dynkin[6] of a nonorthogonal basis for root and weight space based on simple roots, greatly simplifies our work. Though our point of view is different from that of Dynkin,[6] our method is based on the work of that reference. We also make extensive use of the tables developed by McKay and Patera,[7] also based on the work of Dynkin.[6]

In Sec. II we write down the well-known linear and quadratic relations among the $N_{\alpha\beta}$ in the Dynkin basis. We also tabulate the roots of all simple Lie algebras in an orthogonal basis (discussed in Appendix A) and determine the magnitudes of the structure constants.

In Sec. III, we give a simple possible choice of phases of $N_{\alpha\beta}$, consistent with the above relations. These are expressed as constraints on the phases, using the orthogonal bases, in Appendix B.

II. SPECIFICATION OF ROOTS OF LIE GROUPS AND GENERAL RELATIONS BETWEEN STRUCTURE CONSTANTS

We will consider commutation relations for Lie algebras in the Dynkin basis[6,2]:

$$[\mathbf{H}, \mathbf{H}] = 0, \tag{2.1}$$

$$[\mathbf{H}, E_{\pm\alpha}] = \pm \alpha E_{\pm\alpha}, \tag{2.2}$$

$$[E_\alpha, E_{-\alpha}] = \alpha \cdot \mathbf{H} \equiv H_\alpha, \tag{2.3}$$

and the previously given relation

$$[E_\alpha, E_\beta] = N_{\alpha\beta} E_{\alpha+\beta}, \tag{2.4}$$

with α, β roots and $N_{\alpha\beta}$ vanishing unless $\alpha + \beta$ is a root. Comparison of this basis with the Chevalley basis[1,3] immediately gives, for nonzero $N_{\alpha\beta}$,

$$|N_{\alpha\beta}| = (x+1)|\alpha| \;\; |\beta|/\sqrt{2}|\alpha+\beta|, \tag{2.5}$$

where $|\alpha|$ is the length of the root α and x is the maximum positive integer such that $\beta - x\alpha$ is a root.

We conclude this section by giving a list of general relations for structure constants in the Dynkin basis.[1,3] We have first

$$N_{\alpha\beta} = -N_{\beta\alpha}, \tag{2.6}$$

which follows directly from the properties of commutators. We also have, for

$$\alpha + \beta + \gamma = 0, \quad N_{\alpha\beta} = N_{\beta\gamma} = N_{\gamma\alpha}, \tag{2.7}$$

and for

$$\alpha + \beta + \gamma + \delta = 0,$$
$$N_{\alpha\beta} N_{\gamma\delta} + N_{\beta\gamma} N_{\alpha\delta} + N_{\gamma\alpha} N_{\beta\delta} = 0. \tag{2.8}$$

Finally, we have

$$N_{\alpha\beta} N_{-\alpha,-\beta} = -|N_{\alpha\beta}|^2. \tag{2.9}$$

[a)] Supported in part by the National Science Foundation.

 0022-2488/84/051222-08$02.50

TABLE I. The quarks and roots for the classical Lie algebras. n is the rank of the algebra. L_Q^2 and $L_{\mathscr{A}}^2$ are the squares of lengths for nonzero quark weights and of roots. D_Q and $D_{\mathscr{A}}$ are the dimensions of the quark and adjoint representations and N_Φ^G is the number of independent phases of $N_{\alpha\beta}$ [see Eq. (A6)]. See Eqs. (2.13) and (A1)–(A5) in the text for the definitions and properties of λ_p and χ_p.

Algebra, G	Index labels p, q	QUARKS Weights	D_Q	L_Q^2/λ^2	ROOTS	$D_{\mathscr{A}}$	$L_{\mathscr{A}}^2/\lambda^2$	N_Φ^G
$A(n)\equiv SU(n+1)$	$1, \ldots, n+1$	χ_p	$n+1$	$n/n+1$	$\pm(\lambda_p - \lambda_q), p\neq q$	$n(n+2)$	2	$\frac{1}{2}n(n-1)$
$D(n)\equiv SO(2n)$	$1, \ldots, n$	$\pm\lambda_p$	$2n$	1	$\pm\lambda_p \pm \lambda_q, p\neq q$	$n(2n-1)$	2	$n(n-2)$
$B(n)\equiv SO(2n+1)$	$1, \ldots, n$	$\pm\lambda_p$ 0	$2n+1$	1	$\pm\lambda_p \pm \lambda_q, p\neq q;\ \pm\lambda_p$	$n(2n+1)$	2, 1	$n(n-1)$
$C(n)\equiv Sp(2n)$	$1, \ldots, n$	$\pm\lambda_p$	$2n$	1	$\pm\lambda_p \pm \lambda_q, p\neq q;\ \pm 2\lambda_p$	$n(2n+1)$	2, 4	$n(n-1)$

Equations (2.7)–(2.9) are consequences of Jacobi identities. Equation (2.9) is not generally noted. It arises from the Jacobi identity involving E_α, $E_{-\alpha}$, and E_β, where (2.7) is also used. The conventional choice for (2.9) is

$$N_{\alpha\beta} \quad \text{real},$$
$$N_{\alpha\beta} = -N_{-\alpha,-\beta}, \tag{2.10}$$

and we will adopt it in what follows. However, clearly

$$N_{\alpha\beta} = e^{i\Phi_{\alpha\beta}}|N_{\alpha\beta}|, \tag{2.11}$$

with

$$\Phi_{\alpha\beta} + \Phi_{-\alpha,-\beta} = (2n+1)\pi, \quad n = 1,2,\ldots \tag{2.12}$$

is also possible.

The different Lie algebras are completely specified by the roots. These are given in Tables I and II in terms of an orthogonal base of vectors λ_p,

$$\lambda_p \cdot \lambda_q = \delta_{pq}\lambda^2. \tag{2.13}$$

As can be seen from Table I, for the classical algebras the base vectors λ_p are very closely related to the weights of the quarks. Unless otherwise stated, we shall normalize

$$\lambda^2 = 1. \tag{2.14}$$

The derivation of these results is discussed in Appendix A.

The lengths of the structure constants $N_{\alpha\beta}$ are all determined by Eq. (2.5). From the explicit expressions for the roots given in Tables I and II, it is easy to determine the values of α and β (and hence the values of p) which give rise to nonvanishing structure constants. In this way, we arrive at the specific results given below.

Roots of an algebra are classified as "short" if the algebra contains other roots which are longer. The structure constant $N_{\alpha\beta}$ is called short,

$$N_{\alpha\beta} = N_{\alpha\beta}^{(S)}, \tag{2.15}$$

if α, β and $\alpha+\beta$ are *all* short roots. Otherwise $N_{\alpha\beta}$ is "long":

$$N_{\alpha\beta} = N_{\alpha\beta}^{(L)}. \tag{2.16}$$

For[8] $A(n)$, $B(n)$, $D(n)$, $E(6)$, $E(7)$, $E(8)$, $F(4)$, and $G(2)$,

$$|N_{\alpha\beta}^{(L)}| = \lambda. \tag{2.17}$$

$B(n)$ has short roots but no short structure constants.

TABLE II. The quarks and roots for the exceptional Lie algebras. See the caption of Table I for definition of symbols.

Algebra G	Labeling subalgebra	Index labels p	QUARKS Weights	D_Q	L_Q^2/λ^2	ROOTS	$D_{\mathscr{A}}$	$L_{\mathscr{A}}^2/\lambda^2$	N_Φ^G
$E(6)$	$D(5)\oplus U(1)$	$1\cdots 5$	$\pm\lambda_p - (1/\sqrt{3})\bar{\lambda}_6$ $\frac{1}{2}[\Sigma_{(\text{even})}(\pm\lambda_p) + (1/\sqrt{3})\bar{\lambda}_6]$ $(2/\sqrt{3})\bar{\lambda}_6$	27	$\frac{4}{3}$	$\pm\lambda_p \pm \lambda_q, \ p\neq q$ $\frac{1}{2}[\Sigma_{(\text{odd})}(\pm\lambda_p) + \sqrt{3}\bar{\lambda}_6]$ $\frac{1}{2}[\Sigma_{(\text{even})}(\pm\lambda_p) - \sqrt{3}\bar{\lambda}_6]$	78	2	30
$E(7)$	$D(6)\oplus A(1)$	$1\cdots 6$	$\pm\lambda_p \pm (1/\sqrt{2})\bar{\lambda}_7$ $\frac{1}{2}\Sigma_{(\text{odd})}(\pm\lambda_p)$	56	$\frac{3}{2}$	$\pm\lambda_p \pm \lambda_q, \ p\neq q$ $\frac{1}{2}[\Sigma_{(\text{even})}(\pm\lambda_p) \pm \sqrt{2}\bar{\lambda}_7]$ $\pm\sqrt{2}\bar{\lambda}_7$	133	2	56
$E(8)$	$D(8)$	$1\cdots 8$	$\pm\lambda_p \pm \lambda_q, p\neq q$ $\frac{1}{2}\Sigma_{(\text{odd})}(\pm\lambda_p)$ Eight zeros	248	2, 0	$\pm\lambda_p \pm \lambda_q, \ p\neq q$ $\frac{1}{2}\Sigma_{(\text{odd})}(\pm\lambda_p)$	248	2	112
$F(4)$	$B(4)$	$1\cdots 4$	$\pm\lambda_p$ $\frac{1}{2}\Sigma_{(\text{all})}(\pm\lambda_p)$ Two zeros	26	1, 0	$\pm\lambda_p \pm \lambda_q, \ p\neq q$ $\pm\lambda_p$ $\frac{1}{2}\Sigma_{(\text{all})}(\pm\lambda_p)$	52	2, 1	20
$G(2)$	$A(2)$	1, 2, 3	$\pm\chi_p = \pm[\lambda_p - \frac{1}{3}\Sigma_{q=1}^3\lambda_q]$	7	$\frac{2}{3}$, 0	$\pm(\lambda_p - \lambda_q)$ $\pm\chi_p = \pm[\lambda_p - \frac{1}{3}\Sigma_{q=1}^3\lambda_q]$	14	$2, \frac{2}{3}$	4

For $F(4)$,

$$|N^{(S)}_{\alpha\beta}| = \lambda/\sqrt{2}. \tag{2.18}$$

For $G(2)$,

$$|N^{(S)}_{\alpha\beta}| = 2\lambda/\sqrt{3}. \tag{2.19}$$

For $C(n)$,

$$|N^{(L)}_{\alpha\beta}| = \sqrt{2}\lambda, \tag{2.20}$$

$$|N^{(S)}_{\alpha\beta}| = \lambda. \tag{2.21}$$

III. CLASSIFICATION OF COMMUTATION RELATIONS. SPECIFIC PHASES FOR DIFFERENT ALGEBRAS

To simplify the notation, we introduce

$$\lambda_{|p|} \equiv (|p|), \tag{3.1}$$

$$-\lambda_{|p|} \equiv \lambda_{-|p|} = (-|p|), \tag{3.2}$$

$$2\lambda_{|p|} \equiv (2|p|). \tag{3.3}$$

Then, for example, all the roots of $D(n)$ can be written

$$(p-q), \quad p,q = \pm 1, \pm 2, ..., \pm n, \quad \text{i.e.,} \ (\pm|p| \pm |q|). \tag{3.4}$$

Finally, we introduce the notation for the raising and lowering operators:

$$E_{(p-q)} \equiv \{p-q\}. \tag{3.5}$$

Again, using Tables I and II, the nonvanishing commutation relations of the form (2.4) can be classified and the lengths of the structure constants explicitly included. It remains only to determine the phase factors, which we denote by symbols b, c, d, etc. As stated before, we choose the phase factors to be real and to satisfy Eq. (2.10). Hence they all have the values ± 1. These phases have to be chosen to satisfy the conditions (2.7) and (2.8), which are reexpressed in terms of the phase factors in Appendix B. A complete list of commutation relations of the form (2.4), together with a particular choice of phases satisfying (2.7) and (2.8), for all simple Lie algebras is given below. The phase factors are expressed in terms of[9]

$$\epsilon(x) = +1, \quad x>0,$$
$$\epsilon(x) = -1, \quad x<0, \tag{3.6}$$

and we have used the normalization (2.14). We have chosen all of our phases to satisfy Eq. (2.10), as stated above. For classical algebras we have the following.

$D(n)$:

$$[\{p-q\}, \{q-r\}] = d(p,q,r)\{p-r\},$$

where

$$p,q,r = \pm 1, ..., \pm n. \tag{3.7}$$

Here, and in all subsequent formulas for phase factors, if different labels such as p and q are used, their absolute values are always distinct. A simple expression satisfying the conditions (B6)–(B8) is

$$d(p,q,r) = \epsilon(p+q)\epsilon(q+r)\epsilon(r+p). \tag{3.8}$$

$A(n-1)$: As for $D(n)$, except that p, q, r are all positive or all negative so that

$$d(p,q,r) = +1$$

for all positive p, q, r.

$B(n)$: Equation (3.7), together with

$$[\{p-q\}, \{q\}] = b(p,q)\{p\}, \tag{3.9}$$

$$[\{p\}, \{-q\}] = \bar{b}(p,q)\{p-q\}, \tag{3.10}$$

$$p,q,r = \pm 1, ..., \pm n.$$

To satisfy the conditions (B10)–(B12), $d(p,q,r)$ can be taken as in (3.8) and

$$b(p,q) = \bar{b}(p,q) = \epsilon(p+q). \tag{3.11}$$

$C(n)$:

$$[\{p-q\}, \{q-r\}] = c(p,q,r)\{p-r\}, \tag{3.12}$$

$$[\{p-q\}, \{2q\}] = \sqrt{2}\sigma(p,q)\{p+q\}, \tag{3.13}$$

$$[\{p-q\}, \{q+p\}] = \sqrt{2}\bar{\sigma}(p,q)\{2p\}, \tag{3.14}$$

$$p,q,r = \pm 1, ..., \pm n.$$

To satisfy the conditions (B13)–(B17), we can take[10]

$$c(p,q,r) = \tfrac{1}{2}[\epsilon(p) + \epsilon(q) + \epsilon(r) - \epsilon(p)\epsilon(q)\epsilon(r)], \tag{3.15}$$

$$\sigma(p,q) = \bar{\sigma}(p,q) = \epsilon(p). \tag{3.16}$$

The expressions given above for the phase factors are in a form in which it is relatively simple to check that they satisfy the necessary conditions summarized[11] in Appendix B. These expressions can be evaluated explicitly by making use of two other definitions in the phase factors such as $d(p,q,r)$. If we have a specific sign in an index, the negative indices will be written as superscripts and the positive as subscripts. For example,

$$d(-3,2,8) \equiv \equiv d^{3}_{28}. \tag{3.17}$$

Finally, we will define a phase which has its indices ordered. For example,

$$d(-3,2,8) \to \tilde{d}(2,-3,8), \tag{3.18}$$

and

$$d^{3}_{28} \to \tilde{d}_{2}{}^{3}{}_{8}. \tag{3.19}$$

Using this notation, a consistent set of phases of all structure constants for the classical Lie algebras is given in Table III. In this table, we list numerical values of essential phases in the above algebraic relations. By essential phases, we mean the phases from which all other phases follow trivially. For example, $\bar{b}(p,q)$ and $b(-p,-q)$ follow trivially from $b(p,q)$ [cf. Eq. (B11)], therefore no $\bar{b}$'s and not all b's are listed in Table III.

The first entry in line one of Table III suffices for $A(n)$, the first four entries for $D(n)$. All entries in line one are needed for $B(n)$, and all in line two for $C(n)$.

We now turn to the exceptional algebras.

TABLE III. Explicit numerical values of phase choices for classical algebras. Only essential phases are listed and $\pm \equiv \pm 1$. See Eqs. (3.17) and (3.19) for definition of notation and observe that, in the table, $r>p>q>0$.

$\tilde{d}_{pqr} = +$	$\tilde{d}_{p}{}^{qr} = -$	$\tilde{d}_{pq}{}^{r} = +$	$\tilde{d}_{p}{}^{q}{}_{r} = -$	$\tilde{b}_{pq} = +$	$\tilde{b}_{p}{}^{q} = -$
$\tilde{c}_{pqr} = +$	$\tilde{c}_{p}{}^{qr} = -$	$\tilde{c}_{pq}{}^{r} = +$	$\tilde{c}_{p}{}^{q}{}_{r} = +$	$\sigma_{pq} = +$	$\sigma_{p}{}^{q} = +$

$E(8)[E(7), E(6)]$:

$$[\{p-q\}, \{q-r\}] = d(p,q,r)\{p-r\}, \tag{3.20}$$

$$[\{p-q\}, \{\tfrac{1}{2}(-p+q+r+s+t+u+v+w)\}] = e(p,q;r,s,\ldots,w)\{\tfrac{1}{2}(p-q+r+s+\cdots+w), \tag{3.21}$$

$$[\{\tfrac{1}{2}(p-q+r+s+\cdots+w)\}, \{\tfrac{1}{2}(p-q-r-s-\cdots-w)\}] = \bar{e}(p,q;r,s,\ldots,w)\{p-q\}, \tag{3.22}$$

$$p, q, \ldots, \quad w = \pm 1, \ldots, \pm 8,$$

with an even number of negative indices in e (and $\bar{e}$), so that for $E(8)$ there is an odd number in the roots which are spinors in $D(8)$ (see Table II). In order to satisfy the symmetry property of e [Eq. (B20)], it is convenient to identify explicitly the label 8. To satisfy all the conditions (B19)–(B22), we can take

$$d(p,q,r) = \epsilon(p+q)\epsilon(q+r)\epsilon(r+p), \tag{3.23}$$

$$e = \bar{e}, \tag{3.24}$$

and

$$e(p,q;r,s,t,u,v,-8) = -\prod \epsilon(|p|+l)\epsilon(|q|+l), \tag{3.25}$$

where

$$p, q, \ldots, v = \pm 1, \ldots, \pm 7,$$

and

$$l = r, s, t, u, v.$$

By (B20),

$$e(p,q;r,s,t,u,v,8) = -e(p,q;-r,-s,-t,-u,-v,-8) = \prod_l \epsilon(|p|-l)\epsilon(|q|-l). \tag{3.26}$$

We also take

$$e(p,-8;r,s,t,u,v,w,) = e(-p,8;r,s,t,u,v,w) = \prod_{\substack{l,m \\ |l|<|m|}} \epsilon(l+m), \quad p>0, \tag{3.27}$$

$$e(p,8;r,s,t,u,v,w) = e(-p,-8;r,s,t,u,v,w) = (-1)^{p+1} \prod_{\substack{l,m \\ |l|<|m|}} \epsilon(l+m), \quad p>0, \tag{3.28}$$

where

$$p = +1, \ldots, +7, \quad r, s, t, u, v, w, = \pm 1, \ldots, \pm 7,$$

and

$$l, m = r, s, \ldots, w.$$

These expressions can be greatly simplified if it is known how many of the labels take on negative values. For example, using the identities

$$\prod_{\substack{l \\ |l|\neq|p|}} \epsilon(|l|-|p|) = (-1)^{p+1}, \tag{3.29}$$

one can show

$$e(p,-8;-r,s,t,u,v,w) = (-1)^{r+1}\epsilon(p-r), \tag{3.30}$$

$$e(p,8;-r,-s,t,u,v,w) = (-1)^{p+r+s+1}\epsilon(p-r)\epsilon(p-s), \tag{3.31}$$

$$e(p,-8;-r,-s,-t,u,v,w) = (-1)^{r+s+t+1}\epsilon(p-r)\epsilon(p-s)\epsilon(p-t), \text{ etc.,} \tag{3.32}$$

where

$$p, r, s, t, u, v, w > 0$$

in Eqs. (3.30)–(3.32) above.

Other useful relations are

$$\epsilon(|p|(\pm)|q|)\epsilon(|q|(\pm)|p|) = (\pm)\epsilon(p+q), \tag{3.33}$$

and

$$(-1)^{\sum_{|m|=1}^{7} m} = 1. \tag{3.34}$$

The commutation relations for $E(7)$ and $E(6)$ can be obtained from the above equations by the techniques described in Appendix A. Thus to obtain $E(7)$, Eq. (3.20), which defines the d phases, is still valid if the labels are restricted to

$$p, q, r = \pm 1, \ldots, \pm 6.$$

Equations (3.21) and (3.22) apply if either v and w or p and $-q$ appear in the combination $7, -8$ with the magnitudes of the remaining labels taking the values from 1 to 6. Possible e phases are given by

Eq. (3.25) with

$$v = 7, \tag{3.35}$$

Eq. (3.26) with

$$v = -7, \tag{3.36}$$

and Eq. (3.27) with

$$p = 7. \tag{3.37}$$

In $E(6)$, Eq. (3.20), defining the d phases, is still valid, but the labels are now restricted to

$$p, q, r = \pm 1, \ldots, \pm 5.$$

The subset of e phases does not include any of those given by Eqs. (3.27) and (3.28). Of the e phases given in Eqs. (3.25), only

$$e(p,q;r,s,t,6,7,-8) \equiv e(p,q;r,s,t,-\bar{\bar{6}}) \equiv \theta(r,s,t), \tag{3.38}$$

and of those given in (3.26), only

$$e(p,q;r,s,t,-6,-7,8) \equiv e(p,q;r,s,t,\bar{\bar{6}}) = -\theta(-r,-s,-t) \tag{3.39}$$

appear [see Eq. (A12)]. The function θ, defined in Eq. (3.38), is sufficient to describe the e phases. Given r, s, t, $|p|$ and $|q|$ are determined. Further, if the number of negative indices in r, s, t is odd (even), the number of negative indices in p and q is odd (even). For convenience, we simplify the rather cumbersome expression, Eq. (3.25), for $E(6)$. We obtain, using the convention of Eq. (3.17),

TABLE IV. Explicit numerical values of the phases $\theta(r, s, t)$.

r	s	t	$\tilde{\theta}_{rst}$	$\tilde{\theta}^{r}{}_{st}$	$\tilde{\theta}_{r}{}^{s}{}_{t}$	$\tilde{\theta}_{rs}{}^{t}$	$\tilde{\theta}_{r}{}^{st}$	$\tilde{\theta}^{r}{}_{s}{}^{t}$	$\tilde{\theta}^{rs}{}_{t}$	$\tilde{\theta}^{rst}$
3	4	5	−	−	−	−	−	−	−	−
2	4	5	−	+	−	−	−	+	+	+
1	4	5	−	−	−	−	−	−	−	−
2	3	5	−	+	+	−	+	+	−	−
1	3	5	−	−	+	−	+	−	+	+
2	3	4	−	+	+	+	−	−	−	+
1	3	4	−	−	+	+	−	+	+	−
1	2	5	−	−	−	−	−	−	−	−
1	2	4	−	−	−	+	+	+	−	+
1	2	3	−	−	−	−	−	−	−	−

$$
\begin{aligned}
&\theta_{rst} = -1,\\
&\theta^{r}_{st} = -\epsilon(|p|-|r|)\epsilon(|q|-|r|),\\
&\theta^{st}_{r} = (-1)^{p+q}\epsilon(|p|-|r|)\epsilon(|q|-|r|),\\
&\theta^{rst} = (-1)^{p+q}.
\end{aligned}
\tag{3.40}
$$

We give explicit numerical values of these additional essential phases for E (6) (besides those which already appear in Table III), in Table IV.

$F(4)$: $B(4)$ relations, (3.7), (3.9), and (3.10), and

$$
\begin{aligned}
&[\{p-q\}, \{\tfrac{1}{2}(-p+q+r+s)\}]\\
&\quad = f(p,q;r,s)\{\tfrac{1}{2}(p-q+r+s)\},
\end{aligned}
\tag{3.41}
$$

$$
\begin{aligned}
&[\{\tfrac{1}{2}(p-q+r+s)\}, \{\tfrac{1}{2}(p-q-r-s)\}]\\
&\quad = \bar{f}(p,q;r,s)\{p-q\},
\end{aligned}
\tag{3.42}
$$

$$
\begin{aligned}
&[\{p\}, \{\tfrac{1}{2}(-p+q+r+s)\}]\\
&\quad = (1/\sqrt{2})\phi(p;q,r,s)\{\tfrac{1}{2}(p+q+r+s)\},
\end{aligned}
\tag{3.43}
$$

$$
\begin{aligned}
&[\{\tfrac{1}{2}(p+q+r+s)\}, \{\tfrac{1}{2}(p-q-r-s)\}]\\
&\quad = (1/\sqrt{2})\bar{\phi}(p;q,r,s)\{p\},
\end{aligned}
\tag{3.44}
$$

$p, q, r, s = \pm 1, \ldots, \pm 4$.

The spinors have all possible sign distributions.

To satisfy the conditions (B23)–(B28), the phases d, b, and $\bar{b}$ are given in Eqs. (3.8) and (3.11). As for E (8), we here single out the index 4 to give solutions as follows:

$$
\begin{aligned}
&f = \bar{f}, \quad \phi = \bar{\phi},\\
&f(p,q;r,-4) = -\epsilon(|p|+r)\epsilon(|q|+r).
\end{aligned}
\tag{3.45}
$$

By the symmetry condition (B24), this implies

$$
f(p,q;r,4) = \epsilon(|p|-r)\epsilon(|q|-r).
\tag{3.46}
$$

Also,

$$
f(p,-4;r,s) \equiv f(-p,4;r,s) = \epsilon(r+s), \quad p>0
\tag{3.47}
$$

$$
\begin{aligned}
f(p,4;r,s) &= f(-p,-4;r,s)\\
&= (-1)^{p+1}\epsilon(r+s), \quad p>0
\end{aligned}
\tag{3.48}
$$

$$
\phi(p;q,r,-4) = -\epsilon(|p|+q)\epsilon(|p|+r),
\tag{3.49}
$$

$$
\phi(p;q,r,4) = \epsilon(|p|-q)\epsilon(|p|-r),
\tag{3.50}
$$

$$
\phi(\pm 4;p,q,r) = -\epsilon(p+q)\epsilon(q+r)\epsilon(r+p).
\tag{3.51}
$$

$G(2)$: We now have

$$
\{p\} \equiv E_{\chi_p},
\tag{3.52}
$$

and not E_{λ_p}, where $p = 1,2,3$, and correspondingly for $E_{-\chi_p}$. The definition, Eq. (3.5), still holds for the other generators. Further,

$$
\chi_1 + \chi_2 + \chi_3 = 0
\tag{3.53}
$$

so that

$$
\{-r\} = \{p+q\}, \quad p \neq q \neq r.
\tag{3.54}
$$

The commutation relations are given by Eq. (3.7) and

$$
[\{p-q\}, \{q\}] = g(p,q)\{p\},
\tag{3.55}
$$

$$
[\{p\}, \{-q\}] = \bar{g}(p,q)\{p-q\},
\tag{3.56}
$$

$$
[\{p\}, \{q\}] = (2/\sqrt{3})\bar{\bar{g}}(p,q)\{-r\},
\tag{3.57}
$$

where

$$
p, q, r = \pm 1, \pm 2, \pm 3,
$$

taking either all positive or all negative signs.

A convenient choice of independent phases is

$$
\begin{aligned}
&g = \bar{g},\\
&g(1,2) = g(2,3) = g(1,3) = \bar{\bar{g}}(2.1) = +1,
\end{aligned}
\tag{3.58}
$$

the remaining phases being determined by (B29)–(B33). This solution agrees with that of Ref. 2.

IV. SUMMARY AND DISCUSSION

We have succeeded in obtaining sets of simple phase choices for the structure constants of all simple Lie algebras. We have used essentially the same method of construction for all the algebras and have obtained simple algebraic expressions for the phases, which can be easily evaluated and which have closely related structures for all algebras. We have thus achieved our aim of providing a unified and systematic presentation for the $N_{\alpha\beta}$ phases of all Lie algebras.

Our approach takes advantage of an orthogonal basis in root space. Though it is a development of the Dynkin point of view,[6] it does not follow Dynkin in using simple roots as a (nonorthogonal) basis of root space. One can contrast our approach with that of Dynkin in different words, by saying that our focus is on all the quarks (actually the nonzero weights of the lowest-dimensional representations) of the

classical algebras, rather than directly on a subset of weights (the simple roots) of the adjoint representation of each particular algebra. We can thus simply describe all weights and all phases by general formulas, rather than give only a selection of them, with prescriptions which are easy to state but at times tedious to carry out in detail for the remaining ones.

The above considerations are, most of the time, aesthetic ones and/or those of convenience. However, if we wish to relate the algebra to all (or even many) of its maximal nonregular subalgebras simultaneously, the simultaneous consistent choice of all phases is no longer just a matter of aesthetics or convenience, but one of necessity. Enough phases will appear in such cases, even if only the parts of the algebra associated with the simple roots are specified, that the linear and quadratic relations of the $N_{\alpha\beta}$ come into play to restrict our freedom in choosing them. These considerations are not purely formal ones, since the simultaneous study of all maximal subgroups of a given group can be of physical interest, for example, in the symmetry breaking in grand unified theories.[2,12,13]

We have previously given phase solutions[2] for the classical algebras and for $G(2)$ from this point of view. We further simplified our presentation of our previous results in the present paper. However, our main focus in the current work was on the $N_{\alpha\beta}$ phase for the remaining exceptional algebras $E(6)$, $E(7)$, $E(8)$, and $F(4)$, so as to complete the phase analysis of all Lie algebras. Because of our emphasis on the use of orthonormal bases, it is natural to describe the roots of these algebras in terms of the weights of representations of their regular $B(n)$ and $D(n)$ subalgebras, rather that in terms of representations of their regular $A(n)$ subalgebras, as is often done. The representations needed for the description of roots for $E(6,7,8)$ and $F(4)$ are the adjoint and at least one of the spinor representations of the corresponding $B(n)$ or $D(n)$ subalgebras but *no* other representation.

APPENDIX A: THE ORTHOGONAL BASIS FOR QUARKS AND ROOTS

In this Appendix we clarify the choice of orthogonal bases,[14] which is expressed in Tables I and II.

For $B(n)$, $C(n)$, and $D(n)$, the orthogonal λ_p are given by the weights of the smallest representation (the quarks).

In $A(n)$, the quark weights, χ_p, are not orthogonal. However, they can be defined in terms of $n+1$ linearly independent orthogonal λ_p's by

$$\chi_p = \lambda_p - \frac{1}{n+1}\sum_{q=1}^{n+1}\lambda_q \tag{A1}$$

implying

$$\chi^2 = (n/(n+1))\lambda^2. \tag{A2}$$

It follows that the $n+1$ χ_p vectors satisfy the condition

$$\sum_{p=1}^{n+1}\chi_p = 0, \tag{A3}$$

i.e., the χ_p are linearly dependent. In addition,

$$\chi_p \cdot \chi_q = -(1/n)\chi^2. \tag{A4}$$

The roots of $A(n)$ are

$$\pm(\chi_p - \chi_q) = \pm(\lambda_p - \lambda_q), \tag{A5}$$

showing explicitly that they are a subset of the roots of $D(n+1)$. The discussion of Ref. 4 can be reduced to the following formula for the number, N_Φ^G, of $N_{\alpha\beta}$ phases, which can be chosen independently:

$$N_\Phi^G = \tfrac{1}{2}(D_{\mathscr{A}} - 3n), \tag{A6}$$

where $D_{\mathscr{A}}$ is the dimension of the adjoint representation of algebra G (the rank plus the number of roots).

The roots of the exceptional algebras in Table II can be given in terms of the roots and weights of classical subalgebras of these algebras. This means that various options for choices of classical algebras are available. A frequent choice[6,2] [except for $F(4)$] is a suitable $A(n)$ subalgebra. However, because of the simpler orthogonal basis which can be defined for the (rotation) algebras $B(n)$ and $D(n)$, we will find these latter to be a more convenient choice.

We will illustrate our approach with $E(6)$ and summarize the results for all exceptional algebras in Table II.

We know that a maximal regular algebra[6,7] of $E(6)$ is $D(5)$. Indeed, $E(6) \supset D(5) \otimes U(1)$. We consider the quarks first and use Ref. 7 with its notation. We have

$$[\text{Quark}]_{E(6)} = [\text{Quark} + \text{Spinor} + \text{Scalar}]_{D(5)}, \tag{A7}$$

$$(000010)_{E(6)} = [(10000) \oplus (00010) \oplus (00000)]_{D(5)},$$
$$27 = 10 + 16 + 1, \tag{A8}$$

where the numbers below the Dynkin–Patera labels are the dimensions of the representations in question. The nonzero weights of all quark representations of any Lie algebra are the same length. Also the vector sum of the weights of any given representation vanishes. We see from (A8) that the weights of the quark of $E(6)$ must be

$$\pm\lambda_p + a\bar{\bar{\lambda}}_6, \quad \tfrac{1}{2}\sum_{\substack{p=1\\(\text{even})}}^{5}(\pm\lambda_p) + b\bar{\bar{\lambda}}_6, \quad c\bar{\bar{\lambda}}_6, \quad p = 1,\ldots,5, \tag{A9}$$

where λ_p is associated with $D(5)$ and $\bar{\bar{\lambda}}_6$ with $U(1)$, and the (even) under the sum for the spinor of $D(5)$ means that an even number of negative signs are taken. The conjugate representation $\overline{16}$ or $(00001)_{D(5)}$ has an odd number of minus signs in the sum. The symbol (all) in the entry for $F(4)$ later on refers to both even and odd numbers of minus signs. All λ_p, $\bar{\bar{\lambda}}_6$ have length squares equal to λ^2. Use of the two conditions stated above Eq. (A9) serves to determine a, b, and c. The result is given in Table II. Knowing the $D(5)$ decomposition of the adjoint representation,[7] we construct it from the decomposition of $27 \otimes \overline{27}$, recognizing that it must be self-conjugate and the nonzero weights must all have the same length. The result is once again given in Table II. A similar approach can serve for $E(7)$, $E(8)$, and $F(4)$.

Some further comments on Table II are in order. $G(2)$ is rather a special case among the exceptional algebras, and is treated differently from the other exceptional algebras. This presents no problems, since it is of low rank and has an adjoint of only 14 dimensions. It makes no difference whether the $A(2)$ or $B(1) \oplus A(1)$ classical algebras of $G(2)$ are used to define its roots. On the other hand, labeling $E(6)$ with

$A(5) \oplus A(1), E(7)$ with $A(7)$, and $E(8)$ with $A(8)$ leads to very different labeling schemes than the ones given in Table II.

Returning to $E(6,7,8)$, we note that a simple relation exists among the roots and structure constants of these algebras since

$$E(8) \supset E(7) \oplus A(1), \quad E(7) \supset E(6) \oplus U(1), \tag{A10}$$

and the subalgebras are in each case maximal and regular. Comparison of the entries for the roots of $E(6,7,8)$ immediately shows that we can make the identifications

$$\sqrt{2}\,\bar{\lambda}_7 = \lambda_7 - \lambda_8, \tag{A11}$$

and

$$-\sqrt{3}\,\bar{\bar{\lambda}}_6 = \lambda_6 + \lambda_7 - \lambda_8 = \lambda_6 + \sqrt{2}\,\bar{\lambda}_7. \tag{A12}$$

Note the minus sign on the lhs of Eq. (A12) above.

We must emphasize that $\bar{\lambda}_7$ and $\bar{\bar{\lambda}}_6$ are vectors extraneous to the bases for $D(6)$ [labeling $E(7)$] and $D(5)$ [labeling $E(6)$], respectively.

The generators of $E(7)$ are obtained from those of $E(8)$ by dropping H_8 and all the E_α's corresponding to the $E(8)$ roots which depend on either λ_8 or λ_7 unless they occur in the combination $\pm(\lambda_7 - \lambda_8)$. In the latter case, the E_α is retained using the substitution (A11).

Similarly, the generators of $E(6)$ are obtained from those of $E(7)$ by a further dropping of H_7 and the E_α's which depend on either λ_6 or $\lambda_7 - \lambda_8$ unless they occur in the combination $\pm(\lambda_6 + \lambda_7 - \lambda_8)$, in which case they are retained in accordance with (A12).

The generators of $E(7)$ are thus subsets of those of $E(8)$, and the generators of $E(6)$ are, in turn, subsets of those of $E(7)$. It is thus easy to identify which structure constants of $E(8)$ remain as structure constants of $E(7)$ and, in turn, of $E(6)$. We need, therefore, only to provide an explicit choice of structure constants for $E(8)$ in order to give one for $E(7)$ and $E(6)$ as well.

We conclude this Appendix by noting that, excluding $G(2)$, the roots of the exceptional algebras are not only given by the roots of maximal orthogonal subalgebras, but also by the nonzero weights of at least one spinor representation of the corresponding orthogonal algebra.

APPENDIX B: THE PHASE CONDITIONS

In this Appendix, we use the general conditions (2.7) and (2.8) to determine the conditions which have to be satisfied by the phase factors b, c, d, etc., introduced in Sec. III.

Using the notation (3.4) we choose

$$\begin{matrix} \alpha & \beta & \gamma \\ (p-q) & (q-r) & (r-p), \end{matrix} \tag{B1}$$

and obtain, from (2.7),

$$d(p,q,r) = d(q,r,p). \tag{B2}$$

Additional use of (2.6) and (2.10) leads to the result that $d(p,q,r)$ is a symmetric function of its three indices.

It remains to apply the bilinear conditions (2.8). They lead to, at most, two additional independent sets of conditions and no more for the two sets of root choices

$$\begin{matrix} \alpha & \beta & \gamma & \delta \\ (p-q) & (q-r) & (r-s) & (s-p), \end{matrix} \tag{B3}$$

and

$$\begin{matrix} \alpha & \beta & \gamma & \delta \\ (p-q) & (q-r) & (r+q) & (-q-p). \end{matrix} \tag{B4}$$

We follow an approach similar to that just outlined for $D(n)$ for all other Lie algebrs. The resulting phase conditions are summarized below.

For all phases, Φ, $e^{i\Phi} = d, c, b, \ldots$, and all Lie algebras,

$$\Phi(p,q\cdots) = 0 \text{ or } \pi. \tag{B5}$$

For the classical algebras we have the following.

$D(n)$:

$$d(p,q,r) = -d(-p,-q,-r) = \tilde{d}(p,q,r) \tag{B6}$$

[see Eqs. (3.18) and (3.19)], and we have the four-index relation

$$d(q,r,s)d(p,r,s)d(p,q,s)d(p,q,r) = 1, \tag{B7}$$

and the three-index relation

$$d(p,q,r)d(-p,q,r)d(p,-q,r)d(p,q,-r) = -1. \tag{B8}$$

$A(n)$:

Equations (B6) and (B7) but *not* (B8), (B9)

since p, q, r must all be of the same sign.

$B(n)$:

[Equations (B6)–(B8)], (B10)

and

$$b(p,q) = \bar{b}(p,q) = -b(-p,-q) = \tilde{b}(p,q), \tag{B11}$$

$$d(p,q,r) = b(p,q)b(q,r)b(r,p). \tag{B12}$$

We note that (B11) and (B12) are consistent with (B6)–(B8) so that the latter equations are redundant. We have therefore put them in square brackets. We shall also do so with all redundant relations below.

$C(n)$:

$$c(p,q,r) = -c(-p,-q,-r) = \tilde{c}(p,q,r), \tag{B13}$$

$$c(q,r,s)c(p,r,s)c(p,q,s)c(p,q,r) = 1, \tag{B14}$$

$$[c(p,q,r)c(-p,q,r)c(p,-q,r)c(p,q,-r)] = +1, \tag{B15}$$

$$\sigma(p,q) = \sigma(p,-q) = \bar{\sigma}(q,p) = -\sigma(-p,-q), \tag{B16}$$

$$c(p,q,r)c(p,-q,-r) = -\sigma(q,p)\sigma(r,p). \tag{B17}$$

Note that, while (B6) and (B7) for d are formally equivalent to (B13) and (B14) for c, (B8) and (B15) differ by a minus sign on the right-hand side. Further, comparison of (B11) and (B16) shows other important differences between $B(n)$ and $C(n)$: $\sigma(p,q)$ is *not* symmetric in p and q, while $b(p,q)$ is:

$$\sigma(p,q) \neq \sigma(q,p). \tag{B18}$$

Equation (B15) is redundant, since it follows from (B16) and (B17). It is included only to provide a comparison with (B8). The need for defining c independently of d is now made apparent. This need exists in spite of the formal equivalence of the defining equations (3.7) and (3.12) for c and d, respectively, and of the identity of some of the constraint conditions on them [Eqs. (B6) and (B13), (B7) and (B14)].

We turn next to the phase conditions for the exceptional algebras.

$E(8)[E(7), E(6)]$:

[Equations (B6)–(B8)], (B19)

and

$$e(p, q; r, s, \ldots, w) = \bar{e}(p, q; r, s, \ldots, w) = e(-p, -q; r, s, \ldots, w) = -e(p, q; -r, -s, \ldots, -w), \quad \text{(B20)}$$

$$d(p, q, r) = -e(p, q; r, s, \ldots, w)e(q, r; p, s, \ldots, w) \times e(p, r; q, s, \ldots, w), \quad \text{(B21)}$$

$$[e(p, q; r, s, t, \ldots, w)e(p, q; -r, -s, t, \ldots, w) = e(r, s; p, q, t, \ldots, w)e(r, s; -p, -q, t, \ldots, w)]. \quad \text{(B22)}$$

One can first show that (B20) and (B21) imply (B6)–(B8). Hence, the latter equations are bracketed. Next, one can demonstrate that (B7) implies (B22) and so that equation is bracketed as well.

$F(4)$:

[Equations (B11) and (B12)], (B23)

and

$$f(p, q; r, s) = \bar{f}(p, q; r, s) = f(-p, -q; r, s) = -f(p, q; -r, -s), \quad \text{(B24)}$$

$$\phi(p; q, r, s) = \bar{\phi}(p; q, r, s) = \phi(-p; q, r, s) = -\phi(p; -q, -r, -s), \quad \text{(B25)}$$

$$d(p, q, r) = -f(p, q; r, s)f(q, r; p, s)f(p, r; q, s), \quad \text{(B26)}$$

$$b(p, q) = -f(p, q; r, s)\phi(p; q, r, s)\phi(q; p, r, s), \quad \text{(B27)}$$

$$\phi(p; q, r, s)\phi(q; p, r, s) = -\phi(p; -q, r, s)\phi(q; -p, r, s). \quad \text{(B28)}$$

$G(2)$:

[Equations (B6), with p, q, r the same sign], (B29)

$$g(p, q) = \bar{g}(p, q) = -g(-p, -q) = g(q, p), \quad \text{(B30)}$$

$$\bar{\bar{g}}(p, q) = \bar{\bar{g}}(q, r) = \bar{\bar{g}}(r, p) = -\bar{\bar{g}}(-p, -q), \quad \text{(B31)}$$

$$\bar{\bar{g}}(p, q) = -\bar{\bar{g}}(q, p), \quad \text{(B32)}$$

$$d(p, q, r) = g(p, q)g(q, r)g(r, p). \quad \text{(B33)}$$

This completes the specification of all phase constraints for all simple Lie algebras. It must be emphasized that, while manifest redundant expressions have been noted above, this is not to imply that all constraints in the lists above which appear outside of square brackets are independent. Indeed, this is far from the case in general. However, these remaining redundancies are more tedious to specify than the ones which have been eliminated. These constraints, of course, do not lead to unique solutions for the phases $d,c,b, \cdots$. Indeed, as shown in Ref. 4, one can choose N_Φ^G independent phases, where N_Φ^G is given by (A6). The proof of this latter result can also be carried out using the constraint equations given in this Appendix.

[1] B. G. Wybourne, *Classical Groups for Physicists* (Wiley-Interscience, New York, 1974), pp. 90,91.

[2] G. Feldman, T. Fulton, and P. T. Matthews, J. Phys. G: Nucl. Phys. **8**, 295 (1982).

[3] R. W. Carter, *Simple Groups of Lie Type* (Wiley-Interscience, New York, 1972), p. 55.

[4] Reference 3, p. 58. The structure constants which can be independently chosen are associated in Ref. 3 with so-called "extra special" pairs of roots. The specification of these roots in Ref. 3 is far from trivial.

[5] For example, Ref. 1, pp. 85, 86, 91; also B. Gruber and M. T. Samuel, in *Group Theory and Its Applications*, Vol. 3, edited by E. M. Loebl (Academic, New York, 1974), p. 122. [The latter authors make an incorrect choice of phases for the $C(3)$ structure constants, in embedding $C(3)$ in $A(5)$].

[6] E. B. Dynkin, Am. Math. Soc. Transl. **17** (1950); **6**, 111 (1957).

[7] W. McKay and J. Patera, *Tables of Dimensions and Branching Rules for Representations of Simple Lie Algebras* (Marcel Dekker, New York, 1981).

[8] Alternatively, if α_L is a long root, for *all* groups

$$|N^{(L)}|^2 = \tfrac{1}{2}\alpha_L^2,$$

$$|N^{(S)}|^2 = x\,\alpha_L^2,$$

where

$x = \frac{1}{4}$ for $C(n)$ and $F(4)$,

$x = \frac{3}{4}$ for $G(2)$.

Thus $|N^{(L)}| > |N^{(S)}|$ for all groups, except $G(2)$.

[9] A notation very similar to the present one was introduced in Ref. 2, provided we identify the $\epsilon_{x,-y}$ of Ref. 2 with the present form, $\epsilon(x+y)$. Indeed, the phase choices of Ref. 2 for $A(n)$, $B(n)$, $D(n)$, and $G(2)$ are identical with those made in the present paper.

[10] This set of phases for $C(n)$ differs from that given in Ref. 2, which do not satisfy all the required constraints.

[11] The definitions for $d(p, q, r)$ and $c(p, q, r)$, Eqs. (3.7) and (3.12), are formally identical, but these phase factors are parts of different self-consistent sets of phases. That is, a solution for $d(p, q, r)$ which satisfies all constraints in $D(n)$, will not satisfy all constraints present in $C(n)$.

[12] G. Feldman and R. Holman, J. Phys. G: Nucl. Phys. **9**, 7 (1983).

[13] R. Holman, J. Phys. G: Nucl. Phys. **9**, 35 (1983).

[14] Those for the classical algebras have been given in this form in Ref. 2.

The semisimple subalgebras of Lie algebras[a)]

G. Feldman and T. Fulton
Physics Department, The Johns Hopkins University, Baltimore, Maryland 21218
P. T. Matthews
University of Bath, Bath, United Kingdom

(Received 28 December 1982; accepted for publication 10 June 1983)

Techniques for obtaining generators of nonregular maximal subalgebras of Lie algebras, alternative to those of Dynkin, are developed. They exploit the use of orthogonal bases in weight space, which are related to quark weights. The projection from an algebra G to its nonregular subalgebras g is related to an orthogonal matrix. The roots of G which project onto roots of g can be simply specified in the orthogonal bases. The phases of the expansion coefficients of generators of g in terms of generators of G are specified in such a manner that, for a given G, the entire set $\{g\}$ of maximal subalgebras have consistent phases. The condition $e_{-\beta} = e_\beta^\dagger$ is satisfied for all generators of g. The generators of all maximal nonregular subalgebras of all exceptional algebras are exhibited.

PACS numbers: 02.20.Sv

1. INTRODUCTION

The subalgebras of Lie algebras have been discussed by Dynkin,[1,2] using his formalism of simple roots and employing phase and normalization conventions, which are chosen in each particular context to simplify the problem in hand. In this paper we will develop alternative techniques for deriving explicit expressions for the generators of nonregular maximal subalgebras, exploiting the use of the orthogonal base related to the weights of the representations of smallest dimension—which for the sake of brevity will be called "quarks" in what follows.[3,4] This provides a geometrical view of the relationship between algebras and subalgebras which is more perspicuous to physicists accustomed to a quark-oriented approach to group theory. We will also establish phase and reality conditions, related to our previous work.[4] They will be used systematically throughout the discussion.

In the next section, Sec. 2, we review relevant aspects of Lie algebras, with the aim of introducing some notation. Section 3 is a brief excursion into the subject of regular subalgebras, in particular maximal regular subalgebras. The maximal regular subalgebras of all Lie algebras are listed. The convenience of using the maximal regular orthogonal subalgebras of the exceptional algebras to express the roots of these algebras in terms of orthogonal basis vectors is pointed out. The special considerations needed to fit $A(n)$ algebras and subalgebras into the orthogonal basis scheme are reviewed and further discussed. The regular subalgebras are considered, in order to provide a contrast to the nonregular subalgebras, in particular, maximal nonregular (MNR) subalgebras, which are the chief concern of the present paper. These latter algebras are considered in general in Sec. 4. The first part of this secton is devoted to exhibiting the scale factor associated with the projection of a larger dimensional weight and root space, associated with G, to the smaller dimensional space, associated with one of its g_{MNR}. The second part of the section deals with general properties of the projection matrix from G to g in the orthogonal basis, the projection of the G basis,[5] λ_i, to the basis μ_a, the corresponding projection $\alpha(r,l) \to \beta(r)$ for the roots of G and g, respectively, and the relation of the generators $e_{\beta(r)}$ to $E_{\alpha(r,l)}$. Conditions on the expansion coefficients of $e_{\beta(r)}$ in terms of $E_{\alpha(r,l)}$, which follow from the commutation relations of the subalgebra, are given. These conditions are obtained directly with the use of Dynkin-like diagrams and tables of scalar products of the $\alpha(r,l)$.

Section 5 is a summary of our results in tabular form. (If a reader is not interested in a detailed discussion of the methods used to obtain them, Sec. 4 and the specific illustrations on the methods of Sec. 4 in Appendices A and B may be omitted.) In particular, in Sec. 5, we list relevant properties of all MNR subalgebras of all exceptional Lie algebras, without providing any further details of calculations. Some of our results for the expansion of $e_{\beta(r)}$ in terms of $E_{\alpha(r,l)}$ appear to differ significantly from those of Dynkin. Most of these differences are due either to different possible choices of phases of the structure constants of G or to the fact that the equations for the expansion coefficients do not determine them uniquely. In Appendix A we discuss in detail two cases where our results differ significantly from those of Dynkin. These differences are due to errors in Dynkin's work. The two cases are $E(6) \supset G(2)$ and $E(8) \supset A^{1040}(1)$ [$A^{520}(1)$ in Dynkin's notation]. They also serve to illustrate our techniques. In Appendix B we consider in detail the cases of $D(4) \supset A(2)$ and $E(8) \supset A(2) \oplus A(1)$ as examples, in which, by a suitable choice of notation, one can easily find the expansion of all the generators $e_{\beta(r)}$ (not just the simple ones) in terms of $E_{\alpha(r,l)}$. The case $E(8) \supset A(2) \oplus A(1)$ also provides an example for which we derive an alternative solution to that given by Dynkin. In Sec. 6, we conclude with a brief discussion and, in particular, contrast our approach, which is a gloss on the approach of Dynkin and is firmly based upon it, with the original approach of Dynkin.

2. SIMPLE LIE ALGEBRAS

The simple Lie algebras, G, can all be expressed in the Dynkin basis[1-3] (FFM) as

[a)] Supported in part by the National Science Foundation.

 0022-2488/84/051230-23$02.50

$$[\mathbf{H},\mathbf{H}] = 0, \tag{2.1}$$

$$[\boldsymbol{\alpha}(p)\cdot\mathbf{H}, E_{\alpha(q)}] = \boldsymbol{\alpha}(p)\cdot\boldsymbol{\alpha}(q)E_{\alpha(q)}, \tag{2.2}$$

$$[E_{\alpha(p)}, E_{-\alpha(p)}] = \boldsymbol{\alpha}(p)\cdot\mathbf{H}, \tag{2.3}$$

$$[E_{\alpha(p)}, E_{\alpha(q)}] = N_{\alpha(p),\alpha(q)}E_{\alpha(m)}, \tag{2.4}$$

where $N_{\alpha(p),\alpha(q)}$ is nonvanishing for[6]

$$\boldsymbol{\alpha}(m)\equiv\boldsymbol{\alpha}(p) + \boldsymbol{\alpha}(q)$$

a root. The vector $\mathbf{H}$ has n components, where n is the *rank* of the algebra. The different algebras are distinguished by the $N - n$, n-dimensional vectors $\pm\boldsymbol{\alpha}(p)$, the *roots* of the algebra, and N is the number of its generators. The operator $\mathbf{H}$ behaves as the position operator in the space of the roots. Thus, if $\boldsymbol{\Lambda}$ is a position vector[7]—*weight*—in the space of the roots,

$$\mathbf{H}|\boldsymbol{\Lambda}\rangle = \boldsymbol{\Lambda}|\boldsymbol{\Lambda}\rangle. \tag{2.5}$$

The operators $E_{\pm\alpha(p)}$ are step up (step down) operators in weight space,

$$E_{\pm\alpha(p)}|\boldsymbol{\Lambda}, \Phi\rangle = C(\boldsymbol{\Lambda},\boldsymbol{\alpha}(p),\Phi)|\boldsymbol{\Lambda}\pm\boldsymbol{\alpha}(p), \Phi\rangle. \tag{2.6}$$

The weights of the finite-dimensional representations of the algebra are regular finite lattices of points in weight space, which can be generated from each other by the successive action of the operators $E_{\pm\alpha(p)}$. Thus, the roots $\pm\boldsymbol{\alpha}(p)$ are the lattice vectors of the representations, and the boundaries of the representation are determined by the vanishing of the factors $C(\boldsymbol{\Lambda},\boldsymbol{\alpha},\phi)$. It is evident from Eqs. (2.2)–(2.6) that $\mathbf{H}$, $E_{\boldsymbol{\alpha}}$, N^{m}_{pq}, and C, like $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$, have the dimension of "length" in weight space. Thus, for a single group, the scale of length in weight space can be chosen arbitrarily. It is usual to express all quantities in terms of the length of the longest root.[8] Finite dimensional representations of Lie groups are unitary, and thus the operators $\mathbf{H}$ can be taken as Hermitian:

$$\mathbf{H} = \mathbf{H}\dagger. \tag{2.7}$$

The roots $\boldsymbol{\alpha}(p)$ are real, so it follows from (2.3) that the operators $E_{\alpha(p)}$ can be chosen to satisfy[9]

$$E_{-\alpha(p)} = E^{\dagger}_{\alpha(p)}. \tag{2.8}$$

One of the set of conditions for the structure constants which follow from Jacobi identities [see FFM Eq. (2.9)] is

$$N_{\alpha(p),\alpha(q)}N_{-\alpha(p),-\alpha(q)} = -|N_{\alpha(p),\alpha(q)}|^2. \tag{2.9}$$

We will take $N_{\alpha(p),\alpha(q)}$ to be real. It follows that[10]

$$N_{-\alpha(p),-\alpha(q)} = -N_{\alpha(p),\alpha(q)}. \tag{2.10}$$

3. REGULAR SUBALGEBRAS

A subalgebra g of a Lie algebra G exists when a projection of the lattice points of the representations in the weight space of G is identical to the weights of the representations of g for a particular choice of the relative length scales. Since some of the roots, $\boldsymbol{\alpha}(p)$, of G project into the roots $\boldsymbol{\beta}(r)$ of g, the length scale of g (in terms of the longest root $|\boldsymbol{\beta}_L|$) is determined by the projection in terms of $|\boldsymbol{\alpha}_L|$, the scale of G. The square of this ratio of scales,

$$j\equiv|\boldsymbol{\alpha}_L|^2/|\boldsymbol{\beta}_L|^2, \tag{3.1}$$

is a defining property of the subgroup, known as the index.[11]

In discussing the subalgebras of Lie algebras, it is clearly sufficient to consider only maximal subalgebras g, namely, those which are not, themselves, subalgebras of other subalgebras of G. Suppose that the subalgebra g has roots $\boldsymbol{\beta}(r)$ and elements $e_{\pm\beta(r)}$. Regular subalgebras are those for which

$$e_{\pm\beta(r)} = E_{\pm\alpha(r)}, \tag{3.2}$$

where, the labels r are a subset of the labels p so that the set $\{e_{\pm\beta(r)}\}$ is a subset of $\{E_{\pm\alpha(p)}\}$ and the $\{\pm\boldsymbol{\beta}(r)\}$ is a subset of $\{\pm\boldsymbol{\alpha}(p)\}$. The maximal regular subalgebras have the same rank as G and index j equal to unity. Simple and powerful methods have been developed by Dynkin[2] for deriving these regular subalgebras from the algebra of G, by removal of vertices in his extended diagrams.[12] The rules apply to both classical and exceptional algebras and we have nothing to add on this topic.

For our discussion of both regular and nonregular subalgebras, it is convenient to set up an orthogonal basis in weight space. In the algebras $B(n)$, $C(n)$, and $D(n)$, the positive weights of the representation of smallest dimension—the quarks—form just such a basis with weights $\boldsymbol{\lambda}_i$ (see Table I of FFM), where

$$\boldsymbol{\lambda}_i\cdot\boldsymbol{\lambda}_j = \delta_{ij}\,\lambda^2, \quad i,j = 1,...,n. \tag{3.3}$$

For $A(n)$, the weights $\boldsymbol{\chi}_i$ of the $n+1$ quarks are not orthogonal but we can set up an $n+1$ dimensional orthogonal basis $\boldsymbol{\lambda}_i$ (see FFM) such that

$$\boldsymbol{\chi}_i = \boldsymbol{\lambda}_i - \sigma, \quad i = 1,...,n+1, \tag{3.4}$$

where

$$\boldsymbol{\sigma}\equiv\frac{1}{n+1}\sum_{j=1}^{n+1}\boldsymbol{\lambda}_j. \tag{3.5}$$

Note that

$$\boldsymbol{\chi}_i\cdot\boldsymbol{\sigma} = 0. \tag{3.6}$$

Equation (3.6) implies that the weight space of $A(n)$ is the n-dimensional subspace of the $(n+1)$-dimensional $\boldsymbol{\lambda}$-space which is orthogonal to $\boldsymbol{\sigma}$. For the exceptional algebras, the number of quarks greatly exceeds the rank of the algebra, and it is convenient to relate the orthogonal basis to a maximal regular subalgebra which is one of the classical algebras. There are alternative possibilities, but, to avoid the complication of quarks with nonorthogonal weights, we avoid subalgebras which involve $A(n)$ (for $n > 1$). Thus, for example, the orthogonal basis for $E(6)$ is related to the simplest representation of $D(5)\oplus U(1)$ rather than $A(5)\oplus A(1)$. The derivation of these bases is discussed in Appendix A of FFM, and the results are presented in Tables I and II there. The quarks and roots (nonzero weights of the adjoint representation) are given in terms of orthogonal basis vectors $\boldsymbol{\lambda}_i$ for all simple Lie algebras in these tables.

For the classical algebras [even with the slight complication present in $A(n)$, discussed above] we can construct the maximal regular subalgebras by focusing on the direct correspondence between the orthonormal basis vectors and the quarks.

The first stage of the construction is to segregate the quark weights of G into two disjoint sets. The quarks of G go into the quarks of the (in general nonsimple) g. If G is a classical algebra, g will be a classical algebra or a direct sum of two classical algebras. There are *two* potential complica-

tions. The first occurs if one of the two subspaces of the original space is spanned by a subset of $k+1$ λ''s$(\lambda_1',...,\lambda_{k+1}')$ of the original λ's $(\lambda_1,\lambda_2,...,\lambda_n)$ for $B(n)$, $C(n)$, $D(n)$ or $(\lambda_1\cdots\lambda_{n+1})$ for $A(n)$, which are in fact used to construct $A(k)$. In this case, we want a k-dimensional subspace of this $k+1$ dimensional space, defined as the subspace orthogonal to the vector $\mathbf{k}'$, where

$$\mathbf{k}' = \sum_{a=1}^{k+1} \lambda_a'. \tag{3.7}$$

The second complication arises from the fact that the quarks of $C(n)$ and $D(n)$ have the same weight structure. In this case, the geometry of the weight space is not sufficient to determine the maximal subalgebras, and we must observe the restrictions due to the structure of the algebras. These restrictions become immediately apparent from the commutation relations, if the roots are written in terms of the orthogonal λ's (see FFM).

Thus, the maximal regular subalgebras of the classical algebras are[13,14]

$$A(n): \begin{cases} A(n-1)\oplus U(1), \\ A(n-k-1)\oplus A(k)\oplus U(1), \quad k=1,2,...,[(n-1)/2], \end{cases} \tag{3.8}$$

where $[(n-1)/2]$ is the next lowest integer to the number indicated,

$$B(n): \begin{cases} D(n), \\ D(n-k)\oplus B(k), \quad k=1,...,n-1, \end{cases} \tag{3.9}$$

$$C(n): \begin{cases} A(n-1)\oplus U(1), \\ C(n-k)\oplus C(k), \quad k=1,...,[(n-1)/2)], \end{cases} \tag{3.10}$$

$$D(n): \begin{cases} A(n-1)\oplus U(1), \\ D(n-k)\oplus D(k), \quad k=1,...,[(n-1)/2]. \end{cases} \tag{3.11}$$

[Note that $B(n-1)\oplus U(1)$ is included in (3.9) and $D(n-1)\oplus U(1)$ in (3.11) since $D(1)$ is isomorphic to $U(1)$.]

Geometrically, we can think of the procedure carried out to obtain the maximal subalgebras of the classical algebras as a projection to subspaces where the orthogonal basis, labeling the axes, is kept fixed. The same argument as the one just given for the classical algebras does not apply to the exceptional algebras as there is no simple relation between the weights of the quarks and the basis vectors λ_i of the orthogonal base. Nevertheless, for the sake of completeness, we list the maximal regular subalgebras of the exceptional algebras below[13,14]:

$$E(6): \begin{cases} D(5)\oplus U(1), \\ A(5)\oplus A(1), \\ A(2)\oplus A(2)\oplus A(2), \end{cases} \tag{3.12}$$

$$E(7): \begin{cases} E(6)\oplus U(1), \\ D(6)\oplus A(1), \\ A(7), \\ A(5)\oplus A(2), \end{cases} \tag{3.13}$$

$$E(8): \begin{cases} E(7)\oplus A(1), \\ E(6)\oplus A(2), \\ D(8), \\ A(8), \\ A(4)\oplus A(4), \end{cases} \tag{3.14}$$

$$F(4): \begin{cases} B(4), \\ C(3)\oplus A(1), \\ A(2)\oplus A(2), \end{cases} \tag{3.15}$$

$$G(2): \begin{cases} A(2), \\ A(1)\oplus A(1). \end{cases} \tag{3.16}$$

4. NONREGULAR SUBALGEBRAS (S-SUBALGEBRAS)

In this section, we describe the procedures which lead to the main results of this paper, namely the expressions, given in Tables I–VI, for the $e_{\beta(r)}$ in terms of $E_{\alpha(p)}$. The $e_{\beta(r)}$ are the generators of the maximal nonregular subalgebras g and the $E_{\alpha(p)}$ are the generators of the algebras G. To obtain this relationship, we must find which set of roots[15] $\alpha(r,l)$ of G project into a given root $\beta(r)$ of g. This, in turn, requires a knowledge of the projection matrix which takes a given representation of G into a respresentation of g. In the conventional techniques (described by Dynkin[2] and used by McKay and Patera[13] and McKay, Patera, and Sankoff[16]), the matrix which projects the basic weights (as defined by Dynkin[2]) of G into linear combinations of basic weights of g is given. These weights are n_G or n_g in number (where n_G and n_g are the respective ranks) and span the two weight spaces. However, these weights are not orthogonal and therefore the projection matrix is not an orthogonal one. In contrast, our approach is to make use of the orthogonal basis vectors λ_i and a knowledge of how the quarks of G project. Because of the direct relationship between quark weights and the basis vectors for the classical algebras this projection matrix becomes trivially obvious. In the case of the exceptional algebras, the construction of this matrix is only slightly more difficult than for the classical algebras. The procedure is described below. One of the important quantities required for the construction is the scale factor, also derived and defined below.

The rank n_g of an MNR subalgebra g of an algebra G is less than n_G, the rank of G. Denote the orthogonal basis of g by μ_a, $(a=1,...,n_g)$, with

$$\mu_a\cdot\mu_b = \delta_{ab}\,\mu^2. \tag{4.1}$$

The space of g is a subspace of the space G, spanned by λ_i $(i=1,...,n_G)$, but the basis vectors μ_a are rotated with respect to the vectors λ_i, and, as pointed out above, the length scale in the two spaces may be different. That is to say, the "scale factor" [closely related to the index j, Eq. (3.1)], which we define as

$$s^2 = \lambda^2/\mu^2, \tag{4.2}$$

may be different from unity.

A. Scale factor

Suppose the decomposition of an irreducible representation ϕ of G in terms of irreducible representations ψ_w of g to be

$$\phi \rightarrow \psi_1\oplus\psi_2\oplus\cdots\oplus\psi_m \equiv \psi. \tag{4.3}$$

Let Λ be the weight of an element of ϕ and H_i be a component[17] of the vector $\mathbf{H}$ $(i=1,...,n_G)$. Then,

$$H_i|\Lambda,\phi\rangle = \Lambda_i|\Lambda,\phi\rangle, \quad i=1,...,n_G. \tag{4.4}$$

The number of weights Λ (including the zero weights) is the dimension of the representation. The subgroup g will have

n_g of the H_i in common with G. For these H_i, we can also write

$$H_i|\Lambda,\psi\rangle = \Lambda_i|\Lambda,\psi\rangle, \quad i=1,...,n_g. \tag{4.5}$$

It follows that

$$\langle\Lambda,\phi|H_iH_j|\Lambda,\phi\rangle = \Lambda_i\Lambda_j, \quad i,j=1,...,n_G, \tag{4.6}$$

$$\langle\Lambda,\psi|H_iH_j|\Lambda,\psi\rangle = \Lambda_i\Lambda_j, \quad i,j=1,...,n_g, \tag{4.7}$$

where the state vectors of the representations are all normalized to unity. Let us sum (4.6) and (4.7) over all[18] Λ in ϕ and ψ, respectively. That is, we take the trace of H_iH_j in these representations. Thus,

$$\sum_{\Lambda}^{\{\phi\}} \Lambda_i\Lambda_j = \sum_{\Lambda}^{\{\phi\}} \langle\Lambda,\phi|H_iH_j|\Lambda,\phi\rangle = \mathrm{Tr}\, H_iH_j = c^{\phi}\delta_{ij}, \quad i,j=1,...,n_G, \tag{4.8}$$

and

$$\sum_{\Lambda}^{\{\psi\}} \Lambda_i\Lambda_j = \sum_{\Lambda}^{\{\psi\}} \langle\Lambda,\psi|H_iH_j|\Lambda,\psi\rangle = \mathrm{Tr}\, H_iH_j = c^{\psi}\delta_{ij}, \quad i,j=1,...,n_g, \tag{4.9}$$

where the c^{ϕ} (c^{ψ}) are constants[19] which depend on G, ϕ (g, ψ). Traces are independent of the state vectors which span the representation. Thus, the space spanned by the $|\Lambda,\phi\rangle$ is the same as that spanned by $|\Lambda,\psi\rangle$. This allows us to deduce from (4.8) and (4.9) that

$$c^{\phi} = c^{\psi}. \tag{4.10}$$

We use (4.8) and (4.9) again by setting $i=j$ and summing over i in each case. We obtain[20,21]

$$c^{\phi}n_G = \sum_{\Lambda}^{\{\phi\}} \sum_{i=1}^{n_G} \Lambda_i^2 \tag{4.11}$$

and

$$c^{\psi}n_g = \sum_{\Lambda}^{\{\psi\}} \sum_{i=1}^{n_g} \Lambda_i^2. \tag{4.12}$$

If the weights in G are expressed in terms of λ^2 and those in g in terms of μ^2, we get from (4.10)–(4.12)

$$s^2 \equiv \frac{\lambda^2}{\mu^2} = \frac{n_G}{n_g}\left(\sum_{\Lambda}^{\{\psi\}}\sum_{i=1}^{n_g}\frac{\Lambda_i^2}{\mu^2}\Big/\sum_{\Lambda}^{\{\phi\}}\sum_{i=1}^{n_G}\frac{\Lambda_i^2}{\lambda^2}\right). \tag{4.13}$$

If the subalgebra is nonsimple and is of the form

$$G \supset g_1 \oplus g_2, \tag{4.14}$$

where g_ω has rank n_ω ($\omega = 1,2$), then the same argument may be applied to a representation ϕ which decomposes as

$$\phi \to (\psi_1^{(1)}, \psi_1^{(2)}) \oplus (\psi_2^{(1)}, \psi_2^{(2)}) \oplus \cdots \oplus (\psi_m^{(1)}, \psi_m^{(2)}). \tag{4.15}$$

We must now distinguish the components of $\mathbf{H}$ in g_1, from the components of $\mathbf{H}$ in g_2. One can show that

$$c^{\phi} = c^{\psi_1} = c^{\psi_2} = \frac{\Sigma_{\{\phi\}}^{\Lambda}\Sigma_{i=1}^{n_G}\Lambda_i^2}{n_G} = \frac{\Sigma_{\{\psi\}}^{\Lambda}\Sigma_{i=1}^{n_\omega}\Lambda_i^2}{n_\omega}, \quad \omega = 1,2, \tag{4.16}$$

so that

$$s_\omega^2 = \frac{\lambda^2}{(\mu^\omega)^2} = \frac{n_G}{n_\omega}\frac{\Sigma_{\Lambda}^{\{\psi\}}\Sigma_i^{n_\omega}[\Lambda_i^2/(\mu^\omega)^2]}{\Sigma_{\Lambda}^{\{\phi\}}\Sigma_i^{n_G}(\Lambda_i^2/\lambda^2)}. \tag{4.17}$$

B. The projection matrix

Suppose that, in the projection of the weight space G into that of g, the basis vectors $\boldsymbol{\lambda}_i$ of G project according to

$$\boldsymbol{\lambda}_i \to \sum_a M_{ia}\,\boldsymbol{\mu}_a, \quad i=1,...,n_G, \ a=1,...,n_g, \tag{4.18}$$

where the $\boldsymbol{\mu}_a$ label the orthogonal basis vectors in g. Since the basis vectors, λ_i and $\boldsymbol{\mu}_a$, each form orthogonal sets, the rectangular matrix

$$M_{ia}/s \tag{4.19}$$

may be interpreted as the first n_g columns of an $n_G \times n_G$ orthogonal matrix,[22] which rotates the orthonormal system $\boldsymbol{\lambda}_i/\sqrt{\lambda^2}$, to another orthonormal system $\boldsymbol{\mu}_i/\sqrt{\mu^2}$, where the n_g vectors $\boldsymbol{\mu}_a$ in (4.18) are a subset of the vectors $\boldsymbol{\mu}_i$. Thus, we would need an additional $n_G - n_g$ of the $\boldsymbol{\mu}$ vectors to span the whole space of G. Accordingly, the projection of G into g is interpreted geometrically as the rotation of the orthogonal $\boldsymbol{\lambda}_i$ basis to the orthogonal $\boldsymbol{\mu}_i$ basis, followed by setting the coefficients multiplying the additional $(n_G - n_g)$ $\boldsymbol{\mu}$ vectors to zero. Thus, the columns of M_{ia} possess the orthogonal property

$$\sum_{i=1}^{n_G} M_{ia}M_{ib} = s^2\delta_{ab}, \quad a,b=1,...,n_g. \tag{4.20}$$

If g is semisimple, the normalization condition (4.20) must be altered so that the scaling factor s^2 is that associated with the subalgebra factor over which the a,b labels range.

Although the quantities (4.19) are elements of an orthogonal matrix, they are not all the elements. Nevertheless, they allow us to determine the $\boldsymbol{\mu}_a$ in terms of the $\boldsymbol{\lambda}_i$ (but, of course, not vice versa):

$$\boldsymbol{\mu}_a = \frac{1}{s^2}\sum_{i=1}^{n_G}\boldsymbol{\lambda}_i M_{ia}. \tag{4.21}$$

Since we know the roots $\boldsymbol{\beta}(r)$ of g in terms of the $\boldsymbol{\mu}_a$ (see Tables I and II of FFM), we easily determine the $\boldsymbol{\beta}(r)$ in terms of the $\boldsymbol{\lambda}_i$, associated with G. In addition, a knowledge of the $\boldsymbol{\alpha}(p)$ (in terms of the $\boldsymbol{\lambda}_i$) and of the elements M_{ia} allows us to find immediately which set of roots of G [call them $\boldsymbol{\alpha}(r,l)$] project into a given root $\boldsymbol{\beta}(r)$ (labeled by r) of g. That is, we can obtain

$$\boldsymbol{\alpha}(r,l) \to \boldsymbol{\beta}(r), \quad l=1,...,L. \tag{4.22}$$

The value of L is given by the number of times that the given weight $\boldsymbol{\beta}(r)$ appears in the decomposition[13] of the adjoint representation of G in terms of representations of g. This follows because every weight in a representation, ϕ, of G projects into a weight of the representation, ψ, of g.

The projection matrix, M_{ia}, is constructed to ensure that the relations (4.18) project weights of quarks, $(\mathbf{q}_x, x=1,...,N_q)$ in G into weights $(\mathbf{w}_x^{(q)})$ of the corresponding representation, ψ, of the subalgebra g in such a way that the conditions (4.20) are satisfied. Thus,

$$\mathbf{q}_x \equiv \sum_{i=1}^{n_G} Q_{xi}\,\boldsymbol{\lambda}_i \to \mathbf{w}_x^{(q)} = \sum_{a=1}^{n_g} W_{xa}\,\boldsymbol{\mu}_a. \tag{4.23}$$

When G is one of the classical algebras $B(n)$, $C(n)$, $D(n)$, the steps outlined above for constructing M_{ia} are very simple to carry out. This is so, since the n_G $\boldsymbol{\lambda}_i$ themselves are the weights of quarks (more precisely, one-half the total number of nonzero quark weights) and project directly into weights of the subalgebra. There is no complication when G is $A(n)$ [or[23] $G(2)$], even though the quark weights are not the $\boldsymbol{\lambda}_i$ but [see (3.4)]

$$\boldsymbol{\lambda}_i - \frac{1}{n+1}\sum_{j=1}^{n+1}\boldsymbol{\lambda}_j, \quad i = 1,...,n+1. \tag{4.24}$$

We can choose the projection such that

$$\sum_{j=1}^{n+1}\boldsymbol{\lambda}_j \to 0, \tag{4.25}$$

so that (4.23) will be given by (4.18), where the $\Sigma_{a=1}^{n_g} M_{ia}\,\boldsymbol{\mu}_a$ are a set of the $\mathbf{w}_x^{(q)}$ weights.

There is, in constrast, a complication when $A(n)$ [or[23] $G(2)$] is a subalgebra, g, and we embed the n_g dimensional space of g into the very convenient $n_g + 1$ subspace defined by Eqs. (3.4) and (3.5). For example, if g is $A(2)$, the projection $(G \to g)$ is onto a two-dimensional space which we embed in a three-dimensional space with basis vectors $\boldsymbol{\mu}_a$ $(a = 1,2,3)$. The weights $\mathbf{w}_x^{(q)}$ are, however, expressed in terms of the three quark weights of g, $\boldsymbol{\chi}_a$, where

$$\boldsymbol{\chi}_a = \boldsymbol{\mu}_a - \frac{1}{3}\sum_{b=1}^{3}\boldsymbol{\mu}_b, \tag{4.26}$$

and, of course,

$$\sum_{a=1}^{3}\boldsymbol{\chi}_a = 0. \tag{4.27}$$

This implies that the quantities W_{xa} in (4.23) will be constrained to satisfy the subsidiary conditions

$$\sum_{a=1}^{3} W_{xa} = 0. \tag{4.28}$$

Although the projection of the basis vector $\boldsymbol{\lambda}_i$ onto the three-dimensional subspace will be of the form given by (4.18) (where the M_{ia}/s are elements of an orthogonal matrix), the projection onto the two-dimensional subspace will be given in terms of the quantities

$$\bar{M}_{ia} = M_{ia} - \frac{1}{3}\sum_{b=1}^{3} M_{ib}. \tag{4.29}$$

This is the case, since $\bar{M}_{ia}$ are linear combinations of the W_{xa} which satisfy (4.28). Using Eqs. (4.21) and (4.26), we have

$$\boldsymbol{\chi}_a = \frac{1}{s^2}\sum_{i=1}^{n_G}\boldsymbol{\lambda}_i\,\bar{M}_{ia}, \tag{4.30}$$

from which it follows [again using (4.26)] that

$$\sum_i \bar{M}_{ia}\bar{M}_{ib} = s^2(\delta_{ab} - \tfrac{1}{3}), \quad a,b = 1,2,3. \tag{4.31}$$

We shall call the quantities $\bar{M}_{ia}/s$ elements of a pseudo-orthogonal matrix.[24]

When G is an exceptional algebra,[25] the construction of the projection matrix is only slightly more difficult than for the classical algebras. In this case, there are many more quark weights $\mathbf{q}_x$ than there are basis vectors $\boldsymbol{\lambda}_i$. However, in our basis, which makes use of the orthogonal subalgebras of the exceptional algebras, the "vector" weights[26] of q are closely related to the basis vectors. Thus we have for the "vector" weights (see Table II of FFM) of[27]

$$F(4): \quad \pm\boldsymbol{\lambda}_i, \quad i = 1,...,4, \tag{4.32}$$

$$E(6): \quad \pm\boldsymbol{\lambda}_i - \bar{\bar{\lambda}}_6/\sqrt{3}, \quad i = 1,...,5, \quad 2\bar{\bar{\lambda}}_6/\sqrt{3}, \tag{4.33}$$

$$E(7): \quad \pm\boldsymbol{\lambda}_i \pm \bar{\lambda}_7/\sqrt{2}, \quad i = 1,...,6, \tag{4.34}$$

$$E(8): \quad \pm\boldsymbol{\lambda}_i \pm \boldsymbol{\lambda}_j, \quad i,j = 1,...,8. \tag{4.35}$$

From a knowledge of the quark decomposition [i.e., of the quantities W_{xa}, defined in (4.23) and listed in Tables I–VI, in section (ii)], we must find a set of $n_G \times n_g$ numbers with which to construct the elements of the orthogonal (or psuedo-orthogonal) matrix. This set is selected so as to satisfy the conditions (4.20) [or (4.31)]. It is far from unique, but simple solutions can be found almost by inspection. One aid in this construction is to examine the weights $\mathbf{w}^{(\alpha')}$ of g into which the adjoint of G projects. A trial set of M_{ia} must have the property that $\boldsymbol{\lambda}_i + \boldsymbol{\lambda}_j$ (which are always some of the adjoint weights) do not project into weights larger than the maximum of the weights $\mathbf{w}^{(\alpha')}$. Another useful ansatz is to let $2\bar{\bar{\lambda}}_6/\sqrt{3}$ for $E(6)$, $\boldsymbol{\lambda}_6 - \bar{\lambda}_7/\sqrt{2}$ for $E(7)$, and $\boldsymbol{\lambda}_7 - \boldsymbol{\lambda}_8$ for $E(8)$, project into one of the smallest weights $\mathbf{w}_x^{(q)}$. For the simple subalgebras, we can also take $\boldsymbol{\lambda}_6 \to 0$ for $E(7)$ and $\boldsymbol{\lambda}_7$ (or $\boldsymbol{\lambda}_8$) $\to 0$ for $E(8)$. For the nonsimple cases, on projection these vectors can be taken to have no components for at least one of the semisimple factors. Note from the Tables IVA–IVE (ii) that all of the simple maximal regular subalgebras of $E(6)$ contain three zero weights $\mathbf{w}_x^{(q)}$, which means that we could (and did) take

$$\boldsymbol{\lambda}_5, \bar{\bar{\lambda}}_6 \to 0 \tag{4.36}$$

for all of these cases. Similar arguments can be used for the $E(7)$ and $E(8)$ subgroups, although for $E(7)$ we cannot[28] take

$$\bar{\lambda}_7 \to 0, \tag{4.37}$$

since $\pm\sqrt{2}\bar{\lambda}_7 \equiv \pm\boldsymbol{\lambda}_7$ are roots. In these cases it is useful to choose $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, say, to project into weights with the largest possible W_{xa} consistent with (4.20) [or (4.31)] and such that $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2$ projects into a weight which is the largest weight into which the adjoint projects. For the $A(1)$ MNR of $E(7)$ and $E(8)$, we make an additional choice, which allows a simplification in the notation for the roots $\boldsymbol{\alpha}(a,l)$. Once the projections of $\boldsymbol{\lambda}_6$, $\bar{\lambda}_7$ for $E(7)$ and $\boldsymbol{\lambda}_7$, $\boldsymbol{\lambda}_8$ for $E(8)$ have been chosen as outlined above, we choose for the projection of $\boldsymbol{\lambda}_1$ the largest possible M_{ia} consistent with (4.20). The rest of the matrix is then uniquely determined.

As mentioned earlier, once we construct the projection matrix, we can find directly the set of roots, $\boldsymbol{\alpha}(r,l)$ of G which project into $\boldsymbol{\beta}(r)$, a root of g. We can now express the generators $e_{\beta(r)}$, of g as linear combinations of those of G $(E_{\alpha(r,l)})$:

$$e_{\beta(r)} = \frac{1}{s'(r)}\sum_{l=1}^{L} B(r,l)E_{\alpha(r,l)}, \tag{4.38}$$

where the coefficients $B(r,l)$ are determined so as to satisfy the algebra of g. The quantity $s'(r)$, which is directly related to the scaling factor s is extracted so as to make the matrix

which determines the $B(r,l)$ coefficients particularly simple. We take[29]

$$[s'(r)]^2 = s^2 2\mu^2/\beta^2(r). \tag{4.39}$$

In keeping with our physical assumption concerning step-up and step-down operators, we take

$$e_{-\beta(r)} = \frac{1}{s'(r)} \sum_{l=1}^{L} B^*(r,l) E_{-\alpha(r,l)}. \tag{4.40}$$

To determine the coefficients $B(r,l)$, we must substitute (4.38) and (4.40) into the commutation relations (2.2)–(2.4), for the group g. From the first set (2.2)

$$[\beta(t)\cdot\mathbf{H}, e_{\beta(r)}] = \beta(t)\cdot\beta(r) e_{\beta(r)}, \tag{4.41}$$

we find

$$\beta(t)\cdot\alpha(r,l) = \beta(t)\cdot\beta(r), \quad \text{all } l, \tag{4.42}$$

and, in particular,

$$\beta(r)\cdot\alpha(r,l) = \beta^2(r), \quad \text{all } l. \tag{4.43}$$

From the set of relations (2.3),

$$[e_{\beta(r)}, e_{-\beta(r)}] = \beta(r)\cdot\mathbf{H}, \tag{4.44}$$

we have

$$\beta(r)\cdot\mathbf{H} = \frac{1}{[s'(r)]^2} \left\{ \sum_{l=1}^{L} |B(r,l)|^2 \alpha(r,l)\cdot\mathbf{H} + \sum_{l\neq l'} B(r,l) B^*(r,l') [E_{\alpha(r,l)}, E_{-\alpha(r,l')}] + \text{c.c.} \right\}. \tag{4.45}$$

Two sets of relations follow from this condition, the first being

$$\frac{1}{[s'(r)]^2} \sum_{l=1}^{L} |B(r,l)|^2 \alpha(r,l) = \beta(r). \tag{4.46}$$

Multiplying by $\alpha(r,m)$ and using (4.39) and (4.43), we obtain[30]

$$\sum_{l=1}^{L} \alpha(r,m)\cdot\alpha(r,l) |B(r,l)|^2 = 2\lambda^2, \quad \text{all } m. \tag{4.47}$$

Another sum rule for the $|B(r,l)|^2$ is obtained by multiplying (4.46) by $\beta(r)$ and using (4.43):

$$\sum_{l=1}^{L} |B(r,l)|^2 = [s'(r)]^2. \tag{4.48}$$

For $E(6)$, $E(7)$, $E(8)$,

$$\begin{aligned} \alpha(r,m)\cdot\alpha(r,l) &= 2\lambda^2, \quad \text{for } l = m, \\ &= \pm\lambda^2, 0, \quad \text{for } l \neq m, \quad l,m = 1,\ldots,L. \end{aligned} \tag{4.49}$$

For $F(4)$, the values $\pm\frac{1}{2}\lambda^2$ are also possible.

If the set of L roots $\alpha(r,l)$ for a given r are a subset of simple roots, the scalar products in (4.49) for $l \neq m$ must be 0 or $-\lambda^2$. A set or subset of simple roots are all independent, which in turn implies that Eqs. (4.47) determine $|B(r,l)|^2$ uniquely. The implication of scalar products in (4.49) with values $+\lambda^2[+\frac{1}{2}\lambda^2$ for $F(4)]$ is that the set of L roots $\alpha(r,l)$ are not independent. In this case, the set of equations (4.47) do not uniquely determine the $|B(r,l)|^2$. The property (4.49) is indicated in the Tables I–VI (iii) in the manner of Dynkin. The roots $\alpha(r,l)$ are labeled by closed or open circles. Two roots joined by a single continuous line have a scalar product of $-\lambda^2$. Roots not directly joined, or not joined at all, are orthogonal. In addition, we have introduced a new notation: The dashed line connecting two roots indicates that their scalar product is $+\lambda^2$.

There will be a second set of relations which follow from (4.45) if for some pairs of roots $\alpha(r,l_v)$ and $\alpha(r,l'_v)$ (v labels each pair)

$$\alpha(r,l_v) - \alpha(r,l'_v) = \alpha(x),$$

where $\alpha(x)$ is a root (not in the set r), which is the same for all v. For such a case, we must have

$$\sum_v B(r,l_v) B^*(r,l'_v) N^x_{l_v l'_v} = 0, \tag{4.50}$$

where we have used the fact that

$$[E_{\alpha(r,l_v)}, E_{-\alpha(r,l'_v)}] = N^x_{l_v l'_v} E_{\alpha(x)}, \tag{4.51}$$

and where the $N^x_{l_v l'_v}$ are structure constants of G. We will delay the determination of the pairs of roots which contribute to (4.50), because similar arguments are used below where all the generators for simple roots are obtained. As shown by Dynkin, if one uses the commutation relations (4.41) and (4.44) for simple roots $[\beta(a)]$ only and the additional relations

$$[e_{\beta(a)}, e_{-\beta(b)}] = 0, \tag{4.52}$$

the entire subalgebra can be generated. If we substitute expressions (4.38) and (4.40) into (4.52), we obtain restrictions on the coefficients $B(a,l)$, $B(b,m)$ similar in form to Eq. (4.50):

$$\sum_v B(a,l_v) B^*(b,m_v) N^y_{l_v m_v} = 0, \tag{4.53}$$

where

$$\alpha(a,l_v) - \alpha(b,m_v) = \alpha(y). \tag{4.54}$$

Restricting our discussion, for the moment,[31] to the case where G is $E(6)$, $E(7)$, and $E(8)$ (all the roots α for these algebras have length $\alpha^2 = 2\lambda^2$), it follows on taking the square of (4.54) that, for every pair contributing to (4.53)

$$\alpha(a,l_v)\cdot\alpha(b,m_v) = +\lambda^2. \tag{4.55}$$

When two[32] or more pairs contribute to the same root $\alpha(y)$, we have

$$\alpha(a,l_1) - \alpha(b,m_1) = \alpha(a,l_2) - \alpha(b,m_2) = \alpha(y). \tag{4.56}$$

It follows from the fact that all the roots appearing in (4.56) have the same length $2\lambda^2$ and, taking squares of various combinations of the roots in (4.56), that the configuration of scalar products must be as displayed in either Fig. 1(a) or 1(b). The meaning of the dashed line connecting two roots has been discussed previously. If we make the substitution

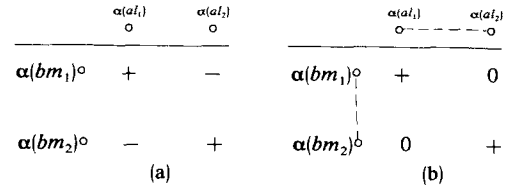


FIG. 1. Scalar products of roots contributing to Eq. (4.53). See text for explanation of symbols.

$$al_v \to rl_v, \quad bm_v \to rl'_v, \tag{4.57}$$

we find the configuration of scalar products for those roots which contribute to (4.50). These are displayed in Fig. 2(a) or 2(b).

We find that restrictions of the form (4.50) exist only for some of the subalgebras of E (7) and E (8) (see Tables VB, VIB, VIC–E). Such contributions are nontrivial only when one or more pairs of $\alpha(r,l)$ have scalar products which are $+\lambda^2$ [see (4.55)]. This is turn implies that not all of the $\alpha(r,l)$, for a given r, are linearly independent, from which it follows that the equations (4.47) do not uniquely determine the $|B(r,l)|^2$.

It is often useful (and straightforward) to generate the entire algebra (and not just the algebra of simple roots) directly. That is, we construct the commutation relations

$$[e_{\beta(r)}, e_{\beta(t)}] = n^u_{rt} e_{-\beta(u)}, \tag{4.58}$$

where $\beta(r)$, $\beta(t)$, $\beta(u)$ are any roots such that

$$\beta(r) + \beta(t) + \beta(u) = 0, \tag{4.59}$$

and the n^u_{rt} are the structure constants of the subgroup g. Substituting (4.38) into (4.58), we obtain in a manner similar to (4.50) and (4.53)

$$\sum_v B(r,l_v)B(t,m_v)N^k_{l_v m_v} = \bar{n}^u_{rt} B^*(u,k), \tag{4.60}$$

where the $\bar{n}$ are the structure constants of g (in units of λ), the magnitudes of which are given by Eqs. (2.17)–(2.21) of FFM. Further, we have

$$\alpha(r,l_v) + \alpha(t,m_v) = -\alpha(u,k), \quad \text{for all } v. \tag{4.61}$$

In this case, no more than one pair need contribute to an equation of form (4.60), and the pairs must satisfy the relation

$$\alpha(r,l)\cdot\alpha(t,m) = -\lambda^2. \tag{4.62}$$

If more than one pair contributes to the same $\alpha(u,k)$, its configuration of scalar products [shown in Fig. 3(a) or 3(b)] is easily obtained from Fig. 2(a) or 2(b) by simply letting

$$\alpha(a,l_v) \to \alpha(r,l_v), \quad \alpha(b,m_v) \to -\alpha(t,m_v). \tag{4.63}$$

As previously mentioned, Dynkin has shown[2] that it is sufficient to determine coefficients $B(a,l)$, which satisfy Eqs. (4.50) and (4.53) for the simple roots. The entire subalgebra can then be generated from the algebra of the simple roots using (4.58) as the definition of $e_{-\beta(u)}$. The phases of the structure constants N^k_{lm}, of the parent algebra, which appear in these equations can be chosen arbitrarily since they are not restricted by either linear constraints [Eqs. (2.7)] of FFM or bilinear constraints [Eqs. (2.8) of FFM]. Dynkin[2] exploits this freedom. The resulting phases generated in this way are consistent for the algebra and subalgebra under considera-

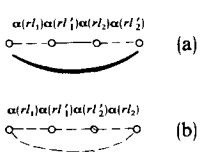


FIG. 2. Scalar products of roots contributing to Eq. (4.50). See text for explanation of symbols.

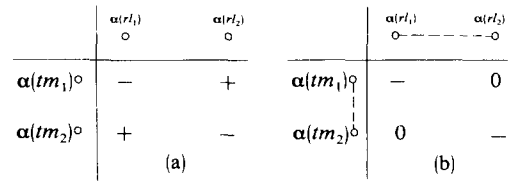


FIG. 3. Scalar products for which more than one pair of roots of type (4.62) contributes to Eq. (4.60). See text for explanation of symbols.

tion but do not necessarily imply a simultaneously consistent phase convention for the treatment of different subalgebras.

In our approach, we have systematically used the phase conventions established in FFM for both the parent algebras and the subalgebras, so that the relation

$$e_{-\beta} = e^\dagger_\beta \tag{4.64}$$

is maintained and all solutions are consistent with each other.

5. SUMMARY OF RESULTS

We are now ready to present a summary of our results for the various MNR subalgebras. Table I gives the relevant information on the single classical Lie algebra we consider in detail in Appendix B, $D(4) \supset A(2)$. Tables II–VI present all MNR subalgebras of all exceptional Lie algebras as follows:

$G(2) \supset A(1)$, Table II;

$F(4) \supset A(1), G(2) \oplus A(1)$, Tables III A and III B, respectively;

$E(6) \supset A(2), G(2), C(4), F(4), G(2) \oplus A(2)$

Tables IV A–IV E in that order;

$E(7) \supset A(1), A'(1), A(2), A(1) \oplus A(1)$,

$G(2) \oplus A(1), C(3) \oplus G(2), F(4) \oplus A(1)$ Tables V A–V G;

$E(8) \supset A(1), A'(1), A''(1), C(2)$,

$A(2) \oplus A(1), F(4) \oplus G(2)$ Tables VIA–VIF.

In addition to using the general presentation of Sec. 4, as illustrated in the examples considered in Appendices A and B (material from which is included in Tables I, IVB, VIC, and VIE), we have relied heavily on the work of Dynkin[2], McKay and Patera,[13] and FFM to prepare the tables. We have used our Dynkin-like diagrams for roots $\alpha(r,l)$ and McKay and Patera's[13] notation for irreducible representations, as well as the structure constant phases chosen in FFM.

Each of the tables is divided into four sections which we label (i), (ii), (iii), (iv). In Section (i), we list the algebra G, the subalgebra g, and the representations of g into which the quark (q) and adjoint ($\mathscr{A}$) of G decompose. (The notation of McKay and Patera[13] is used to label the irreducible representations and their dimensions are listed below the symbols.) The superscript which appears on the subalgebra is the scale factor s^2 defined by Eq. (4.13). Except for $A(1)$ and $C(n)$, s^2 is the same as Dynkin's index j, defined by (3.1). Because of our normalization of roots in terms of quark weights, s^2 is $2j$ for the $A(1)$ and $C(n)$ algebras.[33] The value of s^2 is determined from Eqs. (4.13) or (4.17). This requires a knowledge of all the weights of the quark of G and weights $\mathbf{w}^{(q)}_x$ into which the quark decomposes in g. The dimensions and weights of the quark and adjoint of G are given[34] in Tables I and II of FFM.

We see from Table II of FFM that each nonzero weight of the quarks has a length squared, $\frac{2}{3}\lambda^2, \lambda^2, \frac{4}{3}\lambda^2, \frac{3}{2}\lambda^2, 2\lambda^2$ for the exceptional groups $G(2), F(4), E(6), E(7)$, and $E(8)$, respectively. The tables of section (ii) headed W_{xa} allow one to calculate the length squared of all the weights $\mathbf{w}_x^{(q)}$.

Section (ii) of the tables gives the weight matrices W_{xa}, the projection matrices M_{ia} (or $\bar{M}_{ia}$), and the roots of g, $\beta(r)$, in terms of the basis of G, $\lambda_i \equiv (i)$. The table headed W_{xa} gives the coefficient in the expansions in the g basis of all the weights $\mathbf{w}_x^{(q)}$ into which the quark of G decomposes [see (4.23)]. The column lables μ_a, μ_b, μ_c, etc. imply that all possible permutations of the a,b,c labels are to be taken. The last column gives the number of weights associated with the configuration of W_{xa} elements, indicated in a given row.

In the tables headed M_{ia} ($\bar{M}_{ia}$), the rows are labeled by the indices (i) which stand for the λ_i, the basis vectors of G and the columns by $\mu_1, \mu_2, \ldots \nu_1, \nu_2, \ldots$, where the μ_a and ν_a are the basis vectors of the respective component algebras of the subalgebra g. The projection is always from $\lambda_i \equiv (i)$ space into (μ_a, ν_a) space and is so indicated by an arrow in the table. The matrices have n_G rows [$n_G + 1 = 3$ for $G = G(2)$] and n_g columns [$n_g + 1$ for $A(2)$ and $G(2)$]. When the subalgebra is not $A(2)$ or $G(2)$, the sum of the squares of all entries in a column must equal s^2 and any two different columns must be orthogonal [see (4.20)]. When g is $A(2)$ or $G(2)$, the sum of the squares of all entries in a column must equal $\frac{2}{3}s^2$ and the scalar product of any two different columns must be $-\frac{1}{3}s^2$ [see (4.31)]. To indicate that in these cases the matrices are pseudo-orthogonal, we put an X through the projection arrow.

The last table in section (ii) gives the simple roots of g (more often, all the roots) as expansions in terms of the λ_i, the basis of G. For nonsimple algebras g (in our examples, there are always only two component subalgebras), two such tables are given. The number on the upper left corner of this table indicates the common denominator by which the linear combination of λ_i (indicated by the numerical entries in a given column) are to be divided. Unless otherwise indicated (in some of the cases where all roots of g are given), the $\lambda_i \equiv (i)$ labels of the rows carry over from the M_{ia} tables immediately to the left of these $\beta(r)$ tables. The convention for labeling the $\beta(r)$ is the same as that for the $\alpha(r,l)$, and is discussed below when section (iii) is considered.

In section (iii) of the tables, the roots $\alpha(r,l)$ which project into the simple roots $\beta(a)$ are listed and diagramed. The columns (rows) have two headings, the leading heading being the root[35] $\beta(a)$ of g. Just below (to the right) are the $\alpha(a,l)$ which project into that $\beta(a)$. Associated with the $\alpha(a,l)$ are diagrams which are generalizations of those introduced by Dynkin, in that any two roots in a row (or column) directly connected by a continuous line have a scalar product of $-\lambda^2$, two roots directly connected by a dashed line have a scalar product of $+\lambda^2$, and any two roots not connected, or not directly connected, are orthogonal. The information provided by these column and row headings plus the G phase of $N_{\alpha\beta}$ is sufficient to allow us to write down Eqs. (4.47) and (4.50). The entries in the table give the scalar products between component roots. The information provided by these scalar products allow us to write down Eq. (4.53) [see the discussion following (4.53)], and, in the few cases in which all roots are specified, Eq. (4.60). The roots $\alpha(r,l)$ [$\beta(r)$] are labeled in an obvious way in terms of their λ_i [μ_a or ν_a] decomposition. We have reserved α (β) to label the roots of G (g) and lower (upper) indices indicating a positive (negative) basis vector λ or μ. Thus, for example,

$$\beta_a^b \equiv \mu_a - \mu_b, \tag{5.1}$$

and for corresponding "vector" roots of G

$$\alpha_i^j = \lambda_i - \lambda_j. \tag{5.2}$$

For the spinor roots of $E(6)$ and $E(8)$, we label the α's by the fewest number of positive or negative λ_i which appear. For example, in the case of $E(6)$ we have[27]

$$\alpha_1 \equiv \tfrac{1}{2}(\lambda_1 - \lambda_2 - \lambda_3 - \lambda_4 - \lambda_5 - \sqrt{3}\bar{\lambda}_6), \tag{5.3}$$

[recall that the spinor roots for $E(6)$ or $E(8)$ are such that the total number of negative signs which appear must be odd]. For the case of $E(7)$, we also include the position of the label 7. Thus

$$\alpha_7^{14} = \tfrac{1}{2}(-\lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 + \lambda_5 + \lambda_6 + \sqrt{2}\bar{\lambda}_7). \tag{5.4}$$

We must include the 7, because, although the total number of minus signs in front of $\lambda_1 \cdots \lambda_6$ must be even, we have roots for both signs of $\bar{\lambda}_7$. In addition, we denote all indices 1–6 positive (negative) by a $+$ ($-$) superscript. Thus,

$$\alpha_7^{(\pm)} = \frac{1}{2}\left((\pm)\sum_{i=1}^{6}\lambda_i + \sqrt{2}\bar{\lambda}_7\right). \tag{5.5}$$

The algebra of $F(4)$ has both odd and even numbers of minus signs in the spinor weights. To avoid confusion with vector roots, we label all basis vectors in this case. Thus for $G \equiv F(4)$, by the symbol α_{12}^{34} we mean

$$\alpha_{12}^{34} = \tfrac{1}{2}(\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4), \tag{5.6}$$

and similarly for β_{12}^{34} in terms of μ_a, for $g = F(4)$.

The remaining special cases involve "vector" roots like 2ν, μ_1, $2\mu_3$, λ_7, or χ_a [where χ_a are the $A(2)$ quark weights, Eq. (4.26), and $G(2)$ roots]. These roots are directly so indicated in the tables in which they appear.

In section (iv) of the tables, the results for the simple generators $e_{\beta(a)}$ of g in terms of the generators $E_{\alpha(a,l)}$ of G are given. (Frequently the results for all the generators of g are given. This fact is signaled by the phrase "All e_β" in the tables.) The notation used for generators is obviously related to that for roots. Thus, the generator denoted E_p^q is that associated with the root α_p^q, and the generator E^{pqr} is that associated with the root α^{pqr}, etc. We use lower case letters for the generators of g, and again the notation is almost obvious. Thus, e_a^b is associated with the g root $\beta_a^b \equiv \mu_a - \mu_b$.

For the roots which are explicitly written, like 2ν, χ_a, etc., we also explicitly label the corresponding generators: $e_{2\nu}, e_{\chi_a}$, etc.

We illustrate the use of the tables and figures by discussing the cases of $E(6) \supset G(2)$ and $E(8) \supset A^{1040}(1)$ in Appendix A. We treat these two cases in detail, because they are the ones in which our results differ significantly from those of Dynkin. As described in Appendix A, these differences are due to errors in Dynkin.

In general, other differences between our results and those of Dynkin are due to different choices for the G phases

and are also due to Dynkin's definition of generators other than those associated with simple roots in terms of nested commutators of generators for simple roots:

$$e_{i_1 i_2 \cdots i_k} = [\,[e_{i_1}, e_{i_2}], \ldots, e_{i_k}\,]. \tag{5.7}$$

From his convention [Ref. 2, Eq. (9.10), p. 171] any generators appearing with an even number of k labels (i.e., containing an even number of simple roots) will have a phase of $\pm i$ relative to those with an odd number of such labels.

Other apparent differences which arise for the cases $E(7) \supset A^{462}(1)$, $E(8) \supset C(2)$, $E(8) \supset A(2) \oplus A(1)$, and $E(8) \supset A^{1520}(1)$ are due to the freedom, in these cases, of arbitrarily choosing some of the coefficients $|B(a,l)|^2$. This point is also discussed in Appendix A.

Examples in which we obtain expressions for all the generators $e_{\beta(r)}$ (not just the simple ones) are given in Appendix B.

6. DISCUSSION

The generators for simple roots of MNR subalgebras g of the exceptional algebras G in terms of generators of G have been given by Dynkin.[2] In addition, the $\beta(a)$ of simple roots of g is given in terms of $\alpha(i)$ of G in Ref. 2. Projection matrices which determine the $\alpha(p) \to \beta(r)$ are given in Refs. 13, 16, and 20. In this work we have shown how these results (and some of their generalizations) may be obtained in a more geometrical framework. By making use of the orthogonal bases (the λ_i), we have avoided a number of steps required in the usual procedure. There, one needs to know both the covariant and contravariant bases, and one must specify a certain ordering of weights (and adhere to it). The projection matrices so constructed are equally complicated for the classical as well as the exceptional algebras. In our case, since the bases we use are orthogonal, the projection matrices are parts of orthogonal matrices (or of pseudo-orthogonal, as defined in Sec. 4). For the classical algebras, because of the direct relationship between the quark weights and the basis vectors, the projection matrices are immediately obvious. The situation for the exceptional algebras is only slightly more difficult. For these algebras, we do not have the direct relationship between quark weights and basis vectors that we have for the classical algebras. However, if we use the orthogonal basis associated with the regular maximal $D(n)$ or $B(n)$ subalgebras of the exceptional algebras, we can construct the projection matrix, almost by inspection. The disadvantage of our form of the projection matrix (which we hope to rectify in the future) concerns its use in constructing branching rules for any arbitrary representation of G.

Using the newly constructed projection matrices as a starting point, we have shown how, from Dynkin-type diagrams[36] and a few tables of scalar products of roots, one may derive the equations which determine the $B(a,l)$, the coefficients which relate the subalgebra generators to the generators of the algebra. In this way we have been able to find solutions more symmetric than those given by Dynkin. There are as many nonvanishing coefficients $B(a,l)$ as there are roots $\alpha(a,l)$, a feature not always present in Ref. 2. Also, our set of generators $E_{\alpha(a,l)}$, which enter an $e_{\beta(a)}$ have simple specifications in terms of root labels, given in terms of λ_i. In contrast, Dynkin frequently has to define them indirectly in terms of nested commutators [see Eq. (5.7)]. For the case[37] of $E(8)$, a maximum of eight such nested commutators appear. Two nested commutators occur frequently.

Another advantage of our approach, in addition to the simplified derivation and presentation of results, using orthogonal bases, is its greater convenience for physical applications. It is in general more convenient for computational purposes to specify all roots $\beta(r)$ and their corresponding generators rather than just simple roots and their generators. We have succeeded in doing this for most of the cases we considered. With further effort, other cases could no doubt be similarly presented. In contrast, only simple roots are specified in Ref. 2. The phases of the $B(a,l)$ chosen by Dynkin are frequently zero, or at worst, π. At first glance, this is an advantage of Ref. 2 over the present work. However, in recent physical applications, such as symmetry breaking in grand unified models[3,38,39] one has to deal simultaneously with all of both maximal regular (MR) and MNR subgroups g of G, or a large subset of such g's. In such cases, it is essential to have a consistent set of phase choices for structure constants. This choice, in turn, serves to limit the arbitrariness of the phase choices of the $B(a,l)$, as is indicated by the relations derived in Sec. 4 above. If this condition is not met, the linear and bilinear relations (FFM) between structure constants will not be satisfied in general. Finally, the requirement that all step-up operators be the Hermitian adjoints of step-down operators is a physically useful requirement. We impose this condition in the present paper. Apart from the errors in Ref. 2, this requirement is also satisfied there.[40]

We have not considered nonmaximal regular or nonregular subalgebras of the exceptional algebras in the present work. The extension to these cases is trivial. The $\beta(a)$ and $e_{\beta(a)}$ can be obtained directly for these cases, using the methods we have developed. They can also be obtained indirectly by considering chains of subalgebras, each link of which is a maximal subalgebra of the previous algebra in the chain.

ACKNOWLEDGMENTS

One of the authors (G. F.) would like to thank another (P. T. M.) for his hospitality at the University of Bath where some of the work reported in this paper was carried out.

APPENDIX A

Tables I(iii)–VI(iii) have been constructed so that one may read off directly the equations corresponding to (4.47), (4.50), (4.53), and, when required, (4.60). As an example of how the procedure works, consider the $G(2)$ subalgebra of $E(6)$ (Table IVB). In Table IVB(iii), the column and row headings give the simple roots of $G(2)$, β_1^2, and χ_2, respectively. The secondary column and row labels specify which $\alpha(a,l)$ of $E(6)$ project into the respective simple roots. The Dynkin-like diagram associated with the $\alpha(a,l)$'s indicates scalar products. For the set associated with β_1^2 all scalar products are zero, whereas for those associated with χ_2 not all scalar products are zero. With this information, we can write down Eqs. (4.47) which determine the $|B(a,l)|^2$. They are

$$2|B_1^2|^2 = 2, \quad 2|B_{45}|^2 = 2, \quad 2|B_4^5|^2 = 2 \tag{A1}$$

and

$$2|B^4_1|^2 = 2, \quad 2|B_{25}|^2 - |B_6|^2 = 2, \quad 2|B_5|^2 - |B^5_2|^2 = 2,$$
$$-|B_{25}|^2 + 2|B_6|^2 = 2, \quad -|B_5|^2 + 2|B^5_2|^2 = 2, \tag{A2}$$

from which it immediately follows that the $|B(a,l)|^2$ are determined uniquely to be

$$|B^2_1| = |B_{45}|^2 = |B^5_4|^2 = |B^4_1|^2 = 1 \tag{A3}$$

and

$$|B_{25}|^2 = |B^5_2|^2 = |B_5|^2 = |B_6|^2 = 2. \tag{A4}$$

There are no equations of type (4.50) since there are no dashed line connections among the roots. To find relations of type (4.53), we must find configurations in Table IVB(iii) like those of Figs. 1(a) or 1(b). There is just one such configuration and thus one such equation, namely,

$$d(-2-1-4)B^2_1 B^{4*}_1 + d(4-52)B_{45}B^*_{25} + d(452)B^5_4 B^{5*}_2 = 0, \tag{A5}$$

where the $d(pqr)$ are phases which have been defined in FFM. This last equation can be considered an equation for two relative phases of the $B(a,l)$. We may take all the phases associated with the root χ_2 and the phase of B^2_1 to be zero. This then determines the phases of B_{45} and B^5_4:

$$B_{45} = e^{i\theta}, \quad B^5_4 = e^{-i\theta}, \tag{A6}$$

where

$$e^{i\theta} = (1 + i\sqrt{7})/2\sqrt{2}. \tag{A7}$$

The final results for the roots e^2_1 and e_{χ_2} are listed in Table IVB(iv). These results differ markedly from those to be found in Dynkin.[41] At first sight, it appears that the difference arises because Dynkin does not seem to make the hermiticity assumption implied by our Eqs. (4.38), (4.40), and (4.64). In fact, in place of (4.40), he would write

$$e_{-\beta(a)} = \frac{1}{s'(a)} \sum_{l=1}^{L} B'(a,l) E_{-\alpha(a,l)}, \tag{A8}$$

where $B'(a,l)$ is generally unrelated to $B(a,l)$. This implies that our equations (4.47) for $|B(a,l)|^2$ will become equations for $B(a,l)B'(a,l)$, which in turn implies that for this case [the $G(2)$ subalgebra of $E(6)$] the results (A3) and (A4) will be modified to read

$$B^2_1 B^{2'}_1 = B_{45}B'_{45} = \cdots = 1 \tag{A9}$$

and

$$B_{25}B'_{25} = B^5_2 B^{5'}_2 = \cdots = 2. \tag{A10}$$

Similarly, in place of (A5), which follows from (4.53), we must write

$$d(-2-1-4)B^2_1 B^{4'}_1 + d(4-52)B_{45}B'_{25} + d(452)B^5_4 B^{5'}_2 = 0. \tag{A11}$$

We must now also write the equation which is a consequence of (4.52) when $\beta(a)$, $\beta(b) \to -\beta(a)$, $-\beta(b)$. The result is

$$d(214)B^{2'}_1 B^4_1 + d(-45-2)B'_{45}B_{25} + d(-4-5-2)B'^5_4 B^5_2 = 0. \tag{A12}$$

Proceeding as before, we can choose as unknown phases those associated with B_{45}, B'_{45}, B^5_4, and $B^{5'}_4$. Using (A9)–(A12), we find that, apart from a possible arbitrary overall phase, B_{45} and B^5_4 are determined as in (A6), with

$$B'_{45} = B^*_{45}, \quad B^{5'}_4 = B^{5*}_4.$$

This result does not depend on the phase convention for the $d(pqr)$ but is a consequence of Jacobi identities for phases,[42] namely,

$$d(pqr)d(-p-q-r) = -1. \tag{A13}$$

Dynkin's result, which in our notation implies

$$B_{45} = -B^{5'}_4 = \sqrt{2}, \quad B^{45\prime} = -B^5_4 = 1/\sqrt{2}, \tag{A14}$$

can be obtained by using an incorrect version of (A13) to determine some phases, namely,

$$d(214)d(-2-1-4) = +1. \tag{A15}$$

There are other cases for which our results differ (apart from some trivial phase choices) from those of Dynkin. They are all those cases in which the equations (4.47) do not lead to unique solutions for the $|B(r,l)|^2$. These occur when one or more pairs of the $\alpha(r,l)$ for a given r have a positive scalar product (i.e., they are connected by a dashed line in the Dynkin-like diagrams). In these cases [$E(7) \supset A^{462}(1)$, $E(8) \supset C(2)$, $E(8) \supset A(2) \oplus A(1)$, $E(8) \supset A^{1520}(1)$, $E(8) \supset A^{1040}(1)$] we have looked for and found solutions which we term symmetric. They are those in which the $|B(r,l)|^2$ corresponding to a pair of $\alpha(r,l)$ with positive scalar products are taken to be equal. This inevitably leads to some $B(r,l)$ coefficients which are complex. It appears that Dynkin looks for solutions in which (in our notation for generators) as many $B(r,l)$ as possible are real and some of them are zero. We illustrate the different possible solutions by examining in detail the case of $E(8) \supset A^{1040}(1)$ (Table VIC). This will help to describe our methods further and also to point out another (nontypographical) error we have found in Dynkin's results. That Dynkin's generator[43] e_x for $A^{1040}(1)$ is incorrect is easily seen since his $|B(a,l)|^2$ coefficients do not satisfy the sum rule (4.48). From our Table VIC(iii), we can see that there are 12 $\alpha(a,l)$ which project into the $A(1)$ root 2ν. Again, from the rules outlined in Sec. 4, one may immediately write down the matrix of coefficients and equations [corresponding to (4.47)] which determine the $|B(a,l)|^2$. They are

$$\begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & +1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & +1 & 2 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 2 & +1 & 0 & +1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & +1 & 2 & +1 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & +1 & 2 & +1 & 0 & -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & +1 & 0 & +1 & 2 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 2 & +1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & +1 & 2 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 2 & +1 & -1 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & -1 & +1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} |B_2^3|^2 \\ |B_3^4|^2 \\ |B_3^5|^2 \\ |B_4^6|^2 \\ |B_5^6|^2 \\ |B_5^7|^2 \\ |B_4^7|^2 \\ |B_6^8|^2 \\ |B_7^8|^2 \\ |B_{78}|^2 \\ |B_{68}|^2 \\ |B_1|^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \tag{A16}$$

These equations do not determine the $|B(a,l)|^2$ uniquely. This follows from the presence of $+1$ factors in the matrix (see Sec. 4). However, the sums of those $|B(a,l)|^2$, the roots of which have positive scalar products, are determined. The results are

$$\begin{aligned} &|B_2^3|^2 = 38, \quad |B_3^4|^2 + |B_3^5|^2 = 74, \\ &|B_4^6|^2 + |B_5^6|^2 + |B_5^7|^2 + |B_4^7|^2 = 142, \\ &|B_6^8|^2 + |B_7^8|^2 = 88, \quad |B_{78}|^2 + |B_{68}|^2 = 118, \quad |B_1|^2 = 60. \end{aligned} \tag{A17}$$

Again, because of the existence of these positive scalar products, there will be further restrictions of the type (4.50). These can be obtained by picking out the subdiagrams of the type illustrated in Figs. 2(a) or 2(b) from the structure in Table IV C(iii). Using the $E(8)$ phases[44] of FFM, we find

$$\begin{aligned} &-B_3^4 B_3^{5*} + B_5^6 B_4^{6*} + B_5^7 B_4^{7*} = 0, \\ &B_{78} B_{68}^* + B_7^8 B_6^{8*} - B_5^6 B_5^{7*} - B_4^6 B_4^{7*} = 0. \end{aligned} \tag{A18}$$

The Dynkin solution contains two $B(a,l)$ coefficients which are zero. Such a solution (with some coefficients in common with Dynkin's incorrect solution) can be found by taking

$$B_{68} = B_4^7 = 0. \tag{A19}$$

This choice yields

$$\begin{aligned} e_{2v} = (1/\sqrt{520})\{&\sqrt{38}E_2^3 + \sqrt{4.37}E_3^4 \\ &+ \sqrt{69.63}E_3^5 + \sqrt{38.37}E_4^6 + \sqrt{7.93}E_5^6 \\ &+ \sqrt{95.69}E_5^7 + \sqrt{78.31}E_6^8 \\ &+ \sqrt{9.69}E_7^8 + \sqrt{118}E_{78} + \sqrt{60}E_1\}. \end{aligned} \tag{A20}$$

(The coefficients under the square roots are all ratios of very large integers. They have been evaluated only to two decimal places above.) A somewhat simpler, Dynkin-type solution [with three $B(a,l)$ coefficients equal to zero] can be found by taking

$$B_4^7 = B_5^6 = 0. \tag{A21}$$

From (A18), this choice implies that

$$\text{either} \quad B_3^4 \quad \text{or} \quad B_3^5 = 0. \tag{A22}$$

If we choose B_3^4 to be zero, we find

$$\begin{aligned} e_{2v} = (1/\sqrt{520})\{&\sqrt{38}E_2^3 + \sqrt{74}E_3^5 \\ &+ \sqrt{34}E_4^6 + \sqrt{108}E_5^7 \\ &+ \sqrt{1716/37}E_6^8 + \sqrt{1540/37}E_7^8 \\ &+ \sqrt{3640/37}E_{78} - \sqrt{726/37}E_{68} + \sqrt{60}E_1\}. \end{aligned} \tag{A23}$$

Our symmetric solution, which is listed in Table VI C(iv), is obtained by equating all $|B(a,l)|^2$, the associated roots of which have positive scalar products.

In all of the other cases in which our results differ from those of Dynkin, the Dynkin results can be obtained by setting to zero certain suitably chosen $B(a,l)$ coefficients. Other apparent differences have been discussed in Sec. 5.

APPENDIX B

In this appendix, we show how, by a judicious choice of notation, it is often possible to express all of the generators of g (not just the simple ones) in a uniform manner. We illustrate with two examples.

$D(4) \supset A(2)$

In this case, the quark of $D(4)$ (with weights $\pm\lambda_i$, $i = 1,\ldots,4$) decomposes into the adjoint of $A(2)$ [with weights $\pm(\mu_a - \mu_b)$, $a,b = 1,2,3$ and two zero weights]. Since the subgroup is $A(2)$, the projection matrix will be pseudo-orthogonal, as defined in Sec. 4, and the projection defined by (4.18) is immediately obvious. It yields the matrix of Table I(ii). The value of the index, $s^2(=3)$, follows from (4.13) and it is easy to see that the matrix $\bar{M}_{ia}$ satisfies the conditions (4.31). From this matrix, and using (4.30), one immediately finds for the roots β_b^c,

$$\beta_b^c \equiv \mu_b - \mu_c = \lambda_a - \frac{1}{3}\sum_{b=1}^{3}\lambda_b \quad (a,b,c = 1,2,3 \text{ cyclic}). \tag{B1}$$

It also follows directly from this table that the roots of $D(4)$ which project into $\beta(a)$ [β_b^c] are

$$\alpha_a^{\pm} \equiv \lambda_a \pm \lambda_4, \quad \alpha_a^0 \equiv -\lambda_b - \lambda_c, \quad b,c \neq a. \tag{B2}$$

We have here introduced a notation for our $\alpha(a,l)$ different

from our normal convention. Its usefulness is demonstrated below. The three roots α_a^l ($l = \pm,0$) for a given a are orthogonal and the matrix of scalar products for two different roots is given in Table I(iii). Because of the orthogonality of the α_a^l for different l, Eq. (4.47) and its solution follow immediately:

$$|B_a^l|^2 = 1, \quad \text{all } a,l. \tag{B3}$$

Since there are no two roots connected by a dashed line in the labeling of the rows or columns, there are no restrictions of the type (4.50). There are, of course restrictions of the type (4.53), which follow from (4.52). However, since we wish to generate the entire subalgebra, we proceed directly to restrictions of type (4.60). By looking for scalar product configurations in Table I(ii), given by Figs. 3(a) and 3(b) and using the $D(4)$ phases of FFM, we arrive at the equations

$$B_a^+ B_b^- + B_a^- B_b^+ = (-1)^{a+b} B_c^{0*}, \tag{B4}$$

$$\begin{aligned} (-1)^{b+c} B_a^0 B_b^+ + (-1)^{a+c} B_a^+ B_b^0 &= B_c^{-*}, \\ (-1)^{b+c} B_a^0 B_b^- + (-1)^{a+c} B_a^- B_b^0 &= B_c^{+*}. \end{aligned} \tag{B5}$$

Multiplying by the complex conjugates of the right sides and using (B3), these three equations for all a,b,c become

$$\begin{aligned} B_a^+ B_b^- B_c^0 + B_a^- B_b^+ B_c^0 &= (-1)^{a+b}, \\ (-1)^{b+c} B_a^0 B_b^+ B_c^- + (-1)^{a+c} B_a^+ B_b^0 B_c^- &= 1. \end{aligned} \tag{B6}$$

Recognizing the identities

$$\begin{aligned} \sum_{m=1}^{3} e^{2im\pi/3} &= 0, \\ \sum_{m=1}^{3} (-1)^m e^{im\pi/3} &= 0, \end{aligned} \tag{B7}$$

we can find a complete solution, to (B3) and (B6):

$$B_a^l = (-1)^{a+1} e^{lia\pi/3} \quad (a = 1,2,3) \ (l = \pm,0). \tag{B8}$$

The equation which follows from (4.52), namely,

$$(-1)^{a+b}(B_a^+ B_b^{+*} + B_a^- B_b^{-*}) + B_a^0 B_b^{0*} = 0, \tag{B9}$$

is, of course, satisfied identically by (B8).

$E(8) \supset A(2) \oplus A(1)$

As indicated in Tables VIE(ii), it is convenient to relabel the orthogonal basis λ_i ($i = 1,\ldots,8$) so that $i = a,\dot{a},7,8$, where $a = 1,2,3$, $\dot{a} = \dot{1},\dot{2},\dot{3}$, and $\dot{a} = a + 3$. The scaling factors follow immediately from (4.17), on using the Table VI E(ii) labeled W_{xa}. Thus

$$s_{A(2)}^2 = 6, \quad s_{A(1)}^2 = 32. \tag{B10}$$

Again, since the $A(2)$ part of the projection matrix is pseudo-orthogonal, the sum of the squares of any column labeled by a μ_a must be $\frac{2}{3}s_{A(2)}^2 = 4$ [see (4.31)]. An examination of the weights of W_{xa} indicates immediately that the projection matrix will have the form given. From this table the roots β_b^c and 2ν in terms of the λ_i can be read off and are given in Table VIE(ii). An examination of the adjoint [which is identical with the quark for $E(8)$] decomposition in Table VIE(i) tells us that six $\alpha(a,l)$ project into each β_b^c and eight $\alpha(a,l)$ project into 2ν. These $\alpha(a,l)$ are found from the projection matrix and the results are entered as the labels of the rows and columns of Table VIE(iii).

We next introduce the expansion of the $A(2) \oplus A(1)$ generators defined by (4.38) and (4.40). We find

$$(s'_{A(2)})^2 = s_{A(2)}^2 = 6, \quad (s'_{A(1)})^2 = s_{A(1)}^2/2 = 16. \tag{B11}$$

From the scalar products indicated by the diagram in the row labels of Table VIE(iii), we deduce immediately, using (4.47),

$$|B_c^b|^2 = |B_b^c|^2 = 1, \quad (a,b,c) = (1,2,3,\text{cyclic}), \tag{B12}$$

$$2|B_a^7|^2 + |B_a^8|^2 - |B_7^{\dot{a}}|^2 = 2, \tag{B13}$$

$$2|B_7^{\dot{a}}|^2 + |B_8^{\dot{a}}|^2 - |B_a^7|^2 = 2. \tag{B14}$$

Two more sets of equations are obtained from (B13) and (B14) with the labels 7 and 8 interchanged. Because some scalar products in the set $\alpha(r,l)$ are positive, equations of type (4.50) occur. Looking for configurations as in Fig. 2(a) or 2(b), and using the $E(8)$ phases of FFM, we have

$$B_a^7 B_a^{8*} - B_8^{\dot{a}} B_7^{\dot{a}*} = 0. \tag{B15}$$

Were we interested only in the algebra of the simple roots, we would need the conditions (4.53) which follow from the commutation relations (4.52). Using the entries in Table VIE(iii), we have

$$B_a^7 B_b^{7*} + B_a^8 B_b^{8*} - B_{\dot{c}}^b B_{\dot{c}}^{a*} = 0$$

and (B16)

$$B_8^{\dot{a}} B_8^{\dot{b}*} + B_7^{\dot{a}} B_7^{\dot{b}*} - B_{\dot{b}}^c B_{\dot{a}}^{c*} = 0, \quad (a,b,c) = (1,2,3 \text{ cyclic}).$$

We are, however, interested in constructing the entire $A(2)$ subalgebra. We therefore require conditions (4.60), which follow from the commutations relations (4.58). As outlined previously, we look for configurations of the type in Fig. 3(a) or 3(b) in the table of scalar products, Table VIE(iii). Using the $E(8)$ phases of FFM, we find

$$B_a^7 B_7^{\dot{b}} + B_a^8 B_8^{\dot{b}} = B_{\dot{b}}^{a*}, \tag{B17}$$

$$B_7^{\dot{a}} B_b^7 + B_8^{\dot{a}} B_b^8 = -B_{\dot{a}}^{b*}, \tag{B18}$$

$$B_{\dot{c}}^b B_b^7 - B_{\dot{c}}^a B_a^7 = B_7^{\dot{c}*}, \tag{B19}$$

$$-B_{\dot{b}}^c B_7^{\dot{b}} + B_{\dot{a}}^c B_7^{\dot{a}} = B_c^{7*}, \quad (a,b,c) = (1,2,3,\text{cyclic}). \tag{B20}$$

Two more sets of equations are obtained from (B19) and (B20) by replacing the label 7 by 8.

From the scalar products implied by the diagram of the column labels for the eight roots which project into 2ν [Table VIE(iii)], and, using (4.47), we find immediately that

$$2|B^{abc}|^2 - |B^{\dot a\dot b\dot c}|^2 = 2,$$
$$-|B^{abc}|^2 + 2|B^{\dot a\dot b\dot c}|^2 = 2, \tag{B21}$$

and

$$2|B^{a\dot a 7}|^2 + |B^{a\dot a 8}|^2 - |B^{b\dot b 8}|^2 - |B^{c\dot c 8}|^2 = 2, \tag{B22}$$

and another equation obtained from (B22) by interchanging the labels 7 and 8. Again, since some of the scalar products of the roots which project into 2ν are positive, we have relations of the type (4.50). They are

$$\sum_{a=1}^{3} B^{a\dot a 7} B^{a\dot a 8*} = 0. \tag{B23}$$

Further, there are conditions of the type (4.53) which are consequences of the commutation relation[45]

$$[e_{2\nu}, e_{-\beta(a)}] = 0. \tag{B24}$$

They are

$$B^{a\dot a 7} B^{\dot a *}_{8} - B^{a\dot a 8} B^{\dot a *}_{7} = 0, \tag{B25}$$

$$B^{b\dot b 7} B^{7*}_{a} + B^{b\dot b 8} B^{8*}_{a} + B^{abc} B^{c*}_{b} = 0, \tag{B26}$$

$$B^{c\dot c 7} B^{7*}_{a} + B^{c\dot c 8} B^{8*}_{a} - B^{abc} B^{b*}_{\dot c} = 0, \tag{B27}$$

$$B^{b\dot b 7} B^{b*}_{\dot c} - B^{c\dot c 7} B^{c*}_{b} + B^{\dot a\dot b\dot c} B^{\dot a *}_{7} = 0,$$
$$(a,b,c) = (1,2,3,\text{cyclic}), \tag{B28}$$

and a further equation, obtained from (B28), by replacing 7 by 8.

Except for (B12) and (B21), which lead to unique solutions for $|B^{c}_{b}|^2, |B^{b}_{\dot c}|^2, |B^{abc}|^2, |B^{\dot a\dot b\dot c}|^2$, the other equations, (B13), (B14), etc., (B22), etc., do not lead to unique solutions for the magnitudes appearing in them. This, as we know, follows from the fact that some of the scalar products of the associated roots are positive. However, we can find symmetric solutions in which coefficients associated with roots with positive scalar products are taken equal in magnitude. This ansatz leads to the following solution for equations (B12), (B13), (B14), etc., and (B21), (B22), etc.:

$$|B^{7}_{a}|^2 = |B^{8}_{a}|^2 = |B^{\dot a}_{7}|^2 = |B^{\dot a}_{8}|^2 = |B^{b}_{\dot c}|^2 = |B^{c}_{b}|^2 = 1 \tag{B29}$$

and

$$|B^{abc}|^2 = |B^{\dot a\dot b\dot c}|^2 = |B^{a\dot a 7}|^2 = |B^{a\dot a 8}|^2 = 2. \tag{B30}$$

Recalling the identities (B7), we see that a solution to all of the phase determining equations can be found. Such a solution is

$$B^{7}_{a} = B^{\dot a}_{8} = B^{8*}_{a} = B^{\dot a *}_{7} = e^{-ia\pi/3}, \tag{B31}$$

$$B^{b}_{\dot a} = -B^{a}_{\dot b} = (-1)^{c}, \quad (a,b,c) = (1,2,3,\text{cyclic}), \tag{B32}$$

$$B^{a\dot a 7} = B^{a\dot a 8*} = (-1)^{a} e^{-ia\pi/3}\sqrt{2}, \tag{B33}$$

and

$$B^{abc} = -B^{\dot a\dot b\dot c} = \sqrt{2}. \tag{B34}$$

A very different solution has been given by Dynkin.[37] He is concerned only with the simple roots, and consequently the generators e^{3}_{2}, e^{1}_{3}, and $e_{2\nu}$. For this case, one must find solutions to Eqs. (B12)–(B14), etc., (B15), (B16), (B21), (B22), etc., (B23) and (B25)–(B28), etc. [and is not concerned with (B17)–(B20), etc.], where (a,b,c) take on the values (1,2,3) and (2,3,1). As we already know, Eqs. (B13), (B14), etc., and (B22), etc., do not lead to unique solutions for the magnitudes appearing in them. However, we do have the result that

$$|B^{7}_{a}|^2 = |B^{\dot a}_{7}|^2 = 2 - x_a, \quad |B^{8}_{a}|^2 = |B^{\dot a}_{8}|^2 = x_a, \quad \text{all } a, \tag{B35}$$

with $0 \leqslant x_a \leqslant 2$. We now look for real solutions and substitute these results into (B16) for $(a,b,c) = (1,2,3)$. We find

$$\sqrt{2-x_1}\sqrt{2-x_2} + \sqrt{x_1}\sqrt{x_2} = 1. \tag{B36}$$

We may choose

$$x_1 = 0, \tag{B37}$$

which implies that

$$x_2 = \tfrac{3}{2}. \tag{B38}$$

Use of (B22), etc., (B23), and (B25) allows a determination of the additional magnitudes:

$$|B^{1\dot 1 7}|^2 = 4, \quad |B^{1\dot 1 8}|^2 = 0, \quad |B^{2\dot 2 7}|^2 = |B^{3\dot 3 7}|^2 = 1,$$
$$|B^{2\dot 2 8}|^2 = |B^{3\dot 3 8}|^2 = 3. \tag{B39}$$

If we arbitrarily choose all the signs appearing in the expansion for $e_{\beta(2)}$ to be positive, the remaining equations allow a determination of the other signs. We find a solution which is essentially that given by Dynkin,[37] differences in signs being due to different choices for the $E(8)$ phases used by Dynkin and FFM:

$$e^{3}_{2} = (1/\sqrt{6})\{\sqrt{2}(E^{7}_{\dot 1} - E^{\dot 1}_{7}) + (E^{2}_{\dot 3} - E^{\dot 3}_{2})\},$$
$$e^{1}_{3} = (1/\sqrt{6})\{(1/\sqrt{2})(E^{7}_{\dot 2} + E^{\dot 2}_{7}) + \sqrt{\tfrac{3}{2}}(E^{8}_{\dot 2} + E^{\dot 2}_{8}) + (E^{1}_{\dot 3} + E^{\dot 3}_{1})\}, \tag{B40}$$
$$e_{2\nu} = \tfrac{1}{4}\{\sqrt{2}(E^{123} + E^{\dot 1\dot 2\dot 3}) + 2E^{1\dot 1 7} + (E^{2\dot 2 7} + E^{3\dot 3 7}) + \sqrt{3}(E^{2\dot 2 8} - E^{3\dot 3 8})\}.$$

TABLE I. D(4).

$D(4) \supset A^3(2)$: $q \to (1,1)$; $\mathcal{A} \to (3,0) + (0,3) + (1,1)$ (8, 10, $\overline{10}$, 8) (i)

W_{xa}

μ_a	μ_b	μ_c	No.
$(-)^1$	$(+)^1$	0	6
0	0	0	1

$a \neq b \neq c$, $a, b, c = 1, 2, 3$ (ii)

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	3	β_b^c
(1)	0	1	-1	(a)	2
(2)	-1	0	1	(b)	-1
(3)	1	-1	0	(c)	-1
(4)	0	0	0	(4)	0

	β_b^c	λ_a^4	β_{a4}	α^{bc}
β_c^a	α_b^4	+	-	-
	α_{b4}	-	+	-
	α^{ca}	-	-	+

(a, b, c) = (1, 2, 3, cyclic) (iii)

All e_β:

$$e_b^c = \frac{(-1)^{a+1}}{\sqrt{3}}\{e^{ia\theta}E_{a4} + e^{-ia\theta}E_a^4 + E^{bc}\}; \quad \theta = \frac{\pi}{3} \qquad \text{(iv)}$$

TABLE II.[46] $G(2)$.

$G(2) \supset A^{56}(1)$: $q \to (6)$; $\mathcal{A} \to (10) \oplus (2)$ (7, 11, 3) (i)

W_{xa}

ν	No.
±6	2
±4	2
±2	2
0	1

$\bar{M}_{ia}$

	ν
(1)	-6
(2)	4
(3)	2

28	$2\underline{\nu}$
	-6
	4
	2

(ii)

$\alpha_2^3 \quad \lambda_3 = \frac{1}{3}\{-(1) - (2) + 2(3)\} \to 2\underline{\nu}$ (iii)

Principal subalgebra

All e_β:

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{14}}(\sqrt{5}\,E_2^3 + 3E_{\lambda_3}) \qquad \text{(iv)}$$

TABLE IIIA.[47]

$F(4) \supset A^{312}(1)$: $q \to (16) \oplus (8)$; $\mathcal{A} \to (22) \oplus (14) \oplus (10) \oplus (2)$ (17, 9, 23, 15, 11, 3) (i)

W_{xa}

$\underline{\nu}$	No.
$\pm n_1$	2x4
$\pm n_2$	4x4
0	2

$n_1 = 16, 14, 12, 10$

$n_2 = 8, 6, 4, 2$ (ii)

$\bar{M}_{ia}$

	$\underline{\nu}$
(1)	16
(2)	6
(3)	4
(4)	2

156	$2\underline{\nu}$
	16
	6
	4
	2

$\alpha_2^3 \quad \lambda_3^4 \quad \lambda_4 \quad \alpha_1^{234} \to 2\underline{\nu}$ (iii)

Principal subalgebra

All e_β:

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{78}}(\sqrt{11}\,E_2^3 + \sqrt{21}\,E_3^4 + \sqrt{30}\,E_{\lambda_4} + 4\,E_1^{234}) \qquad \text{(iv)}$$

TABLE IIIB.[48]

$F(4) \supset G^1(2) \oplus A^{16}(1)$: $q \to (0,1)(2) \oplus (0,0)(4)$; $\mathcal{A} \to (0,1)(4) \oplus (1,0)(0) \oplus (0,0)(2)$ (21, 5, 35, 14, 3) (i)

W_{xa}

μ_a	μ_b	μ_c	$\underline{\nu}$	No.
$(+)2/3$	$(\bar{+})1/3$	$(\bar{+})1/3$	± 2	12
$(\bar{+})2/3$	$(+)1/3$	$(\bar{+})1/3$	0	6
0	0	0	± 4	2
0	0	0	± 2	4
0	0	0	0	2

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	$\underline{\nu}$
(1)	2/3	-1/3	-1/3	2
(2)	-1/3	2/3	-1/3	2
(3)	-1/3	-1/3	2/3	2
(4)	0	0	0	2

a,b,c = 1,2,3

1	λ_a	β_b^c	8	$2\underline{\nu}$
(a)	2/3	0		2
(b)	-1/3	1		2
(c)	-1/3	-1		2
(4)	0	0		2

(ii)

		$2\underline{\nu}$	β_a^b
		α_{123}^4	λ_4	λ_a^b
λ_b	α_b^4	+1	-1	-1
	α_{b4}^{ca}	$-\lambda_b$	$+\lambda_b$	-1
λ_a^b	α_a^b	0	0	

(a,b,c) = (1,2,3,cyclic) (iii)

All e_β:

$$e_b^c = E_b^c$$

$$e_{\lambda_a} = \frac{1}{\sqrt{3}}(\sqrt{2}\,E_{a4}^{bc} + E_a^4) \qquad \text{(iv)}$$

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{2}}(E_{\lambda_4} - E_{123}^4)$$

TABLE IVA.[41]

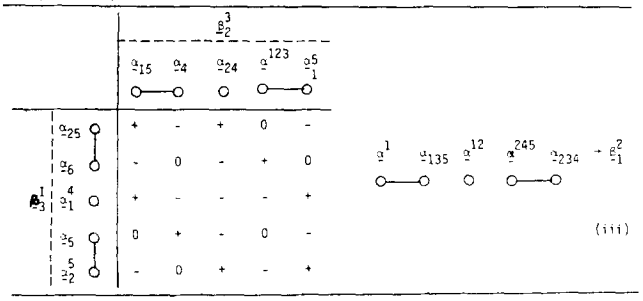
$E(6) \supset A^9(2)$ $\quad q \to (2,2);\ \mathbf{A} \to (1,4)\oplus(4,1)\oplus(1,1)$ $\quad 27 \quad 35 \quad \overline{35} \quad 8$ (i)

W_{xa}				$\bar{W}_{ia}$				9		
$\underline{u}_a$	$\underline{u}_b$	$\underline{u}_c$	No.		$\underline{v}_1$	$\underline{v}_2$	$\underline{v}_3$			
$(\bar{+})^2$	$(\bar{+})^1$	$(\bar{+})^1$	6	(1)	0	1	-1		2	-1	-1
2	-2	0	6	(2)	-1	0	1		-1	2	-1
1	-1	0	12	(3)	2	-2	0		-2	-2	4
				(4)	1	1	-2		3	-3	0
0	0	0	3	(5)	0	0	0		0	0	0
				$(\bar{6})$	0	0	0		0	0	0

(ii)

$a \neq b \neq c$

$a, b, c = 1, 2, 3$



(iii)

All $e_{\underline{\beta}}$:

$$e_2^3 = \tfrac{1}{3}\{E_{24} + \sqrt{2}\,(E_{15} + E_1^5 + E_4 + E^{123})\}$$

$$e_3^1 = \tfrac{1}{3}\{iE_1^4 + \sqrt{2}\,(E_{25} - E_2^5 + e^{-i\theta}E_5 + e^{i\theta}E_6)\}; \qquad \theta = \tfrac{\pi}{4}$$

$$e_1^2 = \tfrac{1}{3}\{-iE^{12} + \sqrt{2}\,(E^1 + E_{234} - e^{i\theta}E_{135} - e^{-i\theta}E^{245})\}$$

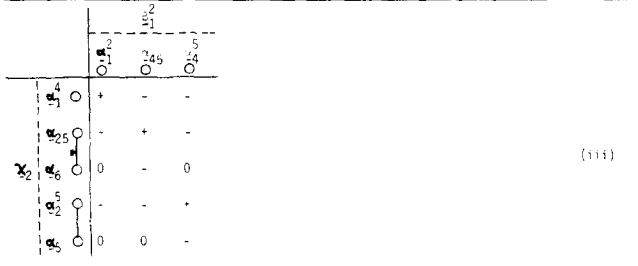
(iv)

TABLE IVB.[41]

$E(6) \supset G^3(2)$ $\quad q \to (0,2)\ ;\ \mathbf{A} \to (1,1)\oplus(1,0)$ $\quad 27 \quad 64 \quad 14$ (i)

W_{xa}				$\bar{W}_{ia}$				3		
$\underline{u}_a$	$\underline{u}_b$	$\underline{u}_c$	No.		$\underline{v}_1$	$\underline{v}_2$	$\underline{v}_3$			
$(\bar{+})2/3$	$(\bar{+})1/3$	$(\bar{+})1/3$	12	(1)	2/3	-1/3	-1/3		-1/3	1
$(\bar{+})4/3$	$(\bar{+})2/3$	$(\bar{+})2/3$	6	(2)	-1/3	2/3	-1/3		2/3	-1
				(3)	-2/3	-2/3	4/3		-2/3	0
$(\bar{+})1$	$(\bar{+})1$	0	6	(4)	1	-1	0		-1	2
0	0	0	3	(5)	0	0	0		0	0
				$(\bar{6})$	0	0	0		0	0

(ii)



(iii)

$$e_{\chi_2} = \tfrac{1}{3}\{E_1^4 + \sqrt{2}\,(E_{25} + E_2^5 + E_5 + E_6)\}$$

$$e_1^2 = \tfrac{1}{\sqrt{3}}\{E_1^2 + e^{i\theta}E_{45} + e^{-i\theta}E_4^5\}; \qquad e^{i\theta} = \frac{1+i}{2\sqrt{2}}$$

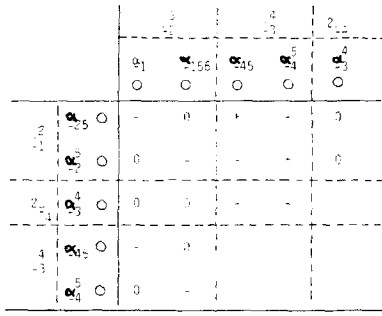
(iv)[44]

TABLE IVC.[41]

$E(6) \supset C^2(4)$ $\quad q \to (0,1,0,0);\ \mathbf{A} \to (0,0,0,1)\oplus(2,0,0,0)$ $\quad 27 \quad 42 \quad 36$ (i)

W_{xa}			$\bar{W}_{ia}$					1				
$\underline{u}_a$	$\underline{u}_b$	No.		$\underline{v}_1$	$\underline{v}_2$	$\underline{v}_3$	$\underline{v}_4$		$\underline{\alpha}_1$	$\underline{\alpha}_2$	$\underline{\alpha}_3$	$\underline{\alpha}_4$
± 1	± 1	24	(1)	1	1	0	0		0	0	0	0
0	0	3	(2)	1	-1	0	0		1	0	0	0
			(3)	0	0	1	1		0	0	0	1
$a,b = 1,\ldots,4$			(4)	0	0	1	-1		0	0	1	0
			(5)	0	0	0	0		0	0	0	0
			$(\bar{6})$	0	0	0	0		0	0	0	0

(iii)



(iv)

$$e_1^2 = \frac{1}{\sqrt{2}}\{E_{25} + E_2^5\} \qquad e_3^4 = \frac{1}{\sqrt{2}}\{E_{45} - E_4^5\}$$

$$e_2^3 = \frac{1}{\sqrt{2}}\{E_1 + E_{156}\} \qquad e_{234} = E_3^4$$

TABLE IVD.[50]

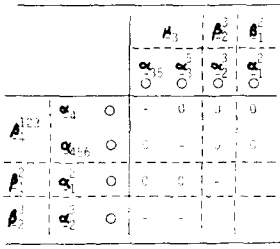
$E(6) \supset F^1(4)$ $\quad q \to (0,0,0,1)\oplus(0,0,0,0);\ \mathbf{A} \to (0,0,0,1)\oplus(1,0,0,0)$ $\quad 26 \quad 1 \quad 26 \quad 52$ (i)

W_{xa}					$\bar{W}_{ia}$					1			
M_a	M_b	M_c	M_d	No.		M_1	M_2	M_3	M_4		$\beta(a,b)$	$\beta(a)$	$\beta(abcd)$
-1	0	0	0	6	(1)	1	0	0	0	(a)	1	1	.
-2	-1	-1	-1	16	(2)	0	1	0	0	(b)	1	0	1
0	0	0	0	3	(3)	0	0	1	0	(c)	0	0	1
					(4)	0	0	0	1	(d)	0	0	1
$a,b,c,d = 1,2,3,4$					(5)	0	0	0	0	(5)	0	0	0
					$(\bar{6})$	0	0	0	0	$(\bar{6})$	0	0	0

$\beta(a,b) = \beta_{ab}$

$\beta(abcd) = \beta(a)\beta(b)\beta(c)\beta(d)$

(ii)



(iii)

All $e_{\underline{\alpha}}$:

$$e(a,b) = E(a,b); \quad e(a) = -\frac{i}{\sqrt{2}}\,E_5(a) + E^6(a);$$

$$e_4(abc) = \frac{1}{\sqrt{2}}\{E_{64}(abc\bar{5}) - d(abc)E_4^6(abc\bar{5})\}$$

$$e^4(abc) = \frac{1}{\sqrt{2}}\{d(abc)\,E_6^4(abc\bar{5}) + E^{64}(abc\bar{5})\}$$

(v)[51]

$$d(abc) = \pm(a+b)(b+c)(a+c)$$

$e_4(a,b,c) = e_{4,3}^{b}$ (etc.

TABLE IVE.[52]

$E(6) \supset G^1(2) \oplus A^2(2)$ | $q \to (0,1)\{0,1\} \oplus (0,0)\{2,0\}$; $\bar{A} \to (0,1)\{1,1\} \oplus (1,0)\{0,0\} \oplus (0,0)\{1,1\}$ (i)

21 6 56 14 8

W_{xa}

μ_a	μ_b	μ_c	ν_a	ν_b	ν_c	No.
0	0	0	$-\frac{4}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	3
0	0	0	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	6
$(-)\frac{2}{3}$	$(+)\frac{1}{3}$	$(+)\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	18

$a \neq b \neq c$, a, b, c = 1, 2, 3

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	ν_1	ν_2	ν_3
(1)	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	0	0
(2)	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$	0	0	0
(3)	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$	0	0	0
(4)	0	0	0	1	-1	0
(5)	0	0	0	0	0	0
($\bar{6}$)	0	0	0	$\frac{1}{\sqrt{3}}$	$\frac{1}{\sqrt{3}}$	$-\frac{2}{\sqrt{3}}$

1	β_b^c	γ_a
(a)	0	$\frac{2}{3}$
(b)	1	$-\frac{1}{3}$
(c)	-1	$-\frac{1}{3}$
(4)	0	0
(5)	0	0
($\bar{6}$)	0	0

2	γ_2^3	γ_3^1	γ_1^2
	0	0	0
	0	0	0
	0	0	0
	$-\frac{1}{2}$	$-\frac{1}{2}$	1
	0	0	0
	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	0

(ii)

		γ_3			β_2^3		γ_1^2	
		α_{35}	α_{23}^5	α_{12}^3	α_2^3		α_{45}	α_4^5
		O	O	O	O		O	O
γ_2^3	α_6^4 O	+	0	-	0		0	-
	α_6 O	-	0	+	0		-	0
γ_1^2	α_{45} O	+	-	0	0			
	α_4^5 O	-	+	0	0			
β_2^3	α_2^3 O	-	-	-				

(iii)

All e_β, e_γ:

G(2):

$$e_b^c = F_b^c$$

$$e_{\gamma_a} = \frac{1}{\sqrt{3}}\{(-1)^a E^{bc} + E_{a5} - E_a^5\}$$

A(2):

$$e_2^3 = \frac{1}{\sqrt{2}}\{E^4 + F_6\}$$

$$e_1^2 = \frac{1}{\sqrt{2}}\{E_{45} + F_4^5\}$$

$$e_2^1 = \frac{1}{\sqrt{2}}\{E_5 + E_{123}\}$$

(iv)

TABLE VA.[53]

$E(7) \supset A^{798}(1)$ | $q \to (27) \oplus (17) \oplus (9)$; $\bar{A} \to (34) \oplus (26) \oplus (22) \oplus (18) \oplus (14) \oplus (10) \oplus (2)$

28 18 10 35 27 23 19 15 11 3

(i)

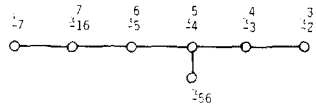
W_{xa}

	No.	
n_1	2x5	$n_1 = 27,25,23,21,19$
n_2	4x4	$n_2 = 17,15,13,11$
n_3	6x5	$n_3 = 9,7,5,3,1$

M_{ia}

		399	2λ
(1)	26		26
(2)	8		8
(3)	6		6
(4)	4		4
(5)	2		2
(6)	0		0
($\bar{7}$)	$\sqrt{2}$		$\sqrt{2}$

(ii)



$\to 2\lambda$

(iii)[27]

$\lambda_7 = \sqrt{2}\,\lambda_7$

Principal subalgebra

All e_λ:

$$e_{2\lambda} = \frac{1}{\sqrt{399}}\{\sqrt{27}\,E_{\lambda_7} + \sqrt{52}\,E_{16}^7 + \sqrt{75}\,E_5^6 + \sqrt{96}\,E_4^5 + \sqrt{66}\,E_3^4 + \sqrt{34}\,E_2^3 + \sqrt{49}\,E_{56}\}$$

(iv)

TABLE VB.[54]

$E(7) \supset A^{462}(1)$

$q \to (21)\oplus(15)\oplus(11)\oplus(5)$; $\mathbf{A} \to (26)\oplus(22)\oplus(18)\oplus(16)\oplus(14)\oplus 2(10)\oplus(6)\oplus(2)$

22 16 12 6 27 23 19 17 15 22 7 3 (i)

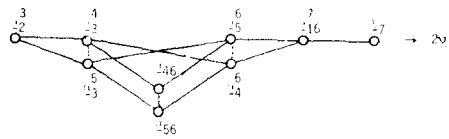
W_{xa}

	No.	
$^{+}n_1$	2×3	$n_1 = 21,19,17$
$^{+}n_2$	4×2	$n_2 = 15,13$
$^{+}n_3$	6×3	$n_3 = 11,9,7$
$^{+}n_4$	8×3	$n_4 = 5,3,1$

M_{ia}

	231	2
(1)	20	20
(2)	6	6
(3)	4	4
(4)	2	2
(5)	2	2
(6)	0	0
(7)	$\sqrt{2}$	$\sqrt{2}$

(ii)



(iii)

All $e_{\underline{a}}$:

$$e_{2_2} = \frac{1}{\sqrt{231}}\{21\,E_{\underline{27}} + 40\,E^{7}_{16} + \sqrt{6}\,(e^{i\theta x}f^{6}_{5} + e^{-i\theta x}f^{6}_{4}) + \sqrt{7}\,(e^{i\theta y}E_{56} + e^{-i\theta y}E_{16}) + \sqrt{2}(E^{4}_{3} + E^{5}_{3}) + 26\,E^{3}_{2}\};$$

$$x = \frac{57}{2}, \quad y = \frac{50}{2}, \quad z = \frac{37}{2}, \qquad \text{(iv)}^{55}$$

$$\sin^2\theta_x = -\left(\frac{y^2 - (x-z)^2}{4xz}\right)^{\frac{1}{2}} \qquad \sin^2\theta_y = \left(\frac{x^2 - (y-z)^2}{4yz}\right)^{\frac{1}{2}} \qquad \theta_y = 2$$

TABLE VC.[50]

$E(7) \supset A^{21}(2)$

$q \to (6,0)\oplus(0,6)$; $\mathbf{A} \to (4,4)\oplus(1,1)$

28 28 125 8 (i)

W_{xa}

μ_a	μ_b	μ_c	No.
$^{+}/_{-}4$	$^{+}/_{-}2$	$^{+}/_{-}2$	6
$^{+}/_{-}3$	$^{+}/_{-}2$	$^{+}/_{-}1$	12
2	−2	0	12
$^{+}/_{-}2$	$^{+}/_{-}1$	$^{+}/_{-}1$	12
1	−1	0	12
0	0	0	2

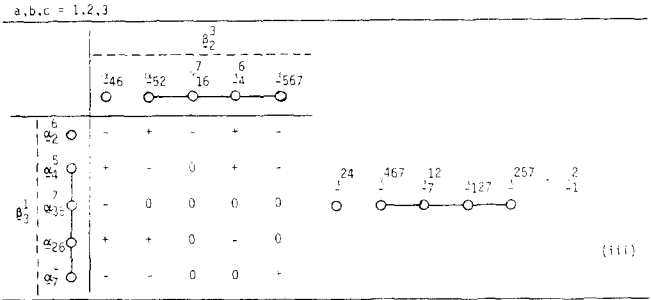
$a \neq b \neq c$

$a,b,c = 1,2,3$

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	21^1_1	β^3_2	β^1_3	β^2_1
(1)	3	−1	−2	1	−5	4	
(2)	−1	0	1	−1	2	−1	
(3)	1	−3	2	−5	1	4	
(4)	0	1	−1	2	−1	−1	
(5)	1	1	−2	3	−3	0	
(6)	0	0	0	0	0	0	
$(\bar{7})$	$\sqrt{2}$	$-\sqrt{2}$	0	$-\sqrt{2}$	$-\sqrt{2}$	$2\sqrt{2}$	

(ii)



(iii)

All $e_{\underline{a}}$:

$$e^3_2 = \frac{1}{\sqrt{21}}\{E_{46} + \sqrt{6}\,(E^6_4 + E^7_{16}) + 2(E_{52} + E_{567})\};$$

$$e^1_3 = \frac{1}{\sqrt{21}}\{E^6_2 + \sqrt{6}\,(-E_{26} + E^7_{35}) + 2(-E^5_4 + E^{-}_{7})\}; \qquad \text{(iv)}$$

$$e^2_1 = \frac{-1}{\sqrt{21}}\{E^{42} + \sqrt{6}\,(E_{127} + E^{12}_{7}) + 2(E^{467} + E^{257})\}.$$

TABLE VD.[56]

$E(7) \supset A^{48}(1)\oplus A^{30}(1)$

$q \to (6)(3)\oplus(4)(1)\oplus(2)(5)$; (i)

28 10 18

$\mathbf{A} \to (8)(2)\oplus(6)(4)\oplus(4)(6)\oplus(4)(2)\oplus(2)(4)\oplus(2)(0)\oplus(0)(2)$

27 35 35 15 15 3 3

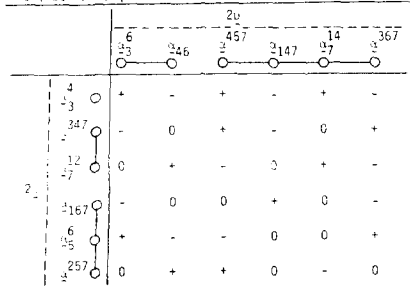
W_{xa}

		No.
+6	±3	4
±6	±1	4
±4	±3	4
+4	±1	8
−2	±5	4
−2	±3	8
−2	±1	12
0	−5	2
0	±3	4
0	±1	6

M_{ia}

	μ	ν	24	2μ	15	2ν
(1)	4	3		4		3
(2)	4	−3		4		−3
(3)	2	1		2		1
(4)	2	−1		2		−1
(5)	0	3		0		3
(6)	0	1		0		1
$(\bar{7})$	$2\sqrt{2}$	0		$2\sqrt{2}$		0

(ii)



(iii)

All $e_{\underline{a}}$:

$$e_{2_\mu} = \frac{1}{\sqrt{12}}\{(E^6_3 + E_{46}) + \sqrt{2}\,(E^{367} - E^{457}) + \sqrt{3}\,(E_{147} + E^{14}_{7})\};$$

$$e_{2_\nu} = \frac{1}{\sqrt{15}}\{(E^4_3 + 2E^6_5 + \sqrt{2}\,(E^{347} + E^{12}_{7}) + \sqrt{3}\,(E^{257} + E_{167})\} \qquad \text{(iv)}$$

TABLE VE.[57]

$E(7) \supset G^2(2) \oplus A^{14}(1)$ | $q \to (1,0)(1) \oplus (0,1)(3)$; $\bar{A} \to (0,2)(2) \oplus (0,1)(4) \oplus (1,0)(0) \oplus (0,0)(2)$ (i)

28 28 81 35 14 3

W_{xa}

ν_a	μ_b	μ_c	ν	No.
$(-)^{\frac{2}{3}}$	$(+)^{\frac{1}{3}}$	$(+)^{\frac{1}{3}}$	± 3	12
$(-)^{\frac{2}{3}}$	$(+)^{\frac{1}{3}}$	$(+)^{\frac{1}{3}}$	± 1	24
1	-1	0	± 1	12
0	0	0	± 3	2
0	0	0	± 1	6

$a \neq b \neq c$
$a,b,c = 1,2,3$

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	ν	7	2ν
(1)	2/3	-1/3	-1/3	1		1
(2)	-1/3	2/3	-1/3	1		1
(3)	-1/3	-1/3	2/3	1		1
(4)	2/3	-1/3	-1/3	-1		-1
(5)	-1/3	2/3	-1/3	-1		-1
(6)	-1/3	-1/3	2/3	-1		-1
($\bar{7}$)	0	0	0	$2\sqrt{2}$		$2\sqrt{2}$

2	ν_a	μ_a^b
(a)	2/3	1
($\dot{a}$)	2/3	1
(b)	-1/3	-1
($\dot{b}$)	-1/3	-1
(c)	-1/3	0
($\dot{c}$)	-1/3	0
($\bar{7}$)	0	0

$a = a + 3$, $(a,b,c) = (1,2,3,\text{cyclic})$ (ii)

		2ν					μ_a^b	
		α_a	α_b	α_c	α_7	$\alpha_{\bar{7}}$	β_a^b	$\beta_{\dot{a}}^{\dot{b}}$
χ_b	$\mu^{a\dot{c}}$	-	0	+	-	+	-	0
	$\beta^{\dot{a}c}$	+	0	-	-	+	0	-
	$\alpha^{\dot{a}\dot{c}7}$	+	0	+	0	-	0	-
	β_7^{ac}	-	0	-	+	0	-	0
μ_a^b	β_a^b	+	-	0	0	0		
	$\beta_{\dot{a}}^{\dot{b}}$	-	+	0	0	0		

(iii)

$\alpha_7^{\dot{} } = \frac{1}{2}\{\pm \sum_{a=1}^{3} (a + \dot{a}) + \sqrt{2}\,(\bar{7})\}$

All $e_{\underline{\beta}}$:

$$e_{\chi_a} = \frac{1}{\sqrt{6}} \{\tfrac{1}{2}(b-c)(E^{\dot{b}c} - E^{b\dot{c}}) + \sqrt{2}((-1)^a E^{\dot{b}\dot{c}7} + E_7^{bc})\}$$

$$e_b^a = \frac{(-1)^c}{\sqrt{2}} (E_b^a + E_{\dot{b}}^{\dot{a}})$$

$$e_{2\mu} = \frac{1}{\sqrt{7}} \{\sum_{a=1}^{3} E_a^{\dot{a}} + \sqrt{2}\,(E_7^+ + E_{\bar{7}}^-)\}$$

(iv)

TABLE VF.[52]

$E(7) \supset C^2(3) \oplus G^1(2)$ | $q \to (1,0,0)(0,1) \oplus (0,0,1)(0,0)$; $\bar{A} \to (0,1,0)(0,1) \oplus (2,0,0)(0,0) \oplus (0,0,0)(1,0)$ (i)

42 14 98 21 14

W_{xa}

μ_a	μ_b	μ_c	ν_a	ν_b	ν_c	No.
± 1	± 1	± 1	0	0	0	8
± 1	0	0	$(-)^{\frac{2}{3}}$	$(+)^{\frac{1}{3}}$	$(+)^{\frac{1}{3}}$	36
± 1	0	0	0	0	0	12

$a \neq b \neq c$
$a,b,c = 1,2,3$

$\bar{M}_{ia}$

	μ_1	μ_2	μ_3	ν_1	ν_2	ν_3
(1)	1	1	0	0	0	0
(2)	1	-1	0	0	0	0
(3)	0	0	0	0	0	0
(4)	0	0	0	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$
(5)	0	0	0	$-\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$
(6)	0	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{2}{3}$
(7)	0	0	$\sqrt{2}$	0	0	0

2	β_1^2	β_2^3	$2\nu_3$
	0	1	0
	2	-1	0
	0	0	0
	0	0	0
	0	0	0
	0	0	0
	0	$-\sqrt{2}$	$2\sqrt{2}$

1	γ_1^2	χ_2
	0	0
	0	0
	0	0
	1	$-\frac{1}{3}$
	-1	$\frac{2}{3}$
	0	$-\frac{1}{3}$
	0	0

(ii)

		β_1^2	$2\nu_3$	λ_1^2	χ_2		
		α_{23}	β_2^3	λ_7	α^{45}	χ_{35}	χ^{46}
β_2^3	β_{13}^7	0	-	-	0	0	-	+
	β^{237}	-	0	-	0	0	+	-
χ_2	χ_{35}	+	-	0	-			
	χ_5^3	-	+	0	-			
	χ^{46}	0	0	0	-			
γ_1^2	γ_4^5	0	0	0				
$2\nu_3$	λ_7	0	0					

(iii)

C(3):

$$e_1^2 = \frac{1}{\sqrt{2}} (E_{23} + E_2^3)$$

$$e_2^3 = \frac{1}{\sqrt{2}} (E_{13}^7 + E^{237})$$

$$e_{2\nu_3} = E_{\lambda_7}$$

G(2):

$$e_1^2 = E_4^5$$

$$e_{\chi_2} = \frac{1}{\sqrt{3}} (E_{35} + E_5^3 + E^{46})$$

(iv)

TABLE VG.[57]

$E(7) \supset F^1(4) \oplus A^6(1)$ | $q = (0,0,0,1)(1) \oplus (0,0,0,0)(3)$ (52, 4)

$\mathcal{A} = (0,0,0,1)(2) \oplus (1,0,0,0)(0) \oplus (0,0,0,0)(2)$ (78, 52, 3) (i)

W_{xa}

ν_a	ν_b	ν_c	ν_d	ν	No.
±1	0	0	0	·1	16
±½	±½	±½	±½	·1	32
0	0	0	0	·3	2
0	0	0	0	·1	6

$a \neq b \neq c \neq d$,
$a,b,c,d = 1,\ldots,4$

M_{ia}

	ν_1	ν_2	ν_3	ν_4	ν
(1)	1	0	0	0	0
(2)	0	1	0	0	0
(3)	0	0	1	0	0
(4)	0	0	0	1	0
(5)	0	0	0	0	0
(6)	0	0	0	0	2
($\tilde{7}$)	0	0	0	0	·2

1	$\underline{\nu}(a,b)$	$\underline{\nu}(a)$	$\underline{\nu}(a,b,c,d)$	3 \ $2\underline{\nu}$
(a)	1	1	½	0
(b)	1	0	½	0
(c)	0	0	½	0
(d)	0	0	½	0
(5)	0	0	0	0
(6)	0	0	0	2
($\tilde{7}$)	0	0	0	·2

$\{\cdot,a\} = \cdot\, \underline{\nu}_a$

$\underline{\beta}(\cdot,a) = \cdot\, \underline{\nu}_a$

$\underline{\beta}(a,b) = \underline{\nu}(a) + \underline{\nu}(b)$

$\underline{\beta}(a,b,c,d) = \tfrac{1}{2}\{\underline{\nu}(a) + \ldots + \underline{\nu}(d)\}$ (ii)

		$2\underline{\nu}$	$\underline{\alpha}_1^2$	$\underline{\alpha}_2^3$	$\underline{\alpha}_3$	
		$\underline{\alpha}_{56}$ ○	$\underline{\alpha}_6^5$ ○	$\underline{\alpha}_7$ ○	$\underline{\alpha}_1^2$ ○	$\underline{\alpha}_2^3$ ○	$\underline{\alpha}_{35}$ ○	$\underline{\alpha}_3^5$ ○
$\underline{\alpha}_4^{123}$	$\underline{\alpha}_{46}^7$ ○	0	+	-	0	0	-	0
	$\underline{\alpha}_{457}$ ○	0	-	+	0	0	0	-
$\underline{\alpha}_3$	$\underline{\alpha}_{35}$ ○	+	-	0	0	-		
	$\underline{\alpha}_3^5$ ○	-	+	0	0	-		
$\underline{\alpha}_{12}^3$	$\underline{\alpha}_{12}^3$ ○	0	0	0	-			
$\underline{\alpha}_1^2$	$\underline{\alpha}_1^2$ ○	0	0	0				

(iii)

All $e_{\underline{\nu}}$:

$e(a,b) = E(a,b)$; $e(a) = -\frac{1}{\sqrt{2}}\{E_5(a) + E^5(a)\}$

$e_4(a,b,c) = \frac{1}{\sqrt{2}}\{E^7_{64}(a,b,c,\cdot 5) - d(a,b,c)\, E^6_{74}(a,b,c,\tilde{7}5)\}$

$e^4(a,b,c) = \frac{1}{\sqrt{2}}\{d(a,b,c)\, E^{74}_6(a,b,c,\tilde{7}5) + E^{64}_7(a,b,c,\cdot 5)\}$

$d(a,b,c) = \cdot\,\epsilon(a+b)\,\epsilon(b+c)\,\epsilon(a+c)$ (iv)

TABLE VIA.[58]

$E(8) \supset A^{2480}(1)$ | $q = \mathcal{A} = (58) \oplus (46) \oplus (38) \oplus (34) \oplus (26) \oplus (22) \oplus (14) \oplus (2)$ (59, 47, 39, 35, 27, 23, 15, 3) (i)

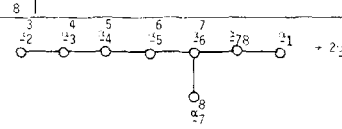
W_{xa}

ν	No.	
$\cdot n_1$	2x6	$n_1 = 58,56,\ldots,50,48$
$\cdot n_2$	4x4	$n_2 = 46,44,42,40$
$\cdot n_3$	6x2	$n_3 = 38,36$
$\cdot n_4$	8x4	$n_4 = 34,32,30,28$
$\cdot n_5$	10x2	$n_5 = 26,24$
$\cdot n_6$	12x4	$n_6 = 22,20,18,16$
$\cdot n_7$	14x6	$n_7 = 14,\ldots,4$
·2	16	
0	8	

M_{ia}

	$\underline{\nu}$	1240 \ $2\underline{\nu}$
(1)	46	46
(2)	12	12
(3)	10	10
(4)	8	8
(5)	6	6
(6)	4	4
(7)	2	2
(8)	0	0

(ii)



(iii)

Principal subalgebra

All $e_{\underline{\nu}}$:

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{620}}\{\sqrt{29}\,E^3_2 + \sqrt{57}\,E^4_3 + \sqrt{84}\,E^5_4 + \sqrt{110}\,E^6_5 + \sqrt{135}\,E^7_6 + \sqrt{91}\,E_{78} + \sqrt{46}\,E_1 + \sqrt{68}\,E^8_7\}$$ (iv)

TABLE VIB.[54]

$E(8) \supset A^{1520}(1)$ | $q = \mathcal{A} = (46) \oplus (38) \oplus (34) \oplus (28) \oplus (26) \oplus (22) \oplus (18) \oplus (14) \oplus (10) \oplus (2)$ (47, 39, 35, 29, 27, 23, 19, 15, 11, 3) (i)

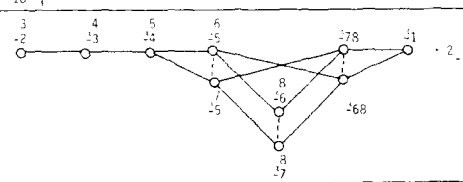
W_{xa}

	No.	
$\cdot n_1$	2x4	$n_1 = 46,44,42,40$
$\cdot n_2$	4x2	$n_2 = 38,36$
$\cdot n_3$	6x3	$n_3 = 34,32,30$
·28	8	
$\cdot n_4$	10x2	$n_4 = 26,24$
$\cdot n_5$	12x2	$n_5 = 22,20$
$\cdot n_6$	14x2	$n_6 = 18,16$
$\cdot n_7$	16x2	$n_7 = 14,12$
$\cdot n_8$	18x4	$n_8 = 10,8,6,4$
·2	20	
0	10	

M_{ia}

	$\underline{\nu}$	760 \ $2\underline{\nu}$
(1)	36	36
(2)	10	10
(3)	8	8
(4)	6	6
(5)	4	4
(6)	2	2
(7)	2	2
(8)	0	0

(ii)



(iii)

All $e_{\underline{\nu}}$:

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{760}}\{\sqrt{46}\,E^3_2 + \sqrt{90}\,E^4_3 + \sqrt{132}\,E^5_4 + \sqrt{x}\,(e^{i\pi x}E_{68} + e^{-i\pi x}E_{78}) + \sqrt{y}\,(e^{i\pi y}E^8_6 + e^{-i\pi y}E^8_7) + \sqrt{z}\,(E^7_5 + E^6_5) + \sqrt{72}E_1\}$$

$x = 71$, $y = 53$, $z = 86$

$\sin\theta_x = \left\{\frac{y^2 - (x-z)^2}{4xz}\right\}^{\frac{1}{2}}$, $\sin\theta_y = \left\{\frac{x^2 - (y-z)^2}{4yz}\right\}^{\frac{1}{2}}$, $-\frac{\pi}{2} \leq \theta_x \leq 0$, $0 \leq \theta_y \leq \frac{\pi}{2}$ (iv)[59]

TABLE VIC.[43]

$E(8) \supset A^{1040}(1)$ | $q = \mathbf{A} + (38) \oplus (34) \oplus (28) \oplus (26) \oplus 2(22) \oplus (18) \oplus (16) \oplus (14) \oplus (10) \oplus (6) \oplus (2)$

39 35 29 27 46 19 17 15 11 7 3 (i)

W_{xa}

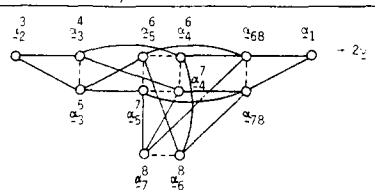
$\underline{\nu}$	No.	$\underline{\nu}$	No.
$\pm n_1$	2x2	± 16	16
$\pm n_2$	4x3	$\pm n_5$	18x2
± 28	6	$\pm n_6$	20x2
$\pm n_3$	8x2	$\pm n_7$	22x2
$\pm n_4$	12x2	± 2	24
± 18	14	0	12

$n_1 = 38,36$; $n_2 = 34,32,30$; $n_3 = 26,24$; $n_4 = 22,20$; $n_5 = 14,12$; $n_6 = 10,8$; $n_7 = 6,4$

M_{ia}

	$\underline{\nu}$	520	$2\underline{\nu}$
(1)	30		30
(2)	8		8
(3)	6		6
(4)	4		4
(5)	4		4
(6)	2		2
(7)	2		2
(8)	0		0

(ii)



(iii)

All $e_{\underline{\theta}}$:

$$e_{2\underline{\nu}} = \frac{1}{\sqrt{520}}\left(\sqrt{39}\,\xi_2^3 + \sqrt{\frac{71}{2}}\left[E_4^6 + E_5^7 + e^{i\zeta}(E_5^6 + E_4^7)\right] + \sqrt{x}(e^{i\theta_x}E_{78} + e^{-i\theta_x}E_{68}) + \sqrt{y}(e^{i\theta_y}E_7^8 + e^{-i\theta_y}E_6^8) + \sqrt{z}(E_3^4 + E_3^5) + \sqrt{60}\,E_1\right)$$

$\cos\zeta = \frac{37}{71}$; $x = 59$, $y = 44$, $z = 37$, (iv)[60]

$$\sin\theta_x = -\left(\frac{y^2 - (x-z)^2}{4xz}\right)^{1/2}, \quad \sin\theta_y = \left(\frac{x^2 - (y-z)^2}{4yz}\right)^{1/2}, \quad -\frac{\pi}{2} \le \theta_x \le 0, \quad 0 \le \theta_y \le \frac{\pi}{2}$$

TABLE VID.[50]

$E(8) \supset C^{24}(2)$ | $q = \mathbf{A} + (6,0) \oplus (2,3) \oplus (2,0)$

84 154 10 (i)

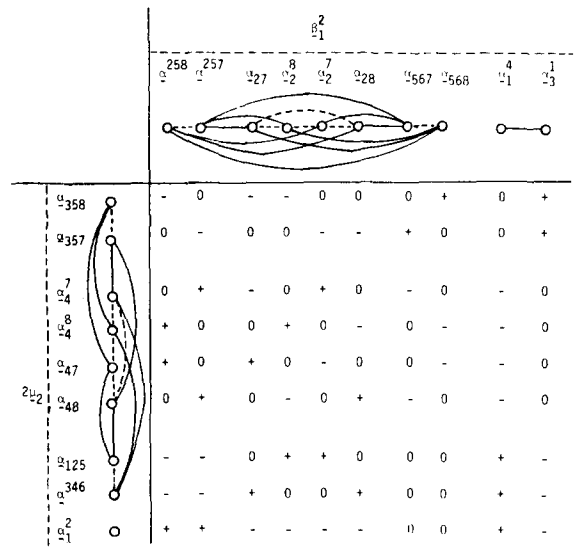
W_{xa}

$\underline{\nu}_a$	$\underline{\nu}_b$	No.
± 6	0	4
± 5	± 3	8
± 5	± 1	16
± 4	± 4	4
± 4	± 2	24
± 4	0	16
± 3	± 3	16
± 3	± 1	48
± 2	± 2	24
± 2	0	36
± 1	± 1	40
0	0	12

M_{ia}

	$\underline{\mu}_1$	$\underline{\mu}_2$	12	$2\underline{\mu}_2$	β_1^2
(1)	1	1		1	0
(2)	1	-1		-1	1
(3)	2	0		0	1
(4)	0	2		2	-1
(5)	3	3		3	0
(6)	3	-3		-3	3
(7)	0	0		0	0
(8)	0	0		0	0

(ii)



-	0	-	-	0	0	0	+	0	+
0	-	0	0	-	-	+	0	0	+
0	+	-	0	+	0	-	0	-	0
+	0	0	+	0	-	0	-	-	0
+	0	+	0	-	0	0	-	-	0
0	+	0	-	0	+	-	0	-	0
-	-	0	+	+	0	0	0	+	-
-	-	+	0	0	+	0	0	+	-
+	+	-	-	-	-	0	0	+	-

(iii)

$$e_{2\underline{\mu}_2} = \frac{1}{\sqrt{12}}\left(\tfrac{1}{2}(1+i)(E_{47} + E_4^7) + \tfrac{1}{2}(1-i)(E_4^8 - E_{48}) + \sqrt{\tfrac{3}{2}}(E_{125} + E^{346} + E_{357} + E_{358}) + E_1^2\right)$$

$$e_1^2 = \frac{1}{\sqrt{12}}\left\{-(E_{27} + E_2^8) + i(E_2^7 + E_{28}) + (E_1^4 - E_3^1) + \sqrt{\tfrac{3}{2}}(E_{567} + E_{568} + E^{258} + E^{257})\right\}$$

(iv)[61]

TABLE VIE.[37]

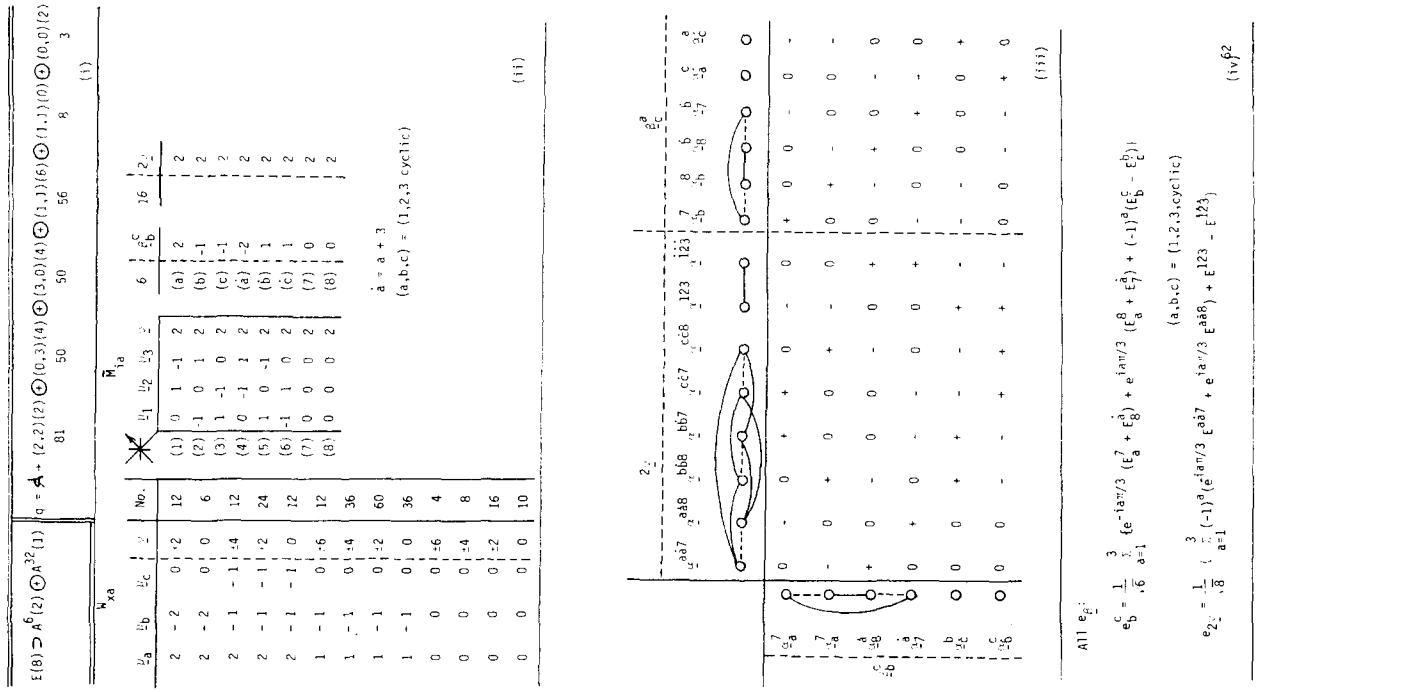
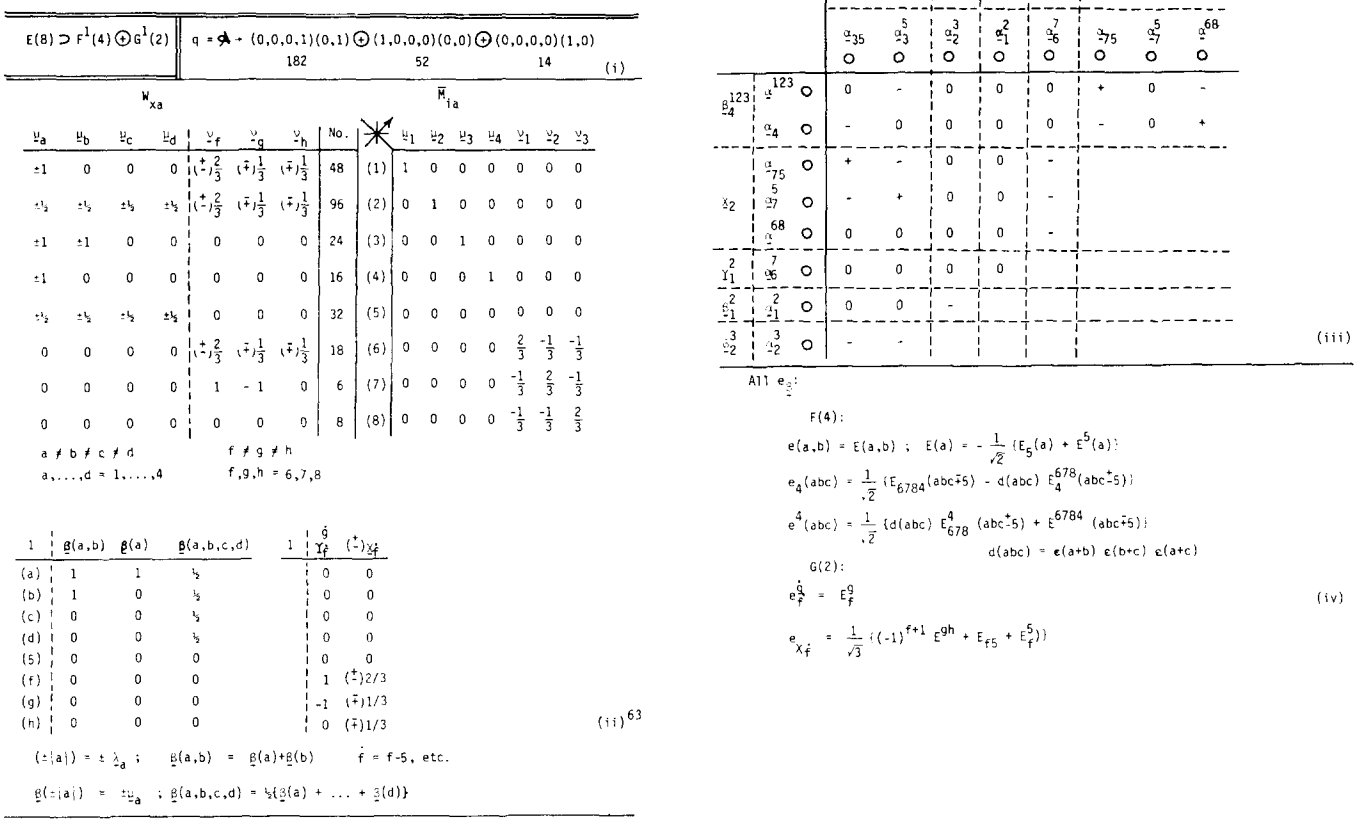


TABLE VIF.[56]



[1]E. B. Dynkin, Am. Math. Soc. Transl. **17**, (1950); maximal regular subalgebras have also been discussed earlier: A. Borel and J. de Siebenthal, Comm. Math. Helv. **23**, 200 (1949).

[2]E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2 **6**, 111, 245 (1957).

[3]G. Feldman, T. Fulton, and P. T. Matthews, J. Phys. G. Nucl. Phys. **8**, 295 (1982).

[4]G. Feldman, T. Fulton, and P. T. Matthews, J. Math. Phys. **25**, 1222 (1984), hereafter referred to as FFM.

[5]The indices i, j will, in what follows, be restricted to range from 1 to n_G (the rank of G), and the α_i, when they appear, will run over a set of simple roots. We will use p,q,m to label any root $\alpha(p)$ etc. of G. The indices a,b,c will generally run from 1 to $n_{\mathbf{g}}$ (the rank of g) [except for $g = A(2)$ and $G(2)$, in which case they run from 1 to $n_g + 1$]. The set β_a of roots of g, when they appear, will be a set of simple roots [except for $A(2)$ and $G(2)$]. The indices r,t,u will be used to label any root of g.

[6]In place of $N_{\alpha(p),\alpha(q)}$, we will often use the notation N^{m}_{pq} and the symbol N^{-m}_{-p-q} will be used for $N_{-\alpha(p),-\alpha(q)}$.

[7]This is the notation of Ref. 2; in Ref. 3, $\Lambda \rightarrow \mathbf{r}$. The vector Λ is not to be confused with the orthogonal basis vectors λ_i, introduced later.

[8]Conventional normalization is to take $\alpha_L^2 = 2$, where $\alpha_L(p)$ are the set of longest roots.

[9]From the results of Dynkin (Tables 15, 24, and 35 of Ref. 2) it appears that some of his expressions for the subgroup generators do not satisfy (2.8). However, except in one case, this is entirely due to misprints or trivial factors of i which can be absorbed into the definitions of some generators (see Sec. 5). The one exception, $E(6) \supset G(2)$, we believe is due to an error on Dynkin's part. This point is discussed in detail in Appendix A.

[10]We have consistently used (2.10) throughout the present work, with the particular phase choices given in FFM. Thus, the generators of all the subgroups of a given group will have consistent phase choices. Dynkin restricts himself to the consideration of simple roots, and their corresponding generators, and to single subalgebras of a given algebra. The simultaneous entertaining of possible inconsistent choices does not create any difficulties for him. Indeed, it makes phase choices an easier process for him than it does for us in what follows below.

[11]Dynkin[2] adopts a more abstract approach than ours. He treats the weight spaces of G and g independently, nomalizes both $|\alpha_L|^2$ *and* $|\beta_L|^2$ to 2, and defines the index in terms of the mapping which relates the two spaces.

[12]See Ref. 2, *R*-subalgebras, Tables 8, 9, and 10, p. 145, 147. Note that, when a regular subalgebra is derived from a Dynkin diagram by the removal of a vertex associated with root $\alpha(p)$, a $U(1)$ term may appear. It is not included by Dynkin in his tables. These terms are essential to the present discussion.

[13]W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representation of Simple Lie Algebras* (Marcel Dekker, New York, 1981).

[14]Cf. Ref. 12 (and also Ref. 2, Table 12a, p. 151), but note the isomorphisms: $A(1) \sim B(1) \sim C(1)$, $D(1) \sim U(1)$, $B(2) \sim C(2)$, $D(2) \sim A(1) \oplus A(1)$, $A(3) \sim D(3)$. Using the quark weight arguments of the present work, the labels we have chosen in Eqs. (3.8)–(3.11) are the natural ones. Note that Ref. 2, Table 12, contains additional subgroups listings to the ones we give here. These are in error. They are not maximal, but subgroups of the groups already listed in Eqs. (3.13)–(3.15). See Borel and Siebenthal, Ref. 1.

[15]As mentioned previously, r is used to label the roots $\beta(r)$ of g. The label l will run over L values where L depends on G, g, and $\beta(r)$. See the discussion following Eq. (4.22) below.

[16]W. McKay, J. Patera, and D. Sankoff, in *Computers in Nonassociative Rings and Algebras*, edited by R. E. Beck and B. Kalmar (Academic, New York, 1977).

[17]Although we use the index i to label the components of the vector operator $\mathbf{H}$ in this case, we do not imply that we are taking components of $\mathbf{H}$ in the λ_i basis. Rather, as presently used, i is a component in any orthogonal basis such that the first n_g components of $\mathbf{H}$ will be the generators common to both g and G.

[18]Care must be taken to include in the sum the degeneracy associated with a given Λ.

[19]c^{ϕ} is related to the index of the representation, l_{ϕ}, defined in Ref. 2, by the equation $c^{\phi} = \frac{1}{2}\alpha_L^2 l_{\phi}$, where α_L^2 is the square of the length of a long root. c^{ϕ} is of dimension length-squared in weight space. In Ref. 2, α_L^2 is set equal to 2.

[20]R. Slansky, Phys. Rep. **79**, 1 (1981).

[21]J. Patera, R. T.Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976).

[22]When g is $A(n)$ or $G(2)$, the space of basis vectors is $(n_g + 1)$-dimensional. The generalization to these cases is described below.

[23]The discussion applies to $G(2)$ because we choose as orthogonal basis vectors the three λ_i's corresponding to the regular $A(2)$ subalgebra of $G(2)$. Alternatively, we could have chosen the two λ_i's corresponding to the regular $A(1) \oplus A(1)$ subalgebra of $G(2)$.

[24]The generalization to the case when g is $A(n)$ is obvious. The right side of (4.31) will read $s^2[\delta_{ab} - 1/(n+1)]$, $a,b = 1,\ldots,n+1$.

[25]We treat the case of $G = G(2)$ separately, since for $G(2)$ we use a labeling[23] which depends on the regular $A(2)$ subalgebra of $G(2)$.

[26]We distinguish the "vector" weights appearing in Table 2 of FFM from the "spinor" weights. The latter are weights of a spinor representation of the relevant $D(n)$ or $B(n)$ used to label the exceptional groups.

[27]We use the symbols $\bar{\lambda}_6$ in $E(6)$ and $\bar{\lambda}_7$ in $E(7)$ to define vectors with "standard" normalizations: $(\bar{\lambda}_6)^2 = (\bar{\lambda}_7)^2 = \lambda^2$. The vectors $\bar{\lambda}_6$ and $\bar{\lambda}_7$ play special roles in $E(6)$ and $E(7)$, respectively, different from those of all other λ_i's. The symbols λ_6 in $E(6)$ and λ_7 in $E(7)$ will also find later use. These vectors have normalizations $(\lambda_6)^2 = 3\lambda^2$ and $(\lambda_7)^2 = 2\lambda^2$. The vectors $\bar{\lambda}_6$ and $\bar{\lambda}_7$ were introduced in FFM [cf. FFM Eqs. (A11) and (A12)] in order to unify the characterizations of $E(6)$, $E(7)$, and $E(8)$ with the aid of the property $E(8) \supset E(7) \supset E(6)$.

[28]For one case, $g = A^{462}_{(1)}$, we could take $\bar{\lambda}_7 \rightarrow 0$, since here two roots of $E(7)$ project to zero. However, for other reasons of simplification, we do not choose to avail ourselves of this opportunity (see Table VB).

[29]If all the roots of g are of the same length, $s'(r) = s$ for all r.

[30]The convenience of introducing the factor $s'(r)$ in the expression (4.38) for $e_{\beta(r)}$ now becomes evident.

[31]The properties (4.55) and the scalar product configurations given by Figs. 1(a) and (b) can be generalized to include the case of G being $F(4)$. However, because $F(4)$ contains both small and large roots, the relevant configurations are more numerous. Since, as seen by the results shown in Tables III A and III B, the $F(4)$ subalgebras are simple enough to work with directly, one need not bother with the general rules for $F(4)$.

[32]At least two pairs must contribute to (4.53); otherwise, one $B(al)$ coefficient would be identically zero.

[33]See Ref. 2, p. 231, bottom. Note the typo in the third row: $A_1^9 \rightarrow A_2^9$.

[34]The zero weights, equal in number to the rank, are not listed for the adjoint in these tables.

[35]In a few cases only the simple roots, $\beta(a)$, are listed. In most cases, all the roots, $\beta(r)$, are given.

[36]These diagrams are neither Dynkin diagrams[2] (except in the case of the principal subalgebras) nor extended Dynkin diagrams.[2] The latter are used by Dynkin to obtain the regular subalgebras.

[37]Cf. Ref. 2, p. 227, Table 35.

[38]G. Feldman and R. Holman, J. Phys. G. Nucl. Phys. **9**, 7 (1983).

[39]R. Holman, J. Phys. G. Nucl. Phys. **9**, 35 (1983).

[40]The cases of $E(6) \supset A(2)$ and $E(7) \supset A(2)$ appear not to satisfy the reality condition (2.8), but this is not so. For $E(6) \supset A(2)$ (Ref. 2, p. 193) there is a misprint in the last coefficient of e_{-y}: the term should be $\mp ie'_{234}$. For $E(7) \supset A(2)$ (Ref. 2, p. 194), there is a misprint in the first coefficient of e_{-y}: the term should be $-2e'_{5432}$. We must, of course, also bear in mind that additional i factors are introduced by the nested commutator definition (5.7) of generators in Ref. 2.

[41]Ref. 2, p. 193, Table 24.

[42]Equation (A13) is the transcription of Eq. (2.9) in terms of the $E(6)$ phases defined in FFM.

[43]Ref. 2, p. 173, Table 15, $A_1^{520} \subset E_8$.

[44]We have chosen our projection matrix so that in fact only $D(8)$ phases appear.

[45]We could replace $-\beta(a)$ by $\beta(a)$ or 2ν by -2ν in (B24).

[46]Cf. Ref. 2, p. 176, Table 16.

[47]Cf. Ref. 2, p. 177, Table 17.

[48]Cf. Ref. 2, p. 224, Table 35. Note the typographical error: $\sqrt{2}e_{1234} \rightarrow e_{1234}$.

[49]See Appendix A for relation between our solution and that of Dynkin.[2]

[50]Cf. Ref. 2, p. 194, Table 24.

[51]The table of scalar products for all the roots (rather than just the simple roots) could have been constructed in (iii). However, to obtain the results in (iv), it is just as direct to make use of the $F(4)$ phase equations (3.7), (3.9), (3.10), (3.41)–(3.44) in FFM.

[52]Cf. Ref. 2, p. 224, Table 35.

[53]Cf. Ref. 2, p. 181, Table 19.

[54]Cf. Ref. 2, p. 173, Table 15.

[55]For an alternative solution, which is less symmetric than the one given here [see Eq. (B40), for a similar case], see Ref. 54. That solution, in our notation, is

$$e_{2v} = (1/\sqrt{231})\big[\sqrt{21}E_{\lambda_7} + \sqrt{40}E^7_{16} + \sqrt{2x}(\cos\theta_x\, E^6_4 - \sin\theta_x\, E^6_5) + \sqrt{2y}(\cos\theta_y\, E_{46} + \sin\theta_y\, E_{56}) + \sqrt{2z}E^5_3 + \sqrt{26}E^3_2\big].$$

Also compare this table (VB) with Tables VIB and VIC.

[56]Cf. Ref. 2, p. 226, Table 35.

[57]Cf. Ref. 2, p. 225, Table 35.

[58]Cf. Ref. 2, p. 185, Table 20.

[59]For an alternative solution, which is less symmetric than the one given here [see Eq. (B40) and Tables VB and VIC for a similar case], see Ref. 43. (Note the typographical error in Ref. 54: $\sqrt{32}\to\sqrt{132}$.) That solution, in our notation, is

$$e_{2v} = (1/\sqrt{760})\big[\sqrt{46}E^3_2 + \sqrt{90}E^4_3 + \sqrt{132}E^5_4 + \sqrt{2x}(\cos\theta_x\, E_{78} - \sin\theta_x\, E_{68}) + \sqrt{2y}(\cos\theta_y\, E^8_7 + \sin\theta_y\, E^8_6) + \sqrt{2z}E^6_5 + \sqrt{72}E_1\big].$$

Also compare this table (VIB) with Tables VB and VIC.

[60]See the second part of Appendix A for a detailed discussion of the relation of this result to the corresponding generator presented by Dynkin[43] and the specific analysis of the error in Ref. 43.

[61]For an alternative, less symmetric solution [recalling that $B(2)\sim C(2)$], see Ref. 50. (Note the typographical error: $3\to\sqrt{3}$ for e_y.) The generators of the simple roots, in our notation, are

$$e_{2\mu_2} = (1/\sqrt{12})\big[2E^8_4 + E_{48} + \sqrt{3}(E^{346} + E_{358}) + E^2_1\big],$$

$$e^2_1 = (1/\sqrt{12})\big[E_{27} - E^8_2 - E^7_2 + E_{28} - E^4_1 + E^1_3 + \sqrt{3}(E_{567} + E^{257})\big].$$

[62]See Appendix B for detailed discussion and comparison with Ref. 37.

[63]In this case [and also for $E(7)\supset F(4)\oplus A(1)$, Table VG], we have projected λ_5 to zero, rather than λ_7 (see Sec. 4) in order to keep these $F(4)$ subgroups [as well as the $E(6)\supset F(4)$, Table IVD] essentially identical.

A unified description of the representations of the graded Lie algebra Gsl(2)

Bruno Gruber and T. S. Santhanam[a)]
Physics Department, Southern Illinois University, Carbondale, Illinois 62901

Raj Wilson[b)]
Mathematics Department, Southern Illinois University, Carbondale, Illinois 62901

(Received 29 January 1983; accepted for publication 5 August 1983)

Indecomposable representations of the graded Lie algebra Gsl(2) are analyzed in detail. It is further shown that the study of the irreducible representations (finite- and infinite-dimensional) is intimately related to the study of these indecomposable representations.

PACS numbers: 02.20.Sv, 11.30.Pb

1. INTRODUCTION

The representation theory of Lie superalgebras has recently been studied extensively. The motivation behind these investigations is that they are mathematical generalizations of ordinary Lie algebras. Besides, supersymmetric theories which are viable candidates for the unified gauge theory of particles are based on these superalgebras.[1] A complete classification of simple Lie superalgebras has already been given.[2]

For the case of Gsl(2) the irreducible representations, both finite-dimensional and infinite-dimensional, have been discussed in the literature.[3] In this article we take a very general point of view and discuss representations of Gsl(2) on the space of its enveloping algebra by using methods successfully used earlier for Lie algebras.[4] Particular emphasis is given to the study of indecomposable (i.e., reducible but not completely reducible) representations of Gsl(2) (Ref. 5). The indecomposable representations of Gsl(2) are shown to contain the finite-dimensional irreducible representations (on quotient spaces) as well as some of the infinite-dimensional irreducible representations (on invariant subspaces). In addition, the known results regarding the class of irreducible representations of Gsl(2), finite- and infinite-dimensional, follow automatically as a consequence of the analysis of indecomposable representations.

In Sec. 2 we define the Lie superalgebra Gsl(2) and give explicitly the action of the generators of this algebra on the basis states. In Sec. 3 we obtain the standard extremal vectors which define the invariant subspaces of the indecomposable representations. In Sec. 4, all representations of Gsl(2), indecomposable and irreducible (finite- as well as infinite-dimensional) which are defined by these extremal sectors are obtained in an explicit form. For the finite-dimensional irreducible representations and some of the infinite-dimensional irreducible representations the standard canonical form is rederived.

2. LIE SUPERALGEBRA Gsl(2)

Let $(l_3,l_\pm)$ be a Weyl basis of sl(2), l_3 being the regular element and $l_\pm$ the corresponding nilpotent elements with the commutation relations[6]

$$[l_3,l_\pm] = \pm l_\pm\,, \quad [l_+,l_-] = l_3\,, \tag{2.1}$$

where the bracket [ ] stands for the commutator. The center of the universal enveloping algebra of sl(2) contains the Casimir element

$$c(Q) = \{l_+,l_-\} + l_3^2\,, \tag{2.2}$$

where the bracket { } denotes the anticommutator. This sl(2) Lie algebra can be graded to a Lie superalgebra Gsl(2) by using the spinor $(V_\pm)$ which satisfies the anticommutation relations

$$\begin{aligned} &\{V_+,V_+\} = \sqrt{2}l_+\,, \quad \{V_-,V_-\} = \sqrt{2}l_-\,, \\ &\{V_+,V_-\} = -l_3\,. \end{aligned} \tag{2.3a}$$

Combining Eqs. (2.1) and (2.3), one obtains the bilinear operations

$$\begin{aligned} &[l_\pm,V_\pm] = 0\,, \quad [l_\pm,V_\mp] = (1/\sqrt{2})V_\pm\,, \\ &[l_3,V_\pm] = \pm\tfrac{1}{2}V_\pm\,. \end{aligned} \tag{2.3b}$$

The center of the universal enveloping algebra of Gsl(2) contains the Casimir element

$$\begin{aligned} \mathrm{Gc}(Q) &= \{l_+,l_-\} + l_3^2 + \tfrac{1}{2}[V_+,V_-] \\ &= \mathrm{c}(Q) + \tfrac{1}{2}[V_+,V_-]\,. \end{aligned} \tag{2.4}$$

The elements $(l_3,l_\pm)$ and $(V_\pm)$ are, respectively, the even and odd elements[7] acting on the Z_2 graded linear vector space V, the latter being the direct sum of two subspaces V_0 and V_1; i.e., $V = V_0 + V_1$, $(l_3,l_\pm)\in V_0$, $V_\pm \in V_1$. A five-dimensional irreducible representation of Gsl(2) can be obtained from the bilinear mappings [Eqs. (2.1)–(2.4)] as follows:

[a)] Visiting member. Permanent address: MATSCIENCE, The Institute of Mathematical Sciences, Madras 600113 India.
[b)] New address: Physics Department, Drexel University, Philadelphia, Pennsylvania 19104.

 0022-2488/84/051253-09$02.50

$$l_3=\begin{bmatrix}\begin{pmatrix}\frac{1}{2}&0\\0&-\frac{1}{2}\end{pmatrix}&&0\\&&\\0&&\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix}\end{bmatrix},$$

$$l_+=\begin{bmatrix}\begin{pmatrix}0&1/\sqrt{2}\\0&0\end{pmatrix}&&\\&&\\0&&\begin{pmatrix}0&0&-1\\0&0&0\\0&1&0\end{pmatrix}\end{bmatrix},$$

$$l_-=\begin{bmatrix}\begin{pmatrix}0&0\\1/\sqrt{2}&0\end{pmatrix}&&0\\&&\\0&&\begin{pmatrix}0&0&0\\0&0&1\\-1&0&0\end{pmatrix}\end{bmatrix}, \tag{2.5}$$

$$V_+=\begin{bmatrix}0&&\begin{pmatrix}0&0-\frac{1}{2}\\0-1/\sqrt{2}&0\end{pmatrix}\\&&\\\begin{pmatrix}\sqrt{2}&0\\0&0\\0&-1\end{pmatrix}&&0\end{bmatrix},$$

$$V_-=\begin{bmatrix}0&&\begin{pmatrix}-1/\sqrt{2}&0&0\\0&0&\frac{1}{2}\end{pmatrix}\\&&\\\begin{pmatrix}0&0\\0&-\sqrt{2}\\-1&0\end{pmatrix}&&0\end{bmatrix}$$

The Casimir element, Eq. (2.4), in this case becomes $\mathrm{Gc}(Q)=\frac{3}{2}\mathbf{1}$ ($\mathbf{1}$ is a 5×5 identity matrix), satisfying the Schur–Kac lemma.[8]

A basis for the universal enveloping algebra Ω of Gsl(2) can be chosen as the following set of ordered elements[9,10] (the product is the ordered tensor product):

$$\begin{aligned}&\Omega\colon\ \{\mathbf{1},V_+^{\,i}\,V_-^{\,k}\,l_+^{\,r}\,l_-^{\,s}\,l_3^{\,t}\ ;\ i,k=0,1;r,s,t\in\mathbb{N}\,,\\&\qquad\mathbb{N}=\text{the set of nonnegative integers}\}\,,\\&\qquad V_+^{\,i}\,V_-^{\,k}\,l_+^{\,r}\,l_-^{\,s}\,l_3^{\,t}\equiv x(i,k;r,s,t)\,,\\&\qquad\mathbf{1}\equiv x(0,0,0,0,0)\,.\end{aligned} \tag{2.6}$$

By straightforward induction one can easily obtain, in Ω, the following relations for $n\in\mathbb{N}$

$$\begin{aligned}&V_-V_+^{2n+1}=-V_+^{2n+1}V_--V_+^{2n}\,l_3-\tfrac{1}{2}nV_+^{2n}\,,\\&V_-V_+^{2n}=V_+^{2n}\,V_--\tfrac{1}{2}nV_+^{2n-1}\,,\\&l_+V_-^{2n+1}=V_-^{2n+1}l_++(1/\sqrt{2})V_-^{2n}\,V_+\\&\qquad\qquad-(n/\sqrt{2})V_-^{2n-1}l_3+(n^2/2\sqrt{2})V_-^{2n-1}\,,\\&l_+V_-^{2n}=-V_-^{2n}\,l_+-(n/\sqrt{2})V_-^{n(n-1)}l_3\\&\qquad\qquad+[n(n-1)/2\sqrt{2}]\,V_-^{2n}\,,\\&l_3V_\pm^{n}=V_\pm^{n}\,l_3\pm\tfrac{1}{2}nV_\pm^{n}\,,\\&l_-l_+^{r}=l_+^{r}\,l_--rl_+^{r-1}l_3-\tfrac{1}{2}r(r-1)l_+^{r-1}\,,\\&l_-l_3^{t}=(l_3+1)^tl_-\,,\\&l_3l_\pm^{r}=l_\pm^{r}\,l_3\pm rl_\pm^{r}\,.\end{aligned} \tag{2.7}$$

The action of the elements of Gsl(2) on $x(i,k;r,s,t)$ can now be determined which yields the following *master representation*:

$$\begin{aligned}&\rho(l_3)x(i,k;r,s,t)=x(i,k;r,s,t+1)\\&\qquad\qquad+[(r-s)+\tfrac{1}{2}(i-k)]x(i,k;r,s,t)\,,\\&\rho(l_+)x(i,k;r,s,t)=x(i,k;r+1,s,t)\\&\qquad\qquad+(k/\sqrt{2})x(i+1,k-1;r,s,t)\\&\rho(l_-)x(i,k;r,s,t)=x(i,k;r,s+1,t)\\&\qquad\qquad-rx(i,k;r-1,s,t+1)\\&\qquad\qquad+r[s-\tfrac{1}{2}(r-1)]x(i,k;r-1,s,t)\\&\qquad\qquad+(i/\sqrt{2})x(i-1,k+1;r,s,t)\,,\\&\rho(V_+)x(i,k;r,s,t)=x(i+1,k;r,s,t)\,,\\&\rho(V_-)x(i,k;r,s,t)=(-1)^ix(i,k+1;r,s,t)\\&\qquad\qquad-ix(i-1,k;r,s,t+1)\\&\qquad\qquad-i(r-s-\tfrac{1}{2}k)x(i-1,k;r,s,t)\,.\end{aligned} \tag{2.8}$$

We find that, in Eq. (2.3), the square of each odd generator ($V_\pm$) is proportional to the even generator ($l_\pm$), i.e., $V_\pm^2=\pm(1/\sqrt{2})l_\pm$. Thus, we have to supplement Eq. (2.8) with the following equations:

$$\begin{aligned}&x(2,0;r,s,t)=(1/\sqrt{2})x(0,0;r+1,s,t)\,,\\&x(2,1;r,s,t)=(1/\sqrt{2})\{x(0,1;r+1,s,t)+(1/\sqrt{2})x(1,0;r,s,t)\}\,,\\&x(0,2;r,s,t)=-(1/\sqrt{2})\{x(0,0;r,s+1,t)\\&\qquad\qquad-rx(0,0;r-1,s,t+1)\\&\qquad\qquad+r[s-\tfrac{1}{2}(r-1)]x(0,0;r-1,s,t)\}\,,\\&x(1,2;r,s,t)=-(1/\sqrt{2})\{x(1,0;r,s+1,t)\\&\qquad\qquad-rx(1,0;r-1,s,t+1)\\&\qquad\qquad+r[s-\tfrac{1}{2}(r-1)]x(1,0;r-1,s,t)\}\,.\end{aligned} \tag{2.9}$$

Equations (2.8) and (2.9) define the representations of Gsl(2). It is important to realize that the Gsl(2) generators acting on $x(i,k;r,s,t)$ keep the (s,t) the same or increase. This allows us to discuss the representations on the invariant subspaces $V(S,T)$, where $V(S,T)$ is the linear span of the set of basis elements,

$$\begin{aligned}&V(S,T)\colon\ \{x(i,k;r,S+s,T+t)\ ;\ i,k=0,1,\\&\qquad r,s,t\in\mathbb{N}\,,\ S,T\in\mathbb{N}^+\}\,,\end{aligned}$$

and also the representations induced on the quotient spaces $\Omega/V(S,T)$. $\mathbb{N}^+$ denotes the set of positive integers.

The relation $\rho(l_3)\mathbf{1}=\Lambda\mathbf{1}$, $\Lambda\in\mathbb{C}$, generates a left ideal I_Λ. The representation (2.8) induces on the quotient space the representation

$$\rho(l_3)x(i,k;r,s) = [\Lambda + r - s + \tfrac{1}{2}(i-k)]x(i,k;r,s)\,,$$
$$\rho(l_+)x(i,k;r,s) = x(i,k;r+1,s) + (k/\sqrt{2})x(i+1,k-1;r,s)\,,$$
$$\rho(l_-)x(i,k;r,s) = x(i,k;r,s+1) - r[\Lambda - s + \tfrac{1}{2}(r-1)] \times x(i,k;r-1,s) + (i/\sqrt{2})x(i-1,k+1;r,s)\,, \quad (2.10)$$
$$\rho(V_+)x(i,k;r,s) = x(i+1,k;r,s)\,,$$
$$\rho(V_-)x(i,k;r,s) = (-1)^i x(i,k+1;r,s) - i\left[\Lambda + r - s - \frac{k}{2}\right]x(i-1,k;r,s)\,,$$

and

$$x(2,0;r,s) = (1/\sqrt{2})x(0,0;r+1,s)\,,$$
$$x(2,1;r,s) = (1/\sqrt{2})\{x(0,1;r+1,s) + (1/\sqrt{2})x(1,0;r,s)\}\,, \quad (2.11)$$
$$x(0,2;r,s) = -(1/\sqrt{2})\{x(0,0;r,s+1) - r[\Lambda - s + \tfrac{1}{2}(r-1)]x(0,0;r-1,s)\}\,,$$
$$x(1,2;r,s) = -(1/\sqrt{2})\{x(1,0;r,s+1) - r[\Lambda - s + \tfrac{1}{2}(r-1)]x(1,0;r-1,s)\}\,.$$

This representation has $\rho(l_3)$ diagonal. It is indecomposable and has invariant subspaces with bases $V(s, T=0)$, $S\in\mathbb{N}^+$.

3. REPRESENTATIONS INDUCED ON QUOTIENT SPACES AND SUBDUCED ON INVARIANT SUBSPACES

A basis for the quotient spaces $\Omega/V(S,T)$ can be chosen as

$$\{V_+^i V_-^k l_+^r l_-^s l_3^t \equiv x(i,k;r,s,t)\,,\ i,k=0,1,\ r\in\mathbb{N}\,,\ s=0,1,\ldots,S-1,\ t=0,1,\ldots,T-1\}\,. \quad (3.1)$$

The representations which are induced on these quotient spaces are obtained from Eqs. (2.8) and (2.9) by formally setting $x(i,k;r,s+S,t+T)$ equal to zero for $S,T\in\mathbb{N}^+$.

Another type of representations is defined by the relation

$$\rho(l_3)\mathbb{1} = \Lambda\mathbb{1}\,, \quad \Lambda\in\mathbb{C}\,,$$

where

$$x(0,0;0,0,0) = \mathbb{1}\,, \quad (3.2)$$

i.e., the relation $(l_3 - \Lambda)\mathbb{1} = 0$ generates a left ideal on Ω. We consider the following representations which are induced by the master representation on the quotient space $\Omega/V(S=1)$ which satisfy Eq. (3.2). We obtain

$$\rho(l_3)x(i,k;r) = [\Lambda + r + \tfrac{1}{2}(i-k)]x(i,k;r)\,,$$
$$\rho(l_+)x(i,k;r) = x(i,k;r+1) + (k/\sqrt{2})x(i+1,k-1;r)$$
$$\rho(l_-)x(i,k;r) = -r[\Lambda + \tfrac{1}{2}(r-1)]x(i,k;r-1) + (i/\sqrt{2})x(i-1,k+1;r)\,,$$
$$\rho(V_+)x(i,k;r) = x(i+1,k;r)\,,$$
$$\rho(V_-)x(i,k;r) = (-1)^i x(i,k+1;r) - i(\Lambda + r - \tfrac{1}{2}k)x(i-1,k;r) \quad (3.3)$$

with

$$x(2,0;r) = x(0,0;r+1)\,,$$
$$x(2,1;r) = (1/\sqrt{2})x(0,1;r+1) + \tfrac{1}{2}x(1,0;r)\,,$$
$$x(0,2;r) = (1/\sqrt{2})r[\Lambda + \tfrac{1}{2}(r-1)]x(0,0;r-1)\,,$$
$$x(1,2;r) = (1/\sqrt{2})r[\Lambda + \tfrac{1}{2}(r-1)]x(1,0;r-1)\,. \quad (3.3')$$

The representations given by Eqs. (3.3) and (3.3') are not necessarily irreducible, but can be indecomposable. In the latter case, they have nontrivial invariant subspaces. These invariant subspaces serve as carrier spaces for the infinite-dimensional subrepresentations subduced by the action of different ρ's while ρ induces finite-dimensional representations on the quotient spaces. The extremal vectors Y which characterize the invariant subspaces satisfy the conditions

$$\rho(V_-)Y = 0\,, \quad \rho(l_-)Y = 0\,. \quad (3.4)$$

An inspection of Eqs. (3.3) and (3.3') reveals that the extremal vectors Y satisfying Eq. (3.4) will exist only when $r=0$ or $\Lambda + \frac{1}{2}(r-1) = 0$, i.e., for $r=0$ or $r = -2\Lambda + 1$. For $r=0$, two extremal vectors

$$y_1 = (1 + [1/(\Lambda - \tfrac{1}{2})]V_+V_-)\,, \quad \Lambda \neq \tfrac{1}{2}$$

and

$$\bar{y}_1 = V_- \quad (3.5)$$

are obtained which correspond to the weights Λ and $\Lambda - \frac{1}{2}$, respectively. If $\Lambda = -\frac{1}{2}(r-1)$, $r\in\mathbb{N}^+$, we get two additional extremal vectors y_2 and $\bar{y}_2$ given by

$$y_2 = [1 - [1/(\Lambda - \tfrac{1}{2})]V_+V_-]l_+^{-2\Lambda+1}\,,$$
$$\bar{y}_2 = V_-l_+^{-2\Lambda+1}\,, \quad (3.6)$$

which correspond to the weights $(-\Lambda + 1)$ and $(-\Lambda + \frac{1}{2})$, respectively. Thus, if $\Lambda \neq -\frac{1}{2}(n-1)$, $n\in\mathbb{N}^+$, the representation (3.3) decomposes into a direct sum of two infinite-dimensional *irreducible* representations which are defined by the extremal vectors y_1, and $\bar{y}_1$. If, however, $\Lambda = -\frac{1}{2}(n-1)$, $n\in\mathbb{N}^+$, then the representation equation (3.3) decomposes into the direct sum of two *indecomposable* representations which are infinite-dimensional. One of these indecomposable representations is defined by the extremal vector y_1 and has a *second* extremal vector $\bar{y}_2$. The other indecomposable representation is defined by the extremal vector $\bar{y}_1$ and has a second extremal vector y_2. Each of these two indecomposable representations induces on the *quotient space* with respect to the invariant subspace defined by the second extremal vector in each of these cases a *finite-dimensional* irreducible representation. In the former case the representation (defined by y_1 with the second extremal vector $\bar{y}_2$) is of dimension $(-4\Lambda + 1)$. In the latter case, the representation (defined by $\bar{y}_1$ with y_2 as the second extremal vector) is of dimension $(-4\Lambda + 3)$. On the invariant subspaces infinite-dimensional irreducible representations are subduced with lowest weight $(-\Lambda + 1)$ and extremal vector $\bar{y}_2$ in the former case (for the indecomposable representation defined by y_1) and with lowest weight $(-\Lambda + \frac{1}{2})$ and extremal vector y_2 in the latter case (for the indecomposable representation defined by $\bar{y}_1$). The indecomposable representation defined by y_1 goes over to the indecomposable representation defined by $\bar{y}_1$ when one makes the substitution $\Lambda \to \Lambda - \frac{1}{2}$. The same result holds for the irreducible finite-dimensional representations induced on the quotient spaces and also for the infi-

nite-dimensional irreducible representations subduced on the invariant subspaces. Therefore, it is sufficient to analyze only one of the two indecomposable representations. In the special case when $\Lambda = \frac{1}{2}$, there exist only two extremal vectors

$$y = V_+V_-\,, \quad \bar{y} = V_-\,, \tag{3.7}$$

with weights $\Lambda = \frac{1}{2}$ and $\Lambda = 0$, respectively. These two extremal vectors result in two infinite-dimensional *irreducible* representations. Equation (3.3) reduces to the sum of two infinite-dimensional irreducible representations. Again the substitution $\Lambda \to \Lambda - \frac{1}{2}$ takes the one defined by y to that defined by $\bar{y}$.

We study below, for purposes of illustration, the nontrivial examples for which $n = 1$ and 2, i.e., the cases for $\Lambda = 0$ and $\Lambda = -\frac{1}{2}$. We shall discuss the two representations which are induced by Eq. (3.3) on the quotient spaces, modulo invariant subspaces.

Case 1: $\Lambda = 0$

The set of elements V_1: $\{x(i,k;1+r),\ i,k = 0,1,\ r \in \mathbb{N}\}$ defines a basis for an invariant subspace of the representation given by Eq. (3.3) for the case $\Lambda = 0$. For the quotient space of $\Omega/V(1,1)$ with respect to the invariant subspace V_1, a basis can be chosen as

$$x(0,0;0)\,, \quad x(1,0;0)\,, \quad x(0,1;0)\,, \quad x(1,1;0)\,. \tag{3.8}$$

The representation equation (3.3) induces on the quotient space a four-dimensional *completely reducible* representation which is given by the matrices

$$\rho(l_3) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\rho(l_+) = \begin{bmatrix} 0 & 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\rho(l_-) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{3.9}$$

$$\rho(V_+) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\rho(V_-) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \frac{1}{2} \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

The Casimir element of Gsl(2) from Eqs. (2.5) and (3.9) is given by the singular matrix

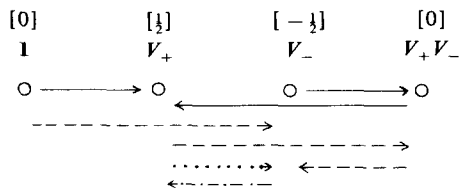


FIG. 1. Graphical representation of the four-dimensional completely reducible representation induced on the quotient space for $\Lambda = 0$.[11]

$$\mathrm{Gc}(2) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 1 & 0 & 0 & \frac{1}{2} \end{bmatrix}. \tag{3.10}$$

There is a degeneracy of the elements $\mathbb{1}$ and V_+V_- and the four-dimensional representation can be broken up into its irreducible constituents. A singlet representation is obtained which consists of the only vector $y_1 = 1 - V_+V_-$, with weight $\Lambda = 0$. That this is extremal can be seen by inspection. The triplet consists of the elements $\{V_+,\ 1 + \frac{1}{2}V_+V_-, V_-\}$ generated by the extremal vector $\bar{y}_1 = V_-$, with weight $\Lambda = -\frac{1}{2}$, as we shall see in the later sections. See Fig. 1.

Case 2: $\Lambda = -\frac{1}{2}$

Here, V_2: $\{x(i,k;2+r),\ i,k = 0,1,\ r \in \mathbb{N}\}$ forms a basis for an invariant subspace of $\Omega/V(1,1)$ with respect to the action of Eq. (3.3). The quotient space $[\Omega/V(1,1)]/V_2$ is eight-dimensional, and a basis can be chosen to consist of the elements

$$X(i,k;0)\,, \quad X(i,k;1)\,, \quad i,k = 0,1\,. \tag{3.11}$$

The representation (3.3) then induces on the quotient space modulo the invariant subspace the following *completely reducible* eight-dimensional representation [with respect to the ordering of basis elements: $x(0,0;0) = \mathbf{1}$, $x(1,0;0) = V_+$, $x(0,1;0) = V_-$, $x(1,1;0) = V_+V_-$, $x(0,0;1) = l_+$, $x(1,0;1) = V_+l_+$, $x(0,1;1) = V_-l_+$, $x(1,1;1) = V_+V_-l_+$ (see Fig. 2)].

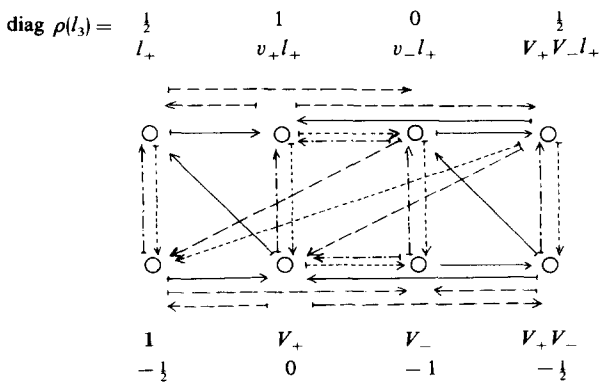


FIG. 2. The indecomposable representation $\Lambda = -\frac{1}{2}$ on the quotient space, yielding a decomposable representation of dimension $8 = 5 + 3$.

$$\rho(l_3)=\left[\begin{array}{cccc|cccc}
-\frac{1}{2} & 0 & 0 & 0 & & & & \\
0 & 0 & 0 & 0 & & & & \\
0 & 0 & -1 & 0 & & 0 & & \\
0 & 0 & 0 & -\frac{1}{2} & & & & \\
\hline
 & & & & \frac{1}{2} & 0 & 0 & 0 \\
 & & & & 0 & 1 & 0 & 0 \\
 & & 0 & & 0 & 0 & 0 & 0 \\
 & & & & 0 & 0 & 0 & \frac{1}{2}
\end{array}\right],$$

$$\rho(l_+)=\left[\begin{array}{cccc|cccc}
0 & 0 & 0 & 0 & & & & \\
0 & 0 & 1/\sqrt{2} & 0 & & 0 & & \\
0 & 0 & 0 & 0 & & & & \\
0 & 0 & 0 & 0 & & & & \\
\hline
1 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\rho(l_-)=\left[\begin{array}{cccc|cccc}
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{4} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\
0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\
\hline
 & & & & 0 & 0 & 0 & 0 \\
 & & & & 0 & 0 & 0 & 0 \\
 & & 0 & & 0 & 1/\sqrt{2} & 0 & 0 \\
 & & & & 0 & 0 & 0 & 0
\end{array}\right],$$

$$\rho(V_+)=\left[\begin{array}{cccc|cccc}
0 & 0 & 0 & 0 & & & & \\
1 & 0 & 0 & \frac{1}{2} & & 0 & & \\
0 & 0 & 0 & 0 & & & & \\
0 & 0 & 1 & 0 & & & & \\
\hline
0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{2} \\
0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}\right],$$

$$\rho(V_-)=\left[\begin{array}{cccc|cccc}
0 & \frac{1}{2} & 0 & 0 & 0 & 0 & -1/2\sqrt{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2\sqrt{2} \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
 & & & & 0 & -\frac{1}{2} & 0 & 0 \\
 & & & & 0 & 0 & 0 & 0 \\
 & & 0 & & & & & \\
 & & & & 1 & 0 & 0 & 0 \\
 & & & & 0 & -1 & 0 & 0
\end{array}\right]. \tag{3.12}$$

The Casimir element of Gsl(2) given in Eq. (2.5) and (3.12) is a nonsingular matrix

$$\mathrm{Gc}(Q)=\left[\begin{array}{cccc|cccc}
\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1/2\sqrt{2} & 0 \\
0 & 0 & \frac{3}{2} & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & \frac{3}{2} & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{4} \\
0 & 0 & 0 & 0 & 0 & \frac{3}{2} & 0 & 0 \\
0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1
\end{array}\right].$$

We can see from Fig. 2 that there is degeneracy of the elements $(1, V_+V_-)$ with $-\frac{1}{2}$ eigenvalue for $\rho(l_3)$, $(l_+, V_+V_-l_+)$ with eigenvalue $\frac{1}{2}$ for $\rho(l_3)$, and (V_+, V_-l_+) with eigenvalue 0 for $\rho(l_3)$. This eight-dimensional completely reducible representation (3.12) can be brought into the block form of three and five dimensions. To do this, we first construct the extremal vectors y_1 and $\bar{y}_1$ for the three- and five-dimensional invariant subspaces. The remaining elements in each case can then be determined by the action of the ρ's on the respective extremal vectors.

In the following we carry out the analysis for the case of a general value of Λ and later we shall specialize to the case when $\Lambda = -\frac{1}{2}$. We define $M = r + \Lambda + \frac{1}{2}(i-k)$ so that M represents, from Eq. (3.3), the eigenvalue of $x(i,k;r)$ for $\rho(l_3)$. Consider the subspaces

$$V_M: \quad \{x(i,k;r),\ r = M - \Lambda + \tfrac{1}{2}(i-k),\ r\in\mathbb{N};\ i,k = 0,1\}. \tag{3.13}$$

The elements of V_M belong to the eigenspace of $\rho(l_3)$ with eigenvalue $(M-\Lambda)$. The two vectors

$$Y = (C_{00} + C_{11}V_+V_-)l_+^{M-\Lambda},\quad (M-\Lambda)\in\mathbb{N},$$
$$\text{and} \tag{3.14}$$
$$\bar{Y} = (C_{01}V_-l_+ + C_{10}V_+)l_+^{M-\Lambda-1/2},\quad M-\Lambda-\tfrac{1}{2}\in\mathbb{N},$$

are elements of V_M. We want to find the conditions for which Y and $\bar{Y}$ become extremal vectors, i.e., $\rho(l_-)Y = 0$, $\rho(V_-)Y = 0$; $\rho(l_-)\bar{Y} = 0$, $(V_-)\bar{Y} = 0$. It turns out that there are four possibilities. Viz.,

$$y_1 = (1 + [1/(\Lambda - \tfrac{1}{2})]V_+V_-),\quad \Lambda \neq \tfrac{1}{2},\quad \text{with weight } \Lambda,$$
$$\bar{y}_1 = V_- \quad \text{with weight } (\Lambda - \tfrac{1}{2}),$$
$$y_2 = (1 - [1/(\Lambda - \tfrac{1}{2})]V_+V_-)l_+^{-2\Lambda+1},$$
$$\Lambda \neq \tfrac{1}{2} \quad \text{with weight } \Lambda + r,$$
$$\bar{y}_2 = V_-l_+^{-2\Lambda+1} \quad \text{with weight } \Lambda - \tfrac{1}{2} + r. \tag{3.15}$$

The extremal vectors y_1 and y_2 exist only when $\Lambda \neq \frac{1}{2}$. The extremal vectors y_2, and $\bar{y}_2$ exist only if $-2\Lambda + 1 \in \mathbb{N}^+$. For the exceptional case of $\Lambda = \frac{1}{2}$, there exist only two extremal vectors, viz.,

$$y = V_+V_-,\quad \bar{y} = V_-.$$

The considerations given above show that if $\Lambda \neq 0, -\frac{1}{2}, -\frac{3}{2}, \ldots$, then the representations given in Eq. (3.3) decompose into a direct sum of two irreducible, infinite-dimensional representations. If $\Lambda = 0, -\frac{1}{2}, -\frac{3}{2}, \ldots$, then the representations given in Eq. (3.3) decompose into the direct sum of two infinite-dimensional, indecomposable representations. One of these has extremal vectors $(y_1, \bar{y}_2)$; the other has $(\bar{y}_1, y_2)$ as

extremal. The extremal vectors $\bar{y}_2$ and y_2 define infinite-dimensional invariant subspaces of the indecomposable representations.

Now, returning to our example of the special case for which $\Lambda = -\frac{1}{2}$, we observe that the action of ρ's on the extremal vector y_1 in Eq. (3.15) yields a three-dimensional irreducible representation on the quotient space $\Omega/V(r=2)$, $\Lambda = -\frac{1}{2}$. The basis for this three-dimensional irreducible representation is given by

$$\{1 - V_+V_-,\ \tfrac{1}{2}l_+ - V_+V_-l_+,\ \tfrac{1}{2}V_+ - V_-l_+\}$$

and the representation is obtained in explicit form as

$$\rho(l_3) = \begin{pmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix},\quad \rho(l_+) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\rho(l_-) = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\rho(V_+) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix},$$
$$\rho(V_-) = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & -1/\sqrt{2} & 0 \end{pmatrix},$$
$$\mathrm{Gc}(Q) = \tfrac{1}{2}. \tag{3.16}$$

Similarly, the action of the ρ's on the extremal vector $\bar{y}_1$ in Eq. (3.15) results in the five-dimensional irreducible representation with the basis $\{V_-,\ V_-l_+ + (1/\sqrt{2})V_+,\ V_+l_+, V_+V_-, V_+V_-l_+ + \frac{1}{2}l_+\}$ and the representation is obtained as

$$\rho(l_3) = \begin{bmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \end{bmatrix},$$
$$\rho(l_+) \begin{bmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \end{bmatrix},$$
$$\rho(l_-) = \begin{bmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} 0 & \frac{1}{2} \\ 0 & 0 \end{pmatrix} \end{bmatrix},$$
$$\rho(V_+) = \begin{bmatrix} 0 & \begin{pmatrix} 0 & 0 \\ 1/\sqrt{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} & 0 \end{bmatrix},$$
$$\rho(V_-) = \begin{bmatrix} 0 & \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & -1/\sqrt{2} & 0 \\ 0 & 0 & -1 \end{pmatrix} & 0 \end{bmatrix},$$
$$\mathrm{Glc}(Q) = \tfrac{3}{2}\mathbb{1}. \tag{3.17}$$

The representations given in Eqs. (3.17) and (2.6) are algebraically equivalent through a similarity transformation.

If one orders the basis elements in the sequence

$$\{V_-,\ V_-l_+ + (1/\sqrt{2})V_+,\ V_+l_+,\ V_+V_-,$$
$$V_+V_-l_+ + \tfrac{1}{2}l_+,\ \mathbb{1} - V_+V_-,\ \tfrac{1}{2}l_+ - V_+V_-l_+,$$
$$-(1/\sqrt{2})V_-l_+ + \tfrac{1}{2}V_+\},$$

then one obtains the representation (3.12) in the form of a direct sum of a three-dimensional and a five-dimensional irreducible representation. Equations (3.16) and (3.17) represent the decomposition of the eight-dimensional (reducible) representation (3.12) into a direct sum of a three-dimensional and five-dimensional irreducible representation. The representations (3.17) and (2.5) are equivalent.

4. CANONICAL AND INDECOMPOSABLE REPRESENTATIONS

In this section we explicitly construct and classify the representations which can be derived from Eq. (3.3). We obtain in an explicit form the matrix elements for the irreducible infinite-dimensional representations which are defined by the extremal vectors $y_1, y_2, \bar{y}_1$, and $\bar{y}_2$. Besides, we also obtain the matrix elements for the indecomposable representations which are defined by the extremal vectors y_1 and y_2 for $\Lambda = 0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$. Finally we derive the matrix elements for the finite-dimensional representations which are defined by the pairs of extremal vectors $(y_1, \bar{y}_2)$ and $(y_2, \bar{y}_1)$, respectively, for $\Lambda = 0, -\frac{1}{2}, -1, -\frac{3}{2}, \ldots$.

A. Representations defined by y_1

The extremal vector y_1 has been defined by

$$y_1 = (1 + \alpha V_+V_-),\quad \alpha \equiv 1/(\Lambda - \tfrac{1}{2}),\quad \Lambda \neq \tfrac{1}{2}.$$

Use of Eq. (3.3) yields the following:

$$\rho(l_+)^n y_1 = [(1 + n\alpha/2) + \alpha V_+V_-]l_+^n,$$
$$\rho(V_+)^{2K+1}\rho(l_+)^n y_1 = (1 + (n+K+1)\alpha/2)(1/\sqrt{2})^K V_+ l_+^{n+K} + \alpha(1/\sqrt{2})^{K+1} V_- l_+^{n+K+1},$$
$$= (1/\sqrt{2})\,\rho(V_+)^{2K-1}\rho(l_+)^{n+1} y_1,$$
$$\rho(V_+)^{2K}\rho(l_+)^n y_1 = (1/\sqrt{2})^K \rho(l_+)^{n+K} y_1 = (1 + (n+K)\alpha/2) \times (1/\sqrt{2})^K l_+^{n+K} + \alpha(1/\sqrt{2})^K V_+V_- l_+^{n+K},$$
$$\rho(l_-)\rho(l_+)^n y_1 = -n[\Lambda + \tfrac{1}{2}(n-1)]\,\rho(l_+)^{n-1} y_1,$$
$$\rho(l_-)\rho(V_+)^{2K+1}\rho(l_+)^n y_1 = -(n+K)[\Lambda + \tfrac{1}{2}(n+K)]\,\rho(V_+)^{2K+1}\rho(l_+)^{n-1} y_1,$$

$$\rho(l_-)\rho(V_+)^{2K}\rho(l_+)^n y_1$$
$$= -(1/\sqrt{2})^K (n+K) \times [\Lambda + \tfrac{1}{2}(n+K-1)]\,\rho(l_+)^{n+K-1} y_1 ,$$
$$\rho(V_-)\rho(l_+)^n y_1 = -(n\alpha/2) V_- l_+^n - (n/\sqrt{2})(1+n\alpha/2) V_+ l_+^{n-1} = -(n/\sqrt{2})\,\rho(V_+)\rho(l_+)^{n-1} y_1 ,$$
$$\rho(V_-)\rho(V_+)^{2K+1}\rho(l_+)^n y_1 = -[\Lambda + \tfrac{1}{2}(n+K)]\,\rho(V_+)^{2K}\rho(l_+)^n y_1 = -[\Lambda + \tfrac{1}{2}(n+K)](1/\sqrt{2})^K \rho(l_+)^{n+K} y_1 ,$$
$$\rho(V_-)\rho(V_+)^{2K}\rho(l_+)^n y_1 = -(n+K)\tfrac{1}{2}\rho(V_+)^{2K-1}\rho(l_+)^n y_1 = -(n+K)(1/\sqrt{2})\,\rho(V_+)^{2K+1}\rho(l_+)^{n-1} y_1 . \tag{4.1}$$

We now define

$$\rho(V_+)^{2K+1}\rho(l_+)^n y_1 \equiv |2K+1,n\rangle = (1/\sqrt{2})^K |1,n+K\rangle ,$$
$$\rho(V_+)^{2K}\rho(l_+)^n y_1 \equiv |2K,n\rangle = (1/\sqrt{2})^K |0,n+K\rangle . \tag{4.2}$$

For $K=0$, we have the basis

$$|0,n\rangle , \ |1,n\rangle , \quad n\in\mathbb{N} .$$

The action of the ρ's on this basis is given by Eq. (4.1):

$$\rho(l_3)|0,n\rangle = (\Lambda+n)|0,n\rangle ,$$
$$\rho(l_3)|1,n\rangle = (\Lambda+\tfrac{1}{2}+n)|1,n\rangle ,$$
$$\rho(l_+)|0,n\rangle = |0,n+1\rangle ; \rho(l_-)|0,n\rangle = -n[\Lambda + \tfrac{1}{2}(n-1)]|0,n-1\rangle ,$$
$$\rho(V_+)|0,n\rangle = |1,n\rangle , \rho(V_-)|0,n\rangle = -(n/\sqrt{2})|1,n-1\rangle$$
$$\rho(V_+)|1,n\rangle = (1/\sqrt{2})|0,n+1\rangle = |2,n\rangle , \quad \rho(V_-)|1,n\rangle = -(\Lambda + \tfrac{1}{2}n)|0,n\rangle . \tag{4.3}$$

Thus, we find that if $\Lambda \neq -n/2$, $n\in\mathbb{N}$, the representation is infinite-dimensional and irreducible. If, however, $\Lambda = -n/2$, $n\in\mathbb{N}$, then the extremal vector $|0,0\rangle$ defines an infinite-dimensional indecomposable representation with a second extremal vector $|1,-2\Lambda\rangle$. This indecomposable representation induces on the quotient space a $(-4\Lambda+1)$-dimensional irreducible representation with basis $\{|0,s\rangle, s=0,1,2,\ldots,-2\Lambda; |1,s\rangle, s=0,1,\ldots,-2\Lambda-1\}$. Moreover, it subduces on the invariant subspace defined by the extremal vector $|1,-2\Lambda\rangle$, an infinite-dimensional irreducible representation with basis:

$$\{|0,s\rangle , \ s = -2\Lambda+1, -2\Lambda+2,\ldots; \quad |1,s\rangle , \ s = -2\Lambda, -2\Lambda+1,\ldots\} .$$

The finite-dimensional irreducible representation which is induced is obtained from Eq. (4.3) by formally setting $|0,-2\Lambda+1\rangle \to 0$, $|1,-2\Lambda\rangle \to 0$. The infinite-dimensional irreducible representation on the invariant subspace is obtained by restricting Eq. (4.3) to the subspace (see Fig. 3).

The connection between the representations given in Eq. (4.3) and the usual finite-dimensional canonical representations on the quotient space modulo the invariant subspace defined by $\bar{y}_2$ can easily be seen by redefining $l = n/2$, $n\geqslant 0$, integer $-\Lambda = l$, $n = l+m$ and $|0,n\rangle = |0,l,m\rangle$, $|0,0\rangle = |0,l,-l\rangle$, $|1,n\rangle = |1,l,m\rangle$, $|1,0\rangle = |1,l,-l\rangle$. One obtains for the canonical basis of the $(4l+1)$-dimensional representation

$$\|0,l,m\rangle \equiv a_{lm}|0,l,m\rangle , \quad m = l, l-1,\ldots,-l+1 ,$$
$$\|1,l,m\rangle = a_{lm}[1/(l-m)]|1,l,m\rangle , \quad m = l-1,\ldots,-l+1 ,$$
$$\|0,l,-l\rangle = |0,l,-l\rangle ,$$
$$\|1,l,-l\rangle = |1,l,-l\rangle , \tag{4.4}$$

where

$$a_{lm} = \left\{ \prod_{K=m}^{-l+1} \frac{1}{2}[l(l+1) - K(K-1)] \right\}^{-1/2} .$$

From Eqs. (4.3) and (4.4), it follows that

$$\rho(l_\pm)\|0,l,m\rangle = \sqrt{\tfrac{1}{2}(l\mp m)(l\pm m+1)}\,\|0,l,m\pm 1\rangle ,$$
$$\rho(V_-)\|0,l,m\rangle = -\sqrt{(l+m)(l-m+1)}\,\|1,l,m-1\rangle ,$$
$$\rho(V_+)\|0,l,m\rangle = (l-m)\|1,l,m\rangle ,$$
$$\rho(l_-)\|1,l,m\rangle = \sqrt{\tfrac{1}{2}(l+m)(l-m+1)}\,\|1,l,m-1\rangle ,$$
$$\rho(l_+)\|1,l,m\rangle = \sqrt{\tfrac{1}{2}(l-m)(l+m+1)} \times [(l-m-1)/(l-m)]\,\|1,l,m+1\rangle ,$$
$$\rho(V_-)\|1,l,m\rangle = \tfrac{1}{2}\|0,l,m\rangle ,$$
$$\rho(V_+)\|1,l,m\rangle = \tfrac{1}{2}\sqrt{(l+m+1)/(l-m)}\,\|0,l,m+1\rangle . \tag{4.5}$$

B. Representations defined by $\bar{y}_1$

The extremal vector $\bar{y}_1 = V_-$. If we now define as we did earlier for y_1,

$$\rho(V_+)^{2K+1}\rho(l_+)^n \bar{y}_1 \equiv |2K+1,n\rangle = (1/\sqrt{2})^K |1,n+K\rangle ,$$
$$\rho(V_+)^{2K}\rho(l_+)^n \bar{y}_1 \equiv |2K,n\rangle = (1/\sqrt{2})^K |0,n+K\rangle , \quad n\in\mathbb{N} , \tag{4.6}$$

it can be seen using Eq. (3.3) that

$$\rho(l_3)|0,n\rangle = (\Lambda+n-\tfrac{1}{2})|0,n\rangle ,$$
$$\rho(l_3)|1,n\rangle = (\Lambda+n)|1,n\rangle ,$$

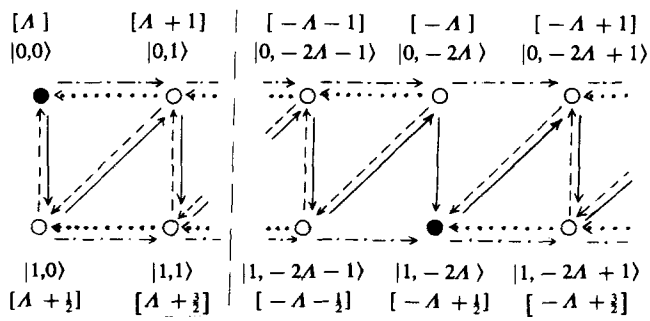


FIG. 3. Finite- and infinite-dimensional (bounded below) representation induced by the two extremal vectors $|0,0\rangle$ and $|1,-2\Lambda\rangle$.[11]

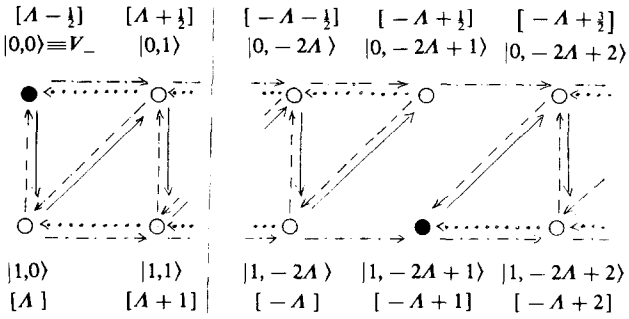


FIG. 4. Finite- and infinite-dimensional (bounded below) representation induced by the two extremal vectors $|0,0\rangle$ and $|1,-2\Lambda+1\rangle$.[11]

$$
\begin{aligned}
&\rho(l_+)|0,n\rangle = |0,n+1\rangle\,,\\
&\rho(l_-)|0,n\rangle = -n(\Lambda+\tfrac{1}{2}(n-2))|0,n-1\rangle\,,\\
&\rho(l_+)|1,n\rangle = |1,n+1\rangle\,,\\
&\rho(l_-)|1,n\rangle = -n[\Lambda+\tfrac{1}{2}(n-1)]|1,n-1\rangle\,,\\
&\rho(V_+)|0,n\rangle = |1,n\rangle\,,\\
&\rho(V_-)|0,n\rangle = -(n/\sqrt{2})|1,n-1\rangle\,,\\
&\rho(V_+)|1,n\rangle = (1/\sqrt{2})|0,n+1\rangle\,,\\
&\rho(V_-)|1,n\rangle = -[\Lambda+\tfrac{1}{2}(n-1)]|0,n\rangle\,.
\end{aligned}
\tag{4.7}
$$

It is easily observed that a replacement $\Lambda\to\Lambda+\frac{1}{2}$ in Eq. (4.7) yields the representation given by Eq. (4.3). Thus, this case gives nothing new (see Fig. 4).

C. Canonical basis for infinite-dimensional irreducible representations subduced on invariant subspaces

In this section, we determine the canonical basis for the infinite-dimensional irreducible representations. In the previous section, it has been shown that the representation defined by the extremal vector y_2 for $\Lambda\to\Lambda+\frac{1}{2}$ is algebraically equivalent to the representation defined by y_1 for the value Λ. Thus, for indecomposable representations (if $\Lambda=0,\ -\frac{1}{2},\ -1,\ -\frac{3}{2},\ldots$), the second extremal vectors which define the invariant subspaces in the two representations can be obtained by the replacement $\Lambda\to\Lambda+\frac{1}{2}$; they yield the same infinite-dimensional irreducible representation. Since it does not matter which of the two extremal vectors we use, we choose

$$y_2=(1+\beta V_+V_-)l^r{}_+\,,\quad \beta=-1/(\Lambda-\tfrac{1}{2})\,,\quad \Lambda\neq\tfrac{1}{2}\,.$$

If we define

$$
\begin{aligned}
|2K+1,n] &= (1/\sqrt{2})^K|1,n+K] \equiv \rho(V_+)^{2K+1}\rho(l_+)^n y_2\,,\\
|2K,n] &= (1/\sqrt{2})^K|0,n+K] \equiv \rho(V_+)^{2K}\rho(l_+)^n y_2\,.
\end{aligned}
\tag{4.8}
$$

the following relations may be worked out:

$$
\begin{aligned}
&\rho(l_3)|0,n] = (-\Lambda+1+n)|0,n]\,,\\
&\rho(l_3)|1,n] = (-\Lambda+\tfrac{3}{2}+n)|1,n]\,,\\
&\rho(l_+)|0,n] = |0,n+1]\,,\\
&\rho(l_-)|0,n] = n[\Lambda-\tfrac{1}{2}(n+1)]|0,n-1]\,,\\
&\rho(l_+)|1,n] = |1,n+1]\,,\\
&\rho(l_-)|1,n] = n[\Lambda-\tfrac{1}{2}(n+2)]|1,n-1]\,,\\
&\rho(V_+)|0,n] = |1,n]\,,\\
&\rho(V_-)|0,n] = -(1/\sqrt{2})n|1,n-1]\,,\\
&\rho(V_+)|1,n] = (1/\sqrt{2})|0,n+1]\,,\\
&\rho(V_-)|1,n] = [\Lambda-\tfrac{1}{2}(n+2)]|0,n]\,.
\end{aligned}
\tag{4.9}
$$

The relation to the standard canonical representation may be obtained by identifying $-\Lambda=l-1$, $n=(m-l)$, $n\in\mathbb{N}$, and by defining the canonical basis as

$$
\begin{aligned}
&\|0,l,l] = |0,0]\,,\\
&\|0,l,m] = c_{lm}|0,n]\,,\quad m=l+1,l+2,\ldots,\\
&\|1,l,l] = |1,0]\,,\\
&\|1,l,m] = c_{lm}[1/(l+m)]|1,n]\,,\quad m=l+1,l+2,\ldots\,,
\end{aligned}
$$

where

$$c_{lm}=\left\{\prod_{K=l+1}^{m}\frac{1}{2}[l(l-1)-K(K-1]\right\}^{-1/2}\,. \tag{4.10}$$

It can be easily verified that

$$
\begin{aligned}
&\rho(l_\mp)\|0,l,m] = \sqrt{\tfrac{1}{2}(l\mp m)(l\pm m-1)}\|0,l,m\pm1]\,,\\
&\rho(V_+)\|0,l,m] = (l+m)\|1,l,m]\,,\\
&\rho(V_-)\|0,l,m] = \sqrt{(l-m)(l+m+1)}\|1,l,m-1]\,,\\
&\rho(l_-)\|1,l,m] = \sqrt{\tfrac{1}{2}(l-m)(l+m-1)}\|1,l,m-1]\,,\\
&\rho(l_+)\|1,l,m] = \sqrt{\tfrac{1}{2}(l+m)(l-m+1)}\\
&\qquad\times[(l+m+1)/(l+m)]\|1,l,m+1]\,,\\
&\rho(V_-)\|1,l,m] = -\tfrac{1}{2}\|0,l,m]\,,\\
&\rho(V_+)\|1,l,m] = \tfrac{1}{2}\sqrt{(l-m-1)/(l+m)}\|0,l,m+1]\,.
\end{aligned}
\tag{4.11}
$$

Here we have characterized the infinite-dimensional irreducible representation which is bounded below by $(-l-1)$ by the weight l of the indecomposable representation from which it has been derived. It should be noted that the two weights $(-l-1)$ and l are related by the action of the discrete Weyl group W, $M=S(l+R)-R$, $S\in W$. Here $R=\frac{1}{2}$, denoting one-half of the sum of the positive roots of Su(2). For $S=1$, the identity, $M=l$, and for $S=-1$, which is the other element of W, we get $M=-l-2R=-l-1$.

D. Exceptional case $\Lambda=\frac{1}{2}$

The exceptional representation when $\Lambda=\frac{1}{2}$, $r=-2\Lambda+1=0$ is defined by the extremal vectors

$$
\begin{aligned}
&y'=V_+V_-=x(1,1;0)=\lim_{\Lambda\to1/2}(\Lambda-\tfrac{1}{2})y_1\,,\\
&y''=V_-=x(0,1;0)=\bar{y}_1\,.
\end{aligned}
$$

The extremal vector y' generates an infinite-dimensional irreducible representation which is bounded below.

In the above we have taken the correspondence from Eq. (4.2) by the replacement $|i,n\rangle\to(\Lambda-\frac{1}{2})|i,n\rangle$, $i=0,1$, $n\in\mathbb{N}$. Now, the extremal vector y'' can be used to generate new representations obtained from y' by the replacement $(\Lambda\leftrightarrow\Lambda+\frac{1}{2})$.

[1] I. Kaplansky, Pac. J. Math. **86**, 93 (1980) and references therein; see also V. Rittenberg, Rockefeller Report No. C00-2232B-141, 1978.
[2] F. A. Berezin and G. I. Kac, Mat. Sb. **82**, 124 (1970) [Math. U.S.S.R. Sb.

11, 311 (1970)]; J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974); A. Salam and J. Strathdee, Phys. Lett. B **49**, 465 (1974).
[3]J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[4]B. Gruber and A. U. Klimyk, J. Math. Phys. **16**, 1816 (1975). Indecomposable representations for parabose algebra (a prototype version of a graded Lie algebra) have been obtained in B. Gruber and T. S. Santhanam, J. Math. Phys. **24**, 1032 (1983). Representations of one-parameter families of states have been discussed in H. D. Doebner and B. Gruber, preprint.
[5]M. Marcu, J. Math. Phys. **21**, 1277 (1980).
[6]J. P. Serre, *Lie Algebras and Lie Groups* (Benjamin, New York, 1965).
[7]F. A. Berezin, Moscow Preprint #ITEP-66, 1977.
[8]V. G. Kac, Commun. Math. Phys. **53**, 31 (1977).
[9]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[10]J. Dixmier, *Algebres eneveloppantes* (Gauthier-Villars, Paris, 1974).
[11]Throughout, we denote the action of $\rho(V_+)$ by ————, the action of $\rho(V_-)$ by — — — — — —, the action of $\rho(l_-)$ by ·····, and the action of $\rho(l_+)$ by -·-·-·-.

On the Cauchy problem for the coupled Schrödinger–Klein–Gordon equations in one space dimension

Szymon Rabsztyn
Institute of Mathematics, Silesian Technical University, Gliwice, Poland

(Received 2 May 1983; accepted for publication 22 July 1983)

The existence, uniqueness, and continuity with respect to initial data of global solutions of the Cauchy problem is proved for the Schrödinger and Klein–Gordon equations with Yukawa coupling in one space dimension. The proof is based on the standard tools for handling abstract nonlinear wave equations.

PACS numbers: 02.30. + g, 03.50.Kk, 11.10.Lm

1. INTRODUCTION

In this paper we shall prove the existence, uniqueness, and continuity with respect to initial data of global solutions of the integral equation corresponding to the Cauchy problem for the Schrödinger and Klein–Gordon equations with Yukawa coupling in one space dimension:

$$i\frac{\partial}{\partial t}\psi(x,t)+\frac{\partial^2}{\partial x^2}\psi(x,t)=g\psi(x,t)\mathrm{Re}\,u(x,t),$$

$$\frac{\partial^2}{\partial x^2}u(x,t)-\frac{\partial^2}{\partial t^2}u(x,t)-m^2u(x,t)=g\psi(x,t)\overline{\psi(x,t)}, \tag{1}$$

$$\psi(x,t_0)=\psi_0(x),$$

$$u(x,t_0)=u_0(x),\quad \frac{\partial}{\partial t}u(x,t_0)=v_0(x),$$

where g, m are real constants and $\psi(x,t)$, $u(x,t)$ are complex-valued functions on $\mathbb{R}$.

This problem will be treated (locally in Sec. 3 and globally in Sec. 4) in the context of a Hilbert space $H^1(\mathbb{R})\oplus H^1(\mathbb{R})\oplus L^2(\mathbb{R})$ by now standard technique for handling abstract nonlinear wave equations (see Sec. 2). In this regard our discussion of Eqs. (1) is simpler and more general than that of Fukuda and Tsutsumi in Ref. 1, where the compactness arguments were used in order to prove the existence of weak solutions for (1).

We note finally that for the integral equation corresponding to the Cauchy problem for the Dirac and Klein–Gordon equations with Yukawa coupling in one space dimension an existence-uniqueness theorem has been proved by Chadam in Ref. 2.

2. PRELIMINARIES

In this section we first introduce some notations, definitions, and results on Sobolev spaces that will be used in the paper (for proofs see, e.g., Ref. 3). Then, for the reader's convenience, we shall outline those aspects in abstract nonlinear equations that are relevant for our purposes. This second part is based on the lectures by Reed[4] to which the reader can be referred for details and further references.

A. We denote by $H^r(\mathbb{R})$, for any nonnegative integer r, the Sobolev space of all complex-valued square-integrable functions f on $\mathbb{R}$ with finite norm $\|f\|_r=\{\int_{\mathbb{R}}(1+k^2)^r|\hat f(k)|^2dk\}^{1/2}$, where $\hat f$ is the Fourier transform of f, defined by

$$(Ff)(k)=\hat f(k)=\underset{N\to\infty}{\text{l.i.m.}}\,(2\pi)^{-1/2}\int_{|x|<N}e^{-ixk}f(x)dx,$$

where l.i.m. denotes the "limit in mean" taken with respect to the norm in $L^2(\mathbb{R})=H^0(\mathbb{R})$. Equivalently, $H^r(\mathbb{R})$ is the set of all complex-valued functions on $\mathbb{R}$ whose distributional derivatives of order smaller or equal to r lie in $L^2(\mathbb{R})$. Hence $H^r(\mathbb{R})$ is a Hilbert space with the scalar product $(f,g)_r=\int_{\mathbb{R}}(1+k^2)^r\overline{\hat f(k)}\hat g(k)dk$ and $H^r(\mathbb{R})\subset H^s(\mathbb{R})$ for $r\geqslant s$. Moreover for $r\geqslant 1$:

(a) $H^r(\mathbb{R})$ is an algebra with respect to pointwise multiplication and there exists a constant α_r such that $\|fg\|_r\leqslant\alpha_r\|f\|_r\|g\|_r$ for all $f,g\in H^r(\mathbb{R})$, $r\geqslant 1$.

(b) $\|f\|_\infty\leqslant\|f\|_r$ for all $f\in H^r(\mathbb{R})$, $r\geqslant 1$, where $\|\cdot\|_\infty$ denotes the norm in $L^\infty(\mathbb{R})$.

For any even nonnegative smooth function j on $\mathbb{R}$ such that $j(x)=0$ for $|x|\geqslant 1$, $\int_{\mathbb{R}}j(x)dx=1$, and any $\epsilon>0$, we denote by J_ϵ a smoothing operator defined by $(J_\epsilon f)(x)=\int_{\mathbb{R}}j_\epsilon(x-y)f(y)dy$ for any f from $H^r(\mathbb{R})$, where $j_\epsilon(x)=(1/\epsilon)j(x/\epsilon)$. The operator J_ϵ has the following properties:

(c) $J_\epsilon\colon H^r(\mathbb{R})\to\bigcap_{r=0}^{\infty}H^r(\mathbb{R})\subset C^\infty(\mathbb{R})$;

(d) $(J_\epsilon f,g)_r=(f,J_\epsilon g)_r$ for all $f,g\in H^r(\mathbb{R})$;

(e) $\|J_\epsilon f\|_r\leqslant\|f\|_r$ for all $f\in H^r(\mathbb{R})$;

(f) $\lim_{\epsilon\to 0}\|J_\epsilon f-f\|_r=0$ for all $f\in H^r(\mathbb{R})$;

(g) $\|J_\epsilon f\|_r\leqslant(2\pi)^{1/2}\|\hat j_\epsilon(1+k^2)^{s/2}\|_\infty\|f\|_{r-s}$ for all $f\in H^r(\mathbb{R})$ and $r\geqslant s$.

B. Let $\mathscr{H}$ be a Hilbert space with the norm $\|\Phi\|$, $\Phi\in\mathscr{H}$ and e^{-itA}, $t\in\mathbb{R}$ be a strongly continuous one-parameter group of unitary operators on $\mathscr{H}$ with the self-adjoint on $\mathscr{D}(A)\subset\mathscr{H}$ generator A. For a mapping K of $\mathscr{H}$ into itself with the property $K(0)=0$ let us list the following assumptions:

(i) $\|K(\Phi)-K(\Psi)\|\leqslant C(\|\Phi\|,\|\Psi\|)\|\Phi-\Psi\|$;

(ii) K takes $\mathscr{D}(A)$ into itself and

$$\|A(K(\Phi)-K(\Psi))\|$$
$$\leqslant C(\|\Phi\|,\|A\Phi\|,\|\Psi\|,\|A\Psi\|)\|A\Phi-A\Psi\|;$$

(iii) $\|AK(\Phi)\|\leqslant C(\|\Phi\|)\|A\Phi\|$

for all $\Phi,\Psi\in\mathscr{D}(A)$, where each C is a monotone increasing (everywhere finite) function of the norms indicated.

Now we can state the basic theorems.

 0022-2488/84/051262-04$02.50

Theorem 1: 1. Let (i) hold. Then, for each $\Phi_0 \in \mathscr{H}$ and $t_0 \in \mathbb{R}$ there is a positive number T such that the integral equation

$$\Phi(t) = e^{-i(t-t_0)A}\Phi_0 + \int_{t_0}^{t} e^{-iA(t-s)}K(\Phi(s))ds \tag{2}$$

has a unique continuous solution $\Phi(t)$ on $(-T+t_0, t_0+T)$. Moreover, we note that T can be taken to be any positive number τ such that $(\|\Phi_0\|+1)C(\|\Phi_0\|+1, \|\Phi_0\|+1) < 1/\tau$ and then $\|\Phi(t)\| \leqslant \|\Phi_0\| + 1$ on $(-\tau + t_0, t_0 + \tau)$.

2. Let (i) and (ii) hold. Then, for each $\Phi_0 \in \mathscr{D}(A)$ and $t_0 \in \mathbb{R}$ there is a positive number T such that the initial-value problem

$$\frac{d\Phi(t)}{dt} = -iA\Phi(t) + K(\Phi(t)),$$
$$\Phi(t_0) = \Phi_0 \tag{3}$$

has a unique strongly continuously differentiable solution $\Phi(t) \in \mathscr{D}(A)$ on $(-T+t_0, t_0+T)$. [Note that $\Phi(t)$ is a solution on $(-T+t_0, t_0+T)$ of (2) too.]

Theorem 2: 1. Let (i) hold. If on every finite interval $(-T+t_0, t_0+T)$ on which a continuous solution $\Phi(t)$ of (2) exists, $\|\Phi(t)\|$ is bounded, then (2) has a unique global continuous solution for all t.

2. Let (i), (ii), and (iii) hold. If on every finite interval $(-T+t_0, t_0+T)$ on which a strongly continuously differentiable solution $\Phi(t)$ of (3) exists, $\|\Phi(t)\|$ is bounded, then (3) has a unique global strongly continuously differentiable solution for all t.

3. THE LOCAL PROBLEM

In order to place (1) in the abstract framework of Theorems 1 and 2, let us rewrite it in the form

$$\frac{d}{dt}\begin{pmatrix}\psi(x,t)\\ u(x,t)\\ v(x,t)\end{pmatrix} = -i\begin{pmatrix}-\Delta+1 & 0 & 0\\ 0 & 0 & i\\ 0 & i(\Delta-1) & 0\end{pmatrix}\begin{pmatrix}\psi(x,t)\\ u(x,t)\\ v(x,t)\end{pmatrix} + \begin{pmatrix}-ig\psi(x,t)\tilde{u}(x,t)+i\psi(x,t)\\ 0\\ \mu u(x,t) - g\psi(x,t)\overline{\psi(x,t)}\end{pmatrix}, \tag{4}$$

$$\begin{pmatrix}\psi(x,t_0)\\ u(x,t_0)\\ v(x,t_0)\end{pmatrix} = \begin{pmatrix}\psi_0(x)\\ u_0(x)\\ v_0(x)\end{pmatrix},$$

where $v(x,t) = \partial u(x,t)/\partial t$, $\tilde{u}(x,t) = \operatorname{Re} u(x,t)$, $\Delta = \partial^2/\partial x^2$, and $\mu = 1 - m^2$.

Now let the Hilbert space $\mathscr{H}$ be the direct sum $H^1(\mathbb{R}) \oplus H^1(\mathbb{R}) \oplus L^2(\mathbb{R})$ with the norm $\|\Phi\| = \{\|\psi\|_1^2 + \|u\|_1^2 + \|v\|_0^2\}^{1/2}$, where

$$\Phi = \begin{pmatrix}\psi\\ u\\ v\end{pmatrix} \quad \text{and} \quad \psi, u \in H^1(\mathbb{R}),\ v \in L^2(\mathbb{R}).$$

We define A in $\mathscr{H}$ as

$$\begin{pmatrix}-\Delta+1 & 0 & 0\\ 0 & 0 & i\\ 0 & i(\Delta-1) & 0\end{pmatrix}$$

acting on the domain $\mathscr{D}(A) = H^3(\mathbb{R}) \oplus H^2(\mathbb{R}) \oplus H^1(\mathbb{R})$. We have

Lemma 1: A is self-adjoint.

Proof: Let $\hat{H}^r(\mathbb{R}) = FH^r(\mathbb{R}) = \{\hat{f}; f \in H^r(\mathbb{R})\}$ be the Hilbert space with the norm $\|\hat{f}\|_r = \|f\|_r$ and $\hat{\mathscr{H}} = F\mathscr{H} = FH^1(\mathbb{R}) \oplus FH^1(\mathbb{R}) \oplus FL^2(\mathbb{R})$ be the Hilbert space with the norm $\|\hat{\Phi}\| = \|\Phi\|$. Let $\hat{A}$ be defined in $\hat{\mathscr{H}}$ as the operator of multiplication by

$$\begin{pmatrix}k^2+1 & 0 & 0\\ 0 & 0 & i\\ 0 & -i(k^2+1) & 0\end{pmatrix}$$

with the domain $\mathscr{D}(\hat{A}) = \{\hat{\Phi} \in \hat{\mathscr{H}};\ \hat{A}\hat{\Phi} \in \hat{\mathscr{H}}\}$. It is easy to see that $F\mathscr{D}(A) = \mathscr{D}(\hat{A})$ and, consequently, that $A = F^{-1}\hat{A}F$. But $\hat{A}$ is self-adjoint and F is an isomorphism of $\mathscr{H}$ onto $\hat{\mathscr{H}}$, so that A is likewise self-adjoint. Q.E.D.

We define now a mapping K and its regularization, which will be useful to derive an *a priori* estimate on solutions in the next section. Let for any $\epsilon > 0$, J_ϵ be a smoothing operator and let J_0 be the identity operator. We define K_ϵ, $\epsilon \geqslant 0$ ($K_0 = K$) as the following mapping of $\mathscr{H}$ into itself:

$$K_\epsilon(\Phi) = \begin{pmatrix}-ig\psi J_\epsilon\tilde{u} + i\psi\\ 0\\ \mu u - gJ_\epsilon(\psi\bar{\psi})\end{pmatrix}.$$

Now in order to apply Theorem 1 we need

Lemma 2: (i) $\|K_\epsilon(\Phi) - K_c(\Psi)\| \leqslant C(\|\Phi\|, \|\Psi\|)\|\Phi - \Psi\|$ for any $\epsilon \geqslant 0$.

(ii) For any $\epsilon > 0$ K_ϵ takes $\mathscr{D}(A)$ into itself and

$$\|A(K_\epsilon(\Phi) - K_\epsilon(\Psi))\| \leqslant C_\epsilon(\|\Phi\|, \|A\Phi\|, \|\Psi\|, \|A\Psi\|)\|A\Phi - A\Psi\|.$$

(iii) $\|AK_\epsilon(\Phi)\| \leqslant C_\epsilon(\|\Phi\|)\|A\Phi\|$, for any $\epsilon > 0$

for all $\Phi, \Psi \in \mathscr{D}(A)$, where C and each C_ϵ is a monotone increasing (everywhere finite for $\epsilon > 0$) function of the norms indicated.

(iv) $\lim_{\epsilon \to 0} \|K_\epsilon(\Phi) - K(\Phi)\| = 0$ for all $\Phi \in \mathscr{H}$.

Proof: Let

$$\Phi = \begin{pmatrix}\psi_1\\ u_1\\ v_1\end{pmatrix}, \quad \Psi = \begin{pmatrix}\psi_2\\ u_2\\ v_2\end{pmatrix}.$$

(i) If $\Phi, \Psi \in \mathscr{H}$, $\epsilon \geqslant 0$ then

$$\begin{aligned}\|K_\epsilon(\Phi) - K_\epsilon(\Psi)\| &\leqslant |g|\|\psi_1 J_\epsilon\tilde{u}_1 - \psi_2 J_\epsilon\tilde{u}_2\|_1 + \|\psi_1 - \psi_2\|_1 + |g|\,\|J_\epsilon(\psi_1\bar{\psi}_1 - \psi_2\bar{\psi}_2)\|_0 + |\mu|\,\|u_1 - u_2\|_0\\ &\leqslant \{1 + |\mu| + 2\alpha_1|g|(\|\Phi\| + \|\Psi\|)\}\|\Phi - \Psi\|,\end{aligned}$$

where we have used (a) and (e) (Sec. 2A).

(ii) If $\Phi, \Psi \in \mathscr{D}(A)$, $\epsilon > 0$ then

$$
\begin{aligned}
&\|A(K_\epsilon(\Phi)-K_\epsilon(\Psi))\| \\
&\quad\leqslant |g|\|\psi_1 J_\epsilon \tilde{u}_1 - \psi_2 J_\epsilon \tilde{u}_2\|_3 \\
&\qquad + \|\psi_1-\psi_2\|_3 + |g|\|J_\epsilon(\psi_1\bar{\psi}_1 - \psi_2\bar{\psi}_2)\|_1 \\
&\qquad + |\mu|\|u_1-u_2\|_1 \leqslant (2\pi)^{1/2}\alpha_3|g|\|\hat{j}_\epsilon(k)(1+k^2)\|_\infty \\
&\qquad \times\{\|u_1\|_1\|\psi_1-\psi_2\|_3 + \|\psi_2\|_3\|u_1-u_2\|_1\} \\
&\qquad + \|\psi_1-\psi_2\|_3 + \alpha_1|g|\|\psi_1-\psi_2\|_1(\|\psi_1\|_1 + \|\psi_2\|_1) \\
&\qquad + |\mu|\|u_1-u_2\|_1 \leqslant \{1+|\mu| + \alpha_1|g|(\|\Phi\| + \|\Psi\|) \\
&\qquad + (2\pi)^{1/2}\alpha_3|g|\|\hat{j}_\epsilon(k)(1+k^2)\|_\infty(\|\Phi\| + \|A\Psi\|)\} \\
&\qquad \times\|A\Phi - A\Psi\|,
\end{aligned}
$$

where we have used (a), (e), and (g) (Sec. 2A).

(iii) This follows from (ii) when we take $\Psi = 0$.

(iv) This is an immediate consequence of (f) (Sec. 2A). Q.E.D.

The last two lemmas show that $\mathscr{H}, A, K_\epsilon$ satisfy the assumptions of Theorem 1.1 (and 1.2 if $\epsilon > 0$) so a local existence-uniqueness theorem for the integral equation corresponding to (4) is proved.

The proof of a global existence-uniqueness theorem will be given in the next section.

4. THE GLOBAL PROBLEM

It is our aim in this section to put Theorem 2.1 to work. Therefore it becomes necessary to show that the norm of the solution is bounded on every finite interval on which the solution exists. To this aim we first establish this bound for the smoothing coupling case (i.e., K_ϵ, $\epsilon > 0$). The result we subsequently need for the Yukawa coupling case (i.e., K) will be obtained thence by limiting arguments. We start with

Lemma 3: Let $\Phi_0 \in \mathscr{H}$, $t_0 \in \mathbb{R}$ and $\epsilon > 0$. If $\Phi_\epsilon(t)$ is a strongly continuously differentiable solution on $(-T+t_0, t_0+T)$ of the initial-value problem

$$
\frac{d\Phi_\epsilon(t)}{dt} = -iA\Phi_\epsilon(t) + K_\epsilon(\Phi_\epsilon(t)), \tag{5}
$$

$$
\Phi_\epsilon(t_0) = J_\epsilon\Phi_0 = \begin{pmatrix} J_\epsilon\psi \\ J_\epsilon u \\ J_\epsilon v \end{pmatrix},
$$

where by (c) (Sec. 2A) $J_\epsilon\Phi_0 \in \mathscr{D}(A)$, then

$$
\|\Phi_\epsilon(t)\| \leqslant \gamma(\|\Phi_0\|)e^{|\mu|T}, \quad t \in (-T+t_0, t_0+T), \tag{6}
$$

where γ is a monotone increasing (everywhere finite) function of the norm indicated.

Proof: Let $|\Phi_\epsilon(t)| = \{2\|\psi_\epsilon(t)\|_1^2 + \|u_\epsilon(t)\|_1^2 + \|v_\epsilon(t)\|_0^2\}^{1/2}$ and $P_\epsilon(t) = 2g(\psi_\epsilon(t)\,\overline{\psi_\epsilon(t)}, J_\epsilon\tilde{u}_\epsilon(t))_0$. Then $|\Phi_\epsilon(t)|^2$ and $P_\epsilon(t)$ are continuously differentiable functions on $(-T+t_0, t_0+T)$ and by straightforward calculation [using Eqs. (5) and (d) (Sec. 2A)] one can check that

$$
\frac{d}{dt}\{|\Phi_\epsilon(t)|^2 + P_\epsilon(t)\} = 2\mu\,\mathrm{Re}(u_\epsilon(t), v_\epsilon(t))_0,
$$

$$
\frac{d}{dt}\|\psi_\epsilon(t)\|_0^2 = 0
$$

or

$$
|\Phi_\epsilon(t)|^2 + P_\epsilon(t) = |J_\epsilon\Phi_0|^2 + P_\epsilon(t_0) + 2\mu\,\mathrm{Re}\int_{t_0}^t (u_\epsilon(s), v_\epsilon(s))_0 ds,
$$

$$
\|\psi_\epsilon(t)\|_0 = \|\psi_\epsilon(t_0)\| \tag{7}
$$

for all $t \in (-T+t_0, t_0+T)$. Now from (b) and (e) (Sec. 2A) it follows that

$$
\begin{aligned}
|P_\epsilon(t)| &\leqslant 2|g|\,\|u_\epsilon(t)\|_1\|\psi_\epsilon(t)\|_0^2 \\
&\leqslant |g|\delta\,\|u_\epsilon(t)\|_1^2 + (|g|/\delta)\|\psi_\epsilon(t)\|_0^4
\end{aligned} \tag{8}
$$

for any $\delta > 0$ and $t \in (-T+t_0, t_0+T)$. Thus (7) and (8) imply

$$
\begin{aligned}
|\Phi_\epsilon(t)|^2 &\leqslant |g|\delta\|u_\epsilon(t)\|_1^2 \\
&\quad + 2|\mu|\left|\int_{t_0}^t (u_\epsilon(s), v_\epsilon(s))_0 ds\right| \\
&\quad + (2+|g|\delta)\|\Phi_0\|^2 + 2\frac{|g|}{\delta}\|\Phi_0\|^4, \\
&\quad t \in (-T+t_0, t_0+T).
\end{aligned}
$$

Therefore, if we choose $\delta > 0$ such that $|g|\delta \leqslant \frac{1}{2}$, then

$$
\begin{aligned}
|\Phi_\epsilon(t)|^2 &\leqslant 2|\mu|\left|\int_{t_0}^t |\Phi_\epsilon(s)|^2 ds\right| \\
&\quad + 5\|\Phi_0\|^2 + (2/\delta^2)\|\Phi_0\|^4, \\
&\quad t \in (-T+t_0, t_0+T).
\end{aligned}
$$

Hence using the Gronwall lemma[3] we have

$$
\|\Phi_\epsilon(t)\| \leqslant |\Phi_\epsilon(t)| \leqslant \gamma(\|\Phi_0\|)e^{|\mu|T},
$$

$$
t \in (-T+t_0, t_0+T). \qquad \text{Q.E.D.}
$$

If we now apply Theorem 2.2 we obtain that the initial-value problem (5) has a unique global solution on $\mathbb{R}$. We remark that K does not map $\mathscr{D}(A)$ into $\mathscr{D}(A)$. Therefore in order to show that (6) remains true for $\epsilon = 0$ we need

Lemma 4: Let $\Phi_0 \in \mathscr{H}$, $t_0 \in \mathbb{R}$, $\epsilon > 0$ and let $\Phi(t)$ be a continuous solution on $(-T+t_0, t_0+T)$ of the integral equation (2) with A and K as in Sec. 3. If $\Phi_\epsilon(t)$ is a strongly continuously differentiable solution of (5) on $\mathbb{R}$, then $\Phi_\epsilon(t)$ converges to $\Phi(t)$ in $\mathscr{H}$, as $\epsilon \to 0$, uniformly on compact subsets of $(-T+t_0, t_0+T)$.

Proof: Since $\lim_{\epsilon\to 0}\|\Phi_\epsilon(t_0) - \Phi(t_0)\| = 0$ [see (f) (Sec. 2A)] we need only verify that there exists a positive number τ such that, if for some $t_1 \in (-T+t_0, t_0+T)$,

$$
\lim_{\epsilon\to 0}\|\Phi_\epsilon(t_1) - \Phi(t_1)\| = 0, \tag{9}
$$

then $\Phi_\epsilon(t)$ converges to $\Phi(t)$ in $\mathscr{H}$, as $\epsilon \to 0$, uniformly on $(-\tau+t_1, t_1+\tau) \cap (-T+t_0, t_0+T)$. To prove that, let τ be chosen so that

$$
\begin{aligned}
&(\gamma(\|\Phi_0\|)e^{|\mu|T} + 1) \\
&\quad \times C(\gamma(\|\Phi_0\|)e^{|\mu|T} + 1, \gamma(\|\Phi_0\|)e^{|\mu|T} + 1) < 1/\tau
\end{aligned} \tag{10}
$$

with C as in Lemma 2(i). From (9) and Lemma 3 we have

$$
\|\Phi(t_1)\| = \lim_{\epsilon\to 0}\|\Phi_\epsilon(t_1)\| \leqslant \gamma(\|\Phi_0\|)e^{|\mu|T}.
$$

Hence

$$
(\|\Phi(t_1)\| + 1)C(\|\Phi(t_1)\| + 1, \|\Phi(t_1)\| + 1) < 1/\tau. \tag{11}
$$

Moreover, it is easy to check that for any $t_1 \in (-T+t_0, t_0+T)$, $\Phi(t)$ [resp. $\Phi_\epsilon(t)$] is a continuous solution on $(-T+t_0, t_0+T)$ (resp. $\mathbb{R}$) of the integral equation of type (2) with t_1 in place of t_0 and with $\Phi(t_1)$ [resp. $\Phi_\epsilon(t_1)$] in place of Φ_0 (resp. $J_\epsilon \Phi_0$). Therefore by (11) and Theorem 1.1 we get

$$\|\Phi(t)\| \leqslant \|\Phi(t_1)\| + 1 \leqslant \gamma(\|\Phi_0\|)e^{|\mu|T} + 1, \tag{12}$$
$$t \in (-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T).$$

Next we observe that for all $t \in (-T+t_0, t_0+T)$,

$$\|\Phi_\epsilon(t) - \Phi(t)\| \leqslant \|\Phi_\epsilon(t_1) - \Phi(t_1)\| + \left| \int_{t_1}^{t} \|K_\epsilon(\Phi_\epsilon(s)) - K(\Phi(s))\| ds \right|.$$

On the other hand, by Lemma 2(i), (6), and (12) we get

$$\left| \int_{t_1}^{t} \|K_\epsilon(\Phi_\epsilon(s)) - K_\epsilon(\Phi(s))\| ds \right| \leqslant \frac{1}{\tau} \left| \int_{t_1}^{t} \|\Phi_\epsilon(s) - \Phi(s)\| ds \right|,$$
$$t \in (-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T).$$

Moreover, if we set

$$\eta_\epsilon(t) = \left| \int_{t_1}^{t} \|K_\epsilon(\Phi(s)) - K(\Phi(s))\| \, ds \right|, \quad t \in (-T+t_0, t_0+T),$$

then the integrand is uniformly bounded on $(-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T)$ [see Lemma 2(i) and (12)] and converges to zero as $\epsilon \to 0$ for each $s \in (-T+t_0, t_0+T)$ [see Lemma 2(iv)]. Consequently, $\eta_\epsilon(t)$ converges to zero, as $\epsilon \to 0$, uniformly on $(-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T)$.

Now, from the last two inequalities it follows that

$$\|\Phi_\epsilon(t) - \Phi(t)\| \leqslant \|\Phi_\epsilon(t_1) - \Phi(t_1)\| + \frac{1}{\tau} \left| \int_{t_1}^{t} \|\Phi_\epsilon(s) - \Phi(s)\| ds \right| + \eta_\epsilon(t),$$
$$t \in (-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T).$$

Therefore, if we apply the Gronwall lemma to the above expression we get

$$\|\Phi_\epsilon(t) - \Phi(t)\| \leqslant e^{(1/\tau)|t - t_1|} \times \{\|\Phi_\epsilon(t_1) - \Phi(t_1)\| + \eta_\epsilon(t)\},$$
$$t \in (-\tau + t_1, t_1 + \tau) \cap (-T + t_0, t_0 + T)$$

which gives the desired result. Q.E.D.

We can now prove the main result of this paper

Proposition: The integrated form of the Yukawa coupled Schrödinger and Klein–Gordon equations (1) has a unique global solution in $H^1(\mathbb{R}) \oplus H^1(\mathbb{R}) \oplus L^2(\mathbb{R})$ which depends continuously on initial data.

Proof: From Lemmas 3 and 4 it follows that the norm of a solution of the integral equation (2) with A and K as in Sec. 3 is bounded on every finite interval $(-T+t_0, t_0+T)$ by $\gamma(\|\Phi_0\|)e^{|\mu|T}$. Therefore, by Theorem 2.1 we obtain the uniqueness and global existence.

Now, let $\Phi^1(t)$ and $\Phi^2(t)$ be solutions with initial data Φ_0^1 and Φ_0^2. Then the continuity with respect to initial data follows from the estimate[4]:

$$\|\Phi^1(t) - \Phi^2(t)\| \leqslant \|\Phi_0^1 - \Phi_0^2\| + \left| \int_{t_0}^{t} \|K(\Phi^1(s)) - K(\Phi^2(s))\| ds \right|$$
$$\leqslant \|\Phi_0^1 - \Phi_0^2\| + C_1 \left| \int_{t_0}^{t} \|\Phi^1(s) - \Phi^2(s)\| ds \right|,$$
$$t \in (-T + t_0, t_0 + T),$$

where

$$C_1 = C(\gamma(\|\Phi_0^1\|)e^{|\mu|T}, \quad \gamma(\|\Phi_0^2\|)e^{|\mu|T}),$$

which by Gronwall lemma gives

$$\|\Phi^1(t) - \Phi^2(t)\| \leqslant \|\Phi_0^1 - \Phi_0^2\| e^{|t - t_0|C_1},$$
$$t \in (-T + t_0, t_0 + T). \quad \text{Q.E.D.}$$

[1] I. Fukuda and M. Tsutsumi, Bull. Sci. Eng. Research Lab. Waseda Univ., No. 69 (1975).
[2] J. M. Chadam, J. Funct. Anal. **13**, 173 (1973).
[3] M. Schechter, *Modern Methods in Partial Differential Equations* (McGraw-Hill, New York, 1977).
[4] M. C. Reed, *Abstract Non-linear Wave Equations, Springer Lecture Notes in Mathematics*, Vol. 507 (Springer, New York, 1976).

Sum rule for products of Bessel functions: Comments on a paper by Newberger

M. Bakker and N. M. Temme
Mathematisch Centrum, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

(Received 12 May 1983; accepted for publication 14 October 1983)

Recently, Newberger considered a series of Bessel functions with as a special case the form $\Sigma(n^j J_n^2(z))/(n+\mu)$. The interesting point is that he obtained new explicit expressions for the sum of the series. In this note we point out that some results of Newberger are not correct, especially the results obtained by the principle of analytic continuation. Our remarks include a correction for his important result for the series $\Sigma J_n(z)J_{n-m}(z)/(n+\mu)$.

PACS numbers: 02.30.Gp, 02.30.Lt

1. INTRODUCTION

Newberger[1] presented a sum rule for the infinite series of the form

$$S=\sum_{n=-\infty}^{\infty}\frac{(-1)^n n^j J_{\alpha-\gamma n}(z)J_{\beta+\gamma n}(z)}{n+\mu}, \tag{1.1}$$

where $j\in\mathbb{N}\cup\{0\}$, $\mu\in\mathbb{C}\setminus\mathbb{Z}$, $\alpha,\beta,z\in\mathbb{C}$, $\gamma\in(0,1]$. Initially, α and β are restricted to $\mathrm{Re}\,(\alpha+\beta)>-1$. Under this last restriction, Newberger found interesting explicit expressions for the sum S. Afterwards he extended his results beyond this range of parameters α and β. As will be shown in this note, this last step yields incorrect results.

2. SYMMETRY RELATION FOR S; $\alpha,\beta\in\mathbb{Z}$, $\gamma=1$

An important observation is that S is not defined for all α and β, as stated after (1.1). This will be proved in Sec. 4. Here we consider $\gamma=1$ and integer values of α and β. Then the series is convergent and there is a symmetry rule. To show this we denote S of (1.1) by $S_j(\alpha,\beta,\gamma,\mu)$. Then we have

$$S_j(-\alpha,-\beta,1,-\mu)=(-1)^{j+\alpha+\beta+1}S_j(\alpha,\beta,1,\mu), \tag{2.1}$$

where we used

$$J_{-n}(z)=(-1)^n J_n(z),\quad n\in\mathbb{Z}.$$

The following important special case is considered by Newberger. We define $T_m(z,\mu)=(-1)^m S_0(m,0,1,\mu)$, or explicitly

$$T_m(z,\mu)=\sum_{n=-\infty}^{\infty}\frac{J_n(z)J_{n-m}(z)}{n+\mu},\quad m\in\mathbb{Z}. \tag{2.2}$$

This function arises in a lot of physical problems, for instance in plasma physics. Applying the symmetry rule (2.1) for this case we obtain

$$T_{-m}(z,\mu)=(-1)^{m+1}T_m(z,-\mu). \tag{2.3}$$

Newberger found [see his result (4.6)]

$$T_m(z,\mu)=\frac{(-1)^m\pi}{\sin\mu\pi}J_{m+\mu}(z)J_{-\mu}(z),\quad m\geqslant 0. \tag{2.4}$$

The addition $m\geqslant 0$ is not given by Newberger, but has to be made. To see this, verify the symmetry rule (2.3) for the above relation. It follows that (2.4) cannot be correct for all $m\in\mathbb{Z}$. The correct relation for negative values is

$$T_m(z,\mu)=\frac{\pi}{\sin\mu\pi}J_{-m-\mu}(z)J_{\mu}(z),\quad m\leqslant 0. \tag{2.5}$$

Observe also that (2.2) is an entire function of z, as are the right-hand sides of (2.4) and (2.5). For $m<0$ (2.4) is not entire in z; for $m>0$ (2.5) is not entire in z.

3. A RECURSION FOR $T_m(z,\mu)$

The fact that (2.4) is no longer valid for negative values of m is also revealed by a recursion for $T_m(z,\mu)$. We recall the well-known identities

$$J_{\nu-1}(z)+J_{\nu+1}(z)=(2\nu/z)J_\nu(z), \tag{3.1}$$

$$\sum_{n=-\infty}^{\infty}J_n(z)J_{n-m}(z)=\delta_{m,0}, \tag{3.2}$$

where Kronecker's symbol is used.[2] By using (3.1) we have

$$T_{m-1}(z,\mu)+T_{m+1}(z,\mu)=\frac{2}{z}\sum_{n=-\infty}^{\infty}\frac{J_n(z)(n-m)J_{n-m}(z)}{n+\mu}.$$

Writing $n-m=(n+\mu)-(\mu+m)$ we obtain with (3.2)

$$T_{m-1}(z,\mu)+T_{m+1}(z,\mu)=-\frac{2(m+\mu)}{z}T_m(z,\mu)+\frac{2}{z}\delta_{m,0}. \tag{3.3}$$

Observe that this recursion relation is an inhomogeneous version of (3.1). Without the term $(2/z)\delta_{m,0}$ a solution would be $(-1)^m J_{m+\mu}(z)$ ($m\in\mathbb{Z}$) times a factor not depending on m. Hence, since $T_m(z,\mu)$ satisfies (3.3) ($m\in\mathbb{Z}$) and the right-hand side of (2.4) satisfies the homogeneous version of (3.3) for $m\in\mathbb{Z}$, it follows once again that (2.4) cannot be true for all $m\in\mathbb{Z}$. For $m\neq 0$ (3.3) gives the proper recursion for both (2.4) and (2.5). On the other hand we have, using (2.4), (2.5), and (3.1),

$$T_{-1}(z,\mu)+T_1(z,\mu)=\frac{-\pi}{\sin\mu\pi}\left[J_\mu(z)J_{-\mu-1}(z)+J_{-\mu}(z)J_{\mu+1}(z)\right]-\frac{2\mu\pi}{z\sin\mu\pi}J_\mu(z)J_{-\mu}(z).$$

Interpreting the cross product of Bessel functions as a well-known Wronskian relation for these functions[3] we obtain

$$T_{-1}(z,\mu)+T_1(z,\mu)=2/z-(2\mu/z)T_0(z,\mu),$$

which confirms (3.3) for $m=0$.

 0022-2488/84/051266-02$02.50

4. CONVERGENCE OF THE SERIES (1.1)

The convergence of the series (1.1) follows from the asymptotic expansions

$$J_\nu(z) \sim (\tfrac{1}{2}z)^\nu/\Gamma(\nu+1), \qquad \mathrm{Re}\ \nu \to \infty, \tag{4.1}$$

$$J_{-\nu}(z) \sim (1/\pi)(\tfrac{1}{2}z)^{-\nu}\Gamma(\nu)\sin\nu\pi,$$

The second line holds for noninteger values of ν; otherwise we use $J_{-n}(z) = (-1)^n J_n(z)$, $n \in \mathbb{Z}$. Using (4.1) for the terms of (1.1) we obtain

$$\frac{n^j}{n+\mu} J_{\alpha-\gamma n}(z) J_{\beta+\gamma n}(z) \sim n^{j-1}(\tfrac{1}{2}z)^{\alpha+\beta} \times \sin[(\gamma n - \alpha)\pi] \frac{\Gamma(\gamma n - \alpha)}{\Gamma(\gamma n + \beta + 1)}.$$

Using $\Gamma(z+a)/\Gamma(z+b) \sim z^{a-b}$, $\mathrm{Re}\ z \to \infty$, we conclude that the series diverges when

$$\mathrm{Re}(\alpha+\beta) < j-2,$$

unless $\alpha, \beta \in \mathbb{Z}$, $\gamma = 1$. In general, large terms for $n \to \pm\infty$ will not cancel each other. Thence there is no chance that the divergence at $n = +\infty$ combined with that at $n = -\infty$ is removed.

The series (1.1) is absolutely convergent when $\mathrm{Re}(\alpha+\beta) > j-1$. This condition is sufficient to make the sum holomorphic with respect to α and β in this domain.

It follows that S, as a function of the complex parameters α and β, is defined and holomorphic for $\mathrm{Re}(\alpha+\beta) > j-1$. Possibly there is an analytic continuation of $S(\alpha, \beta, \mu)$ with respect to $\mathrm{Re}(\alpha, \beta) \leqslant j-1$, but it is not clear what this continuation looks like. For $\alpha, \beta \in \mathbb{Z}$, $\gamma = 1$, the symmetry rule (2.1) gives the value for negative α and β.

Newberger used the splitting

$$S = (-\mu)^j S_1 + S_2 \tag{4.2}$$

with S_1 equal to (1.1) with $j = 0$. He evaluated this expression in the form [see his formula (2.8)]

$$S_1 = \frac{\pi}{\sin\mu\pi} J_{\alpha+\gamma\mu}(z) J_{\beta-\gamma\mu}(z). \tag{4.3}$$

His proof is correct for the range $\mathrm{Re}(\alpha+\beta) > -1$. The right-hand side is entire in α and β, whereas from the above remarks it follows that some combinations of α and β yield a divergent series. Extension of (4.3) to all complex α and β (and γ) is therefore not allowed.

5. A FINAL REMARK

The second part of (4.2), i.e., S_2, is also evaluated in terms of derivatives of Bessel functions. Starting point is the evaluation of

$$\hat{S}_2 = \sum_{n=-\infty}^{\infty} (-1)^n n^p J_{\alpha-\gamma n}(z) J_{\beta+\gamma n}(z), \tag{5.1}$$

where p is an integer, $0 \leqslant p \leqslant j-1$.

As admitted by Newberger, the analysis for deriving the sum rule for S_2 is quite formal, with an appeal to the theory of generalized functions. However, an approach without distributions is possible here. For instance, application of (3.1) gives a recursion relation [denote (5.1) by $\hat{S}_2(\alpha, \beta, p)$]

$$\gamma\hat{S}_2(\alpha,\beta,p) = \alpha\hat{S}_2(\alpha,\beta,p-1) - \tfrac{1}{2}z[\hat{S}_2(\alpha+1,\beta,p-1) + \hat{S}_2(\alpha-1,\beta,p-1)].$$

Repeated application reduces (5.1) to the case $p = 0$. One further step makes the Fourier series in Newberger's formulas (2.12) and (2.13) convergent.

ACKNOWLEDGMENT

The remarks by B. S. Newberger and J. Boersma on a first version of the paper are much appreciated.

[1] B. S. Newberger, "New sum rule for products of Bessel functions with application to Plasma Physics," J. Math. Phys. **23**, 1278–1281 (1982).
[2] G. N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed. (Cambridge U. P., London, 1952), Sec. 5.3.
[3] Reference 2, p. 46.

Fractional approximations for the spherically symmetric Coulomb scattering wave functions

Eduardo Chalbaud and Pablo Martín
Universidad Simón Bolívar, Departamento de Física, Apdo. 80659, Caracas 1080A, Venezuela

(Received 26 July 1983; accepted for publication 28 October 1983)

An improved method is presented for obtaining fractional approximations. The fractional parameters are now solutions of a set of linear equations, and no nonlinear equations are involved as in the previous procedure. Excellent fractional approximations are presented for the Coulomb functions for $\eta = 0.5$, 1, 2, and 5. The accuracy is sufficient for most of the computations where this function is used. The straightforward extension to higher orders is indicated.

PACS numbers: 02.30.Hq, 02.30.Mv, 02.60.Lj, 03.80. + r

I. INTRODUCTION

The Coulomb scattering wave functions are of fundamental importance to a wide range of problems in atomic and nuclear physics.[1] Nevertheless, calculations in scattering problems with charged particles can lead to errors as large as 100%, as in the plane wave Born approximation, or to formidable computational difficulties, as in the distorted wave Born approximation.[2] We thus felt that a simple, manageable polynomial fraction that could be employed in integral forms with ease would be useful in such and other calculations. From the mathematical point of view, our interest in the Schrödinger radial differential equation with a Coulomb potential, and its solution, lies in its relation, through a simple transformation, to Kummer's equation and thus to the confluent hypergeometric functions.[3,4]

The method of fractional approximations used here is based on the one originated by Martín and Zamudio-Cristi[5] for linear first-order differential equations [$Z(s)$, $Ei(z)$], and extended to second order by Martín and Guerrero[6] [$J_0(x)$, $J_1(x)$, $Ai(x)$]. We have introduced improvements and new developments in the general method of generating fractional approximations. The existence of a complex exponent in the multiplicative factor has led to an added difficulty in the treatment. That difficulty was reduced by employing a 1 rather than a complex parameter α, since there is thus no ambiguity in the principal value used; further, the resulting equations have a simpler form, and thus the extension to higher order is straightforward. Here, after extracting an exponential and power multiplicative term, we determine a power series and an asymptotic expansion for the solution of the transformed equation in terms of two complex constants and of the strength parameter η of the Coulomb field. We then determine the parameters of the fractional approximation as solutions of a set of linear algebraic equations whose coefficients are the coefficients of both series. This procedure simplifies and systematizes the calculation of the parameters of the approximation, since the system of algebraic equations is now linear. No equations of second or higher degree appear as in the previous paper,[6] and higher-order approximations can be easily obtained. Finally, a further improvement has consisted of including an imaginary part in the power series which permits greater accuracy of the fractional approximation for a given order. The phase of this complex constant is almost independent of the degree of the approximation, so once it is determined for the simplest approximation, the same phase is valid, with minor adjustments, for approximations of higher degree.

Our results show excellent approximations to the Coulomb function for an ample range of values of η from 0 to 5. For instance, the maximum error for $\eta = 1$ is as low as 0.015, equivalent to 7% of relative error. This maximum error occurs in a narrow range of the variable ρ, and soon decreases more than an order of magnitude. Clearly, this accuracy would be enough for most of the computations where the Coulomb function is used.

The paper has been arranged in six sections. Section II reviews the general method for obtaining fractional approximations of linear differential equations illustrating it for the Coulomb equation, and presents the new developments and improvements. Section III discusses the determination of the values of the parameters for the fractional approximation of the Coulomb scattering wave function. Sections IV and V present the results for $\eta = 1$ and other η's respectively, together with a graphical analysis. The conclusions are discussed in Sec. VI.

II. THEORETICAL TREATMENT AND IMPROVEMENTS

The radial Coulomb scattering equation for $L = 0$ is[4]

$$\rho F'' + (\rho - 2\eta)F = 0. \tag{1}$$

Now we should obtain a fractional approximation of the type $P_n(\rho)/Q_m(\rho)$ valid for small as well as for large values of ρ. The procedure already described[5] consists in substituting $P_n(\rho)/Q_m(\rho)$, rationalizing and equating to zero the higher and lower powers of the resulting polynomial. Clearly, if this procedure is done directly in the differential equation, the largest power in the first term is ρ^{2m+n-1}, and the largest power in the second term ρ^{2m+n+1}. As a consequence, incompatibilities appear in the equations obtained from the highest power. To avoid these inconsistencies, suitable transformations of the dependent and independent variables should be done. In this case, however, modification of the independent variable was found to be not necessary; on the other hand, the dependent variable will be modified with a power and exponential factor in the following way:

$$F(\rho) = e^{\mu\rho}(1+\rho)^s v(\rho). \tag{2}$$

This transformation is similar to the ones described in a pre-

 0022-2488/84/051268-06$02.50

vious paper,[6] but with the choice of α as 1. This will introduce a very important economy in the calculation of the parameters of the fractional approximation due to a strong simplification of the equations used to determine these parameters. Furthermore, the computation of the approximation is also simplified since we do not have to worry about the principal value of the power term $(1+\rho)^s$. In order to find ρ and s, it is better to proceed in two steps. First we make the transformation

$$F(\rho) = e^{\mu\rho} u(\rho), \tag{3}$$

and the differential equation becomes

$$\rho u'' + 2\mu\rho u' + \{(\mu^2+1)\rho - 2\eta\}u = 0. \tag{4}$$

Here μ should be $\pm i$. Since, to obtain the real Coulomb function, we add one function and its complex conjugate, we will only pay attention to $\mu = i$. The second transformation is

$$u(\rho) = (1+\rho)^s v(\rho), \tag{5}$$

giving now the equation

$$\rho^2 v'' + (2s\rho + 2i\rho^2)v' + \{s(s-1) + 2is\rho - 2\eta\rho\}v = 0. \tag{6}$$

Here, by using the fractional approximation $P_n(\rho)/Q_m(\rho)$ and rationalizing, the highest power in ρ gives the equation

$$2i(n-m) + 2is - 2\eta = 0. \tag{7}$$

The simplest choice is $n = m$ and $s = -i\eta$, and this leads to an equation for v of the form

$$(\rho^3 + 2\rho^2 + \rho)v'' + \{2i\rho^3 + (4i - 2i\eta)\rho^2 + (2i - 2i\eta)\rho\}v' + \{(i\eta - 2\eta - \eta^2)\rho - 2\eta\}v = 0. \tag{8}$$

Now the differential equation has a form suitable for a fractional approximation that can be used for small and large values of the dependent variable ρ. Thus, we can approximate $v(\rho)$ by the fraction $\tilde{v}(\rho)$ defined as

$$\tilde{v}(\rho) = \left(\sum_{k=1}^{n} p_k \rho^k\right)\left(1 + \sum_{k=1}^{n} q_k \rho^k\right). \tag{9}$$

Thus the approximation for the Coulomb function will be of the form

$$\tilde{F}(\rho) = \tilde{v}(\rho)\exp(i\{\rho - \eta\ln(1+\rho)\}) + \text{c.c.} \tag{10}$$

It is interesting to point out that we have started the numerator of the fraction with the first power of the variable, rather than with a constant, because of the behavior of the Coulomb function. In this paper we have gone only to fourth power in the variable, but the extension to higher powers is indicated and does not present any new difficulties.

In order to avoid nonlinear algebraic systems of equations, we have modified the procedure described in a previous paper[5] by computing now the power series and asymptotic expansion for v. Because of the indicial equation, we write for v a power series of the form

$$v = \rho \sum_{k=0}^{\infty} a_k \rho^k, \tag{11}$$

which we introduce in the above Eq. (8), and, setting the coefficients of the subsequently increasing powers of the variable equal to zero, we find the a_k in terms of a_0. The first three coefficients are

$$a_1 = \{\eta + i(\eta - 1)\}a_0, \tag{12a}$$

$$a_2 = \{-\eta^2 + 6\eta - 4 + i(6\eta^2 - 9\eta)\}a_0/6, \tag{12b}$$

$$a_3 = -\{8\eta^3 - 27\eta^2 + 20\eta + i(-3\eta^3 + 6\eta^2 + 6\eta - 6)\}a_0/18. \tag{12c}$$

A similar treatment is made for the asymptotic expansion. We first divide Eq. (8) by ρ^3 and introduce in it a series for v of the form

$$v = \sum_{k=0}^{\infty} b_k \left(\frac{1}{\rho}\right)^k. \tag{13}$$

Setting the coefficients of the subsequently decreasing powers of the variable equal to zero, we obtain the b_k in terms of b_0. The first three coefficients are

$$b_1 = \{\eta + i(\eta^2 + 2\eta)\}b_0/2, \tag{14a}$$

$$b_2 = \{\eta^2 - 4\eta^3 - \eta^4 + i(4\eta^3 + 4\eta^2 - 6\eta)\}b_0/8, \tag{14b}$$

$$b_3 = \{-9\eta^5 - 24\eta^4 + 51\eta^3 + 36\eta^2 - 12\eta + i(-\eta^6 - 6\eta^5 + 19\eta^4 + 22\eta^3 - 52\eta^2 + 16\eta)\}b_0/48. \tag{14c}$$

The coefficients a_0 and b_0 are determined by the behavior of the function at zero and infinity, respectively. For subsequent a_k and b_k, one can use

$$a_k = [-1/k(k+1)]\{a_{k-1}(2k^2 - 2k + 2ik - 2i\eta k - 2\eta) + a_{k-2}\{(k-1)(k-2) + (4i - 2i\eta)(k-1) + (i\eta - 2\eta - \eta^2)\} + a_{k-3}(2ik - 4i)\} \tag{12'}$$

and

$$b_k = (1/2ik)\{b_{k-1} \times \{(k-1)k - (4i - 2i\eta)(k-1) + (i\eta - 2\eta - \eta^2)\} + b_{k-2}\{2(k-2)(k-1) - (2i - 2i\eta)(k-2) - 2\eta\} + b_{k-3}(k-2)(k-3)\}. \tag{14'}$$

In order to determine the parameters p_l and q_j of the fractional approximation, the power series and asymptotic expansion of the fractional approximation should coincide with the terms of the power series and asymptotic expansion already computed. Since we want only a linear system of equations, it is convenient to rationalize and write out these conditions as follows:

$$\sum_{l=1}^{n} p_l \rho^l \cong \left(1 + \sum_{j=1}^{n} q_j \rho^j\right)\left(\rho \sum_{k=0}^{\infty} a_k \rho^k\right), \tag{15}$$

$$\sum_{l=0}^{n-1} p_{n-l}\left(\frac{1}{\rho}\right)^l \cong \left(\sum_{j=0}^{n-1} q_{n-j}\left(\frac{1}{\rho}\right)^j + \left(\frac{1}{\rho}\right)^n\right)\left(\sum_{k=0}^{\infty} b_k \left(\frac{1}{\rho}\right)^k\right). \tag{16}$$

Here we get a linear system of equations to determine the p_l and q_j. Clearly, we could not identify all of the terms in both Eqs. (15) and (16) since we have to use only $2n$ equations. There are several possibilities depending on the kind of aproximation wanted, i.e., for better results at zero than at infinity, one should include in general more terms from the power series than from the asymptotic expansion. In our case we have used the same number of terms from the power

series as from the asymptotic expansion. The important point now is that all the resulting equations for q_j and p_l are linear on these parameters, and no second or higher degree equations appear, as they did in the previous procedure.[6]

III. DETERMINATION OF THE PARAMETERS FOR COULOMB FUNCTION

The leading term in the asymptotic expansion of the Coulomb functions goes as

$$\sin(\rho - \eta \ln 2\rho + \sigma) = Ae^{i(\rho - \eta \ln \rho)} + \text{c.c.}, \tag{17}$$

where

$$A = \tfrac{1}{2} \exp[i(\sigma - \eta \ln 2 - \pi/2)] \tag{18}$$

and

$$\sigma = \arg \Gamma(1 + i\eta); \tag{19}$$

thus, if we want our approximation to have the same asymptotic form, we must choose $b_0 = A$. Near zero, the exact function behaves as $C\rho$, where

$$C = [2\pi\eta/(e^{2\pi\eta} - 1)]^{1/2}. \tag{20}$$

Here, instead of choosing $a_0 = C/2$, which does not lead to the most accurate approximations, we include an imaginary part and write

$$a_0 = \tfrac{1}{2} C + \tfrac{1}{2} iC \tan \delta, \tag{21}$$

where the phase of a_0,δ, can have an arbitrary value between $-\pi/2$ and $+\pi/2$. Since we have to add the complex conjugate term, clearly the exact function is not modified by the inclusion of any phase δ in the parameter a_0. However, the approximation depends strongly on δ.

With the values of a_0 and b_0, we can proceed to determine the other a's and b's using Eqs. (12) and (14) and then the p_l and q_j of our fractional approximation, as was explained in the previous section. We have found that the best approximation is obtained if we take an equal number of equations from the power series as from the asymptotic expansion. For the fourth-order approximation, the eight equations, in their simplest form, that determine the p_l and q_j are

$$p_1 = a_0, \quad p_2 = a_0 q_1 + a_1, \quad p_3 = b_0 q_3 + b_1 q_3, \quad p_4 = b_0 q_4, \tag{22}$$

$$(a) = M(q), \tag{23a}$$

where (a) and (q) are the column matrices (a_0,a_1,a_2,a_3) and (q_1,q_2,q_3,q_4), and M is the square matrix

$$M = \begin{Bmatrix} b_0 & b_1 & b_2 & b_3 \\ -a_0 & b_0 & b_1 & b_2 \\ -a_1 & -a_0 & b_0 & b_1 \\ -a_2 & -a_1 & -a_0 & b_0 \end{Bmatrix}. \tag{23b}$$

The first four equations [Eqs. (22)] are obtained directly from the first two terms of Eq. (15) and (16), respectively. Equations (23) are, in matrix form, two by two combinations of equations, one from the preceding group, and the other from the third and fourth terms of Eqs. (15) and (16). The combination is done in the way of deleting the p_l. We can thus solve Eq. (23) for the q_j using standard techniques and then find the p_l using the four Eqs. (22). These equations can be used for fractional approximations of fourth degree or lower; in the last case, we must set some of the parameters equal to zero. If we want to go to higher order, we must include the higher terms in Eqs. (15) and (16). The symmetry of Eqs. (22) and (23) can guide us in writing those equations without doing the algebra. For first order the values of the fractional parameters are

$$q_1 = a_0/b_0 \text{ and } p_1 = a_0.$$

For the second order, the fractional parameters are

$$q_1 = a_0(2 - \{\eta + i(\eta - 1)\}\{\eta + i(\eta^2 + 2\eta)\})/(2b_0 + a_0\{\eta + i(\eta^2 + 2\eta)\}), \tag{24a}$$

$$q_2 = 2(a_0/b_0)(a_0 + b_0\{\eta + i(\eta - 1)\})/(2b_0 + a_0\{\eta + i(\eta^2 + 2\eta)\}), \tag{24b}$$

$$p_1 = a_0, \tag{24c}$$

$$p_2 = b_0 q_2. \tag{24d}$$

For higher order, the analytic expressions in terms of η become too cumbersome, and it is best to solve for a particular η by first evaluating the a_k and b_k substituting in Eqs. (22) and (23) and solving numerically.

IV. RESULTS FOR $\eta = 1$

Figure 1 shows the fourth degree fractional approximation for $\eta = 1$, and for the best value of δ which in this case is -1.46 rad, together with the exact function and the absolute error. The maximum error is less than 0.015 and occurs for small values of ρ ($\rho \simeq 1$), around the point of inflection and the first maximum. This absolute error corresponds to a relative error of about 6%. Beyond $\rho = 7$ the error decreases by more than an order of magnitude. The approximation in Fig. 1 is so good that we cannot really see any difference with the exact function on this scale. The absolute error is included in the figure in order to show the relative error of the approximation. A plot of the relative error is not adequate because the function has several zeros, and there the relative error will be infinite.

Table I shows the values of the fractional parameters

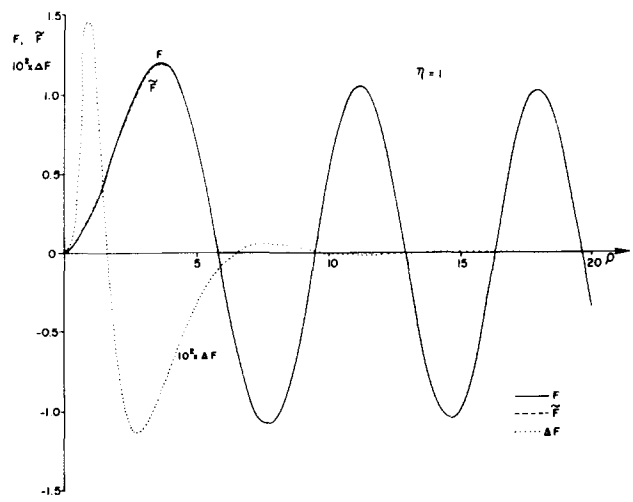


FIG. 1. Fourth degree approximation $\tilde{F}$ (Table I), exact Coulomb function F and absolute error ΔF amplified by a factor of 100, for $\eta = 1$.

TABLE I. Fractional parameters for the fourth degree approximation for $\eta = 1$ and $\delta = -1.46$.

$q_1 =$	$0.635\,52 - 0.371\,44i$
$q_2 =$	$0.149\,57 + 0.602\,79i$
$q_3 =$	$0.631\,96 - 0.720\,79i$
$q_4 =$	$0.866\,45 + 0.274\,04i$
$p_1 =$	$0.054\,211 - 0.487\,28i$
$p_2 =$	$-0.092\,335 - 0.817\,10i$
$p_3 =$	$-0.079\,315 - 0.478\,34i$
$p_4 =$	$-0.288\,69 - 0.350\,88i$

TABLE II. Fractional parameters for $\eta = 1$.

Third degree $\delta = -1.53$		Second degree $\delta = -1.49$	
$q_1 =$	$0.500\,52 + 2.1691i$	$q_1 =$	$-0.582\,6 - 1.180i$
$q_2 =$	$1.574\,2 - 2.1943i$	$q_2 =$	$1.662 - 0.2601i$
$q_3 =$	$4.239\,9 + 4.4131i$	$p_1 =$	$0.054\,21 - 0.6695i$
$p_1 =$	$0.054\,211 - 1.3281i$	$p_2 =$	$-0.767\,6 - 0.3434i$
$p_2 =$	$2.962\,1 - 1.8752i$		
$p_3 =$	$-0.576\,03 - 3.0052i$		

First degree $\delta = -1.39$

$q_1 = 0.232 + 0.556i$
$p_1 = 0.0542 - 0.297i$

for the fourth degree approximation and for the best value of the phase δ. When this phase is varied from its optimum value in either direction, the approximation worsens. This phase is very important since the accuracy of the approximation can change by a factor larger than 100 for positive phases (near $+\pi/2$). We illustrate this in Fig. 2. Here we show a semilogorithmic plot of the absolute error in the fourth degree approximation as a function of the phase, and we can appreciate how important this phase is for the accuracy of the approximation. For instance, the maximum error is more than 100 times larger for positive values of δ larger than 1.3 rad.

In Table II we give the values of the fractional parameters for first, second, and third degree approximations and the best δ in each case. Clearly this phase changes very little with the degree and accuracy of the approximation. In this table, as in the previous one, the parameters are given with the smallest number of decimals needed for the accuracy of the approximation. We could in both tables suppress a decimal figure or two without increasing the maximum error, but that would decrease the accuracy for large values of ρ, which, for instance, by $\rho = 20$ is as high as four digits in Table I. Figures 3, 4, and 5 show the approximated function for first, second, and third degree, respectively, together with the exact function. Figure 3 is the simplest approximation possible in this method, and it already reproduces the location of the maxima and minima of the exact function, as well as the zeros. The worst part of the approximation is from near $\rho = 0$ to $\rho = 5$. After that the approximation gets progressively better. The maximum error in this approximation is 0.5 and occurs just before the first maximum at $\rho = 2.8$, the relative error there is 44%. Figure 4 shows a marked improvement in accuracy with respect to the first degree approximation. Here, as in the previous case, the worst part is near $\rho = 0$ through the first maximum; but now the maximum error is 0.1 at $\rho = 3$, and there the relative error is about 10%. This maximum error occurs near the first maximum and decreases, for $\rho > 15$, more than an order of magnitude. Figure 5 shows that the third degree approximation is excellent and almost as good as the fourth degree (see Fig. 1). Here the maximum error occurs also right before the first maximum, it is 0.042 at $\rho = 2$, and the corresponding relative error is about 6%. The approximation gets progressively better; for instance, by $\rho = 13$ its accuracy is as high as three digits, and by $\rho = 19$ it is as high as four digits.

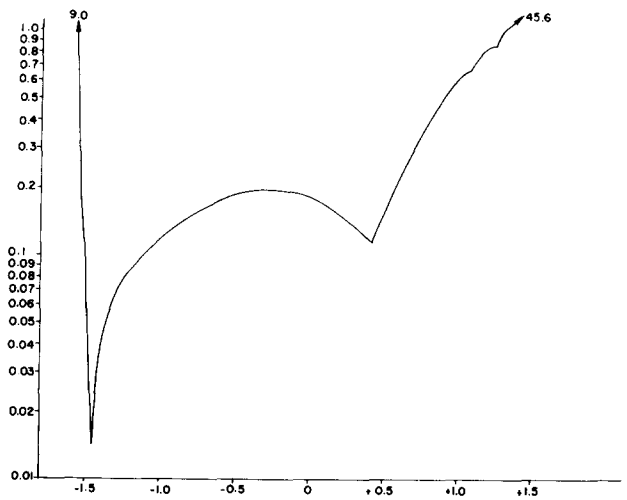


FIG. 2. Semi-log plot of absolute error in fourth degree approximation for $\eta = 1$ as a function of the phase δ.

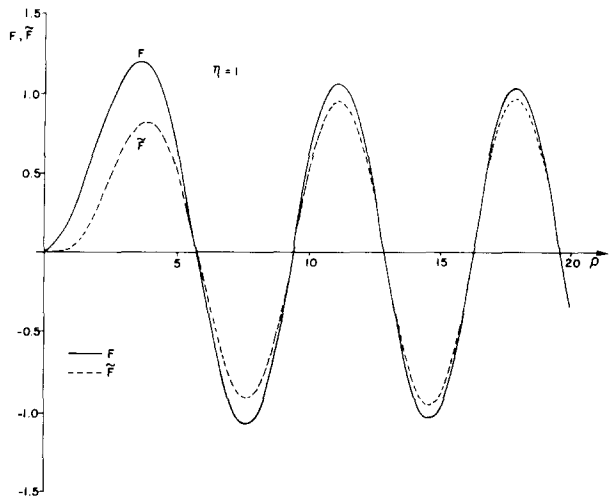


FIG. 3. First degree approximation $\tilde{F}$ (Table II) and exact Coulomb function F for $\eta = 1$.

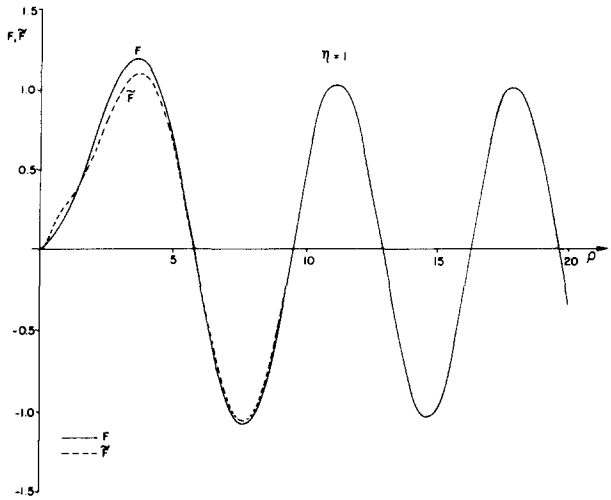


FIG. 4. Second degree approximation $\tilde{F}$ (Table II) and exact Coulomb function F for $\eta = 1$.

TABLE III. Fractional parameters for fourth degree approximations for $\eta = 0.5$, $\eta = 2$, and $\eta = 5$.

$\eta = 0.5, \quad \delta = -1.40$		$\eta = 2, \quad \delta = -1.46$	
$q_1 =$	$1.177\,9 - 0.208\,62i$	$q_1 =$	$0.189\,1 - 0.276\,5i$
$q_2 =$	$2.167\,3 + 2.049\,8i$	$q_2 =$	$-0.063\,62 + 0.012\,99i$
$q_3 =$	$2.364\,2 - 1.234\,2i$	$q_3 =$	$0.048\,80 - 0.073\,23i$
$q_4 =$	$8.887\,5 + 3.161\,1i$	$q_4 =$	$0.032\,22 + 0.025\,97i$
$p_1 =$	$0.188\,34 - 1.092\,0i$	$p_1 =$	$0.003\,31 - 0.029\,75i$
$p_2 =$	$-0.457\,7 - 1.965\,8i$	$p_2 =$	$0.028\,77 - 0.062\,74i$
$p_3 =$	$1.395\,6 - 2.507\,1i$	$p_3 =$	$0.023\,47 - 0.035\,28i$
$p_4 =$	$-1.161\,9 - 4.571\,1i$	$p_4 =$	$-0.011\,31 - 0.017\,33i$

$\eta = 5, \quad \delta = 1.46$

$$
\begin{aligned}
q_1 &= 0.186\,3 + 0.187\,8i \\
q_2 &= -0.021\,17 + 0.034\,85i \\
q_3 &= -3.762\times10^{-3} - 2.033\times10^{-3}i \\
q_4 &= 1.361\times10^{-4} - 1.917\times10^{-4}i \\
p_1 &= 4.223\times10^{-7} + 3.796\times10^{-6}i \\
p_2 &= -1.371\times10^{-5} + 2.146\times10^{-5}i \\
p_3 &= -7.282\times10^{-5} + 9.353\times10^{-6}i \\
p_4 &= -6.667\times10^{-5} - 9.680\times10^{-5}i
\end{aligned}
$$

We should also note that this approximation is much better in the region near $\rho = 0$ than the two previous ones. We see that as the degree of the fraction increases by 1, the accuracy increases and the maximum error decreases by roughly one third of the maximum error of the previous degree.

V. RESULTS FOR OTHER η

Table III shows the fractional parameters for the fourth degree approximations for $\eta = 0.5$, 2, and 5 and the best δ for which each approximation is obtained. In this, as in previous tables, we give the smallest number of digits for the parameters consistent with the accuracy of the approximation. Figures 6–8 illustrate the approximations obtained with this table, together with the exact function and the absolute error. The approximation for $\eta = 0.5$ (Fig. 6) is excellent, and in the scale of the figure we cannot really distinguish between the exact function and the approximation. The maximum error here is 0.02, corresponding to a relative error of about 7%, and occurs around $\rho = 0.7$. For ρ larger than 7 the accuracy is as high as four digits. This case is typical of all cases where $\eta < 1$, where excellent approximations are obtained. For $\eta = 0$, our approximation gives a sine function with a maximum error smaller than 10^{-5}; the p's and q's in this case, however, are very large.

The fourth degree approximation for $\eta = 2$ (Fig. 7) is very good and, after the first maximum, practically indistinguishable from the exact function in the scale of the figure. The accuracy, though high, is not as good as in the cases $\eta = 0.5$ or 1. The maximum error is 0.08 and occurs for

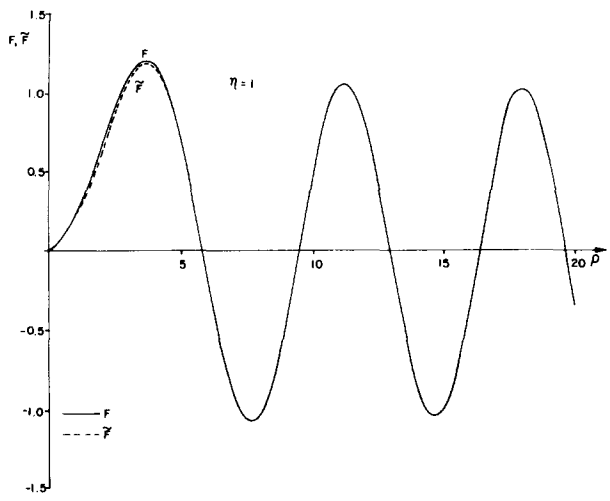


FIG. 5. Third degree approximation $\tilde{F}$ (Table II) and exact Coulomb function F for $\eta = 1$.

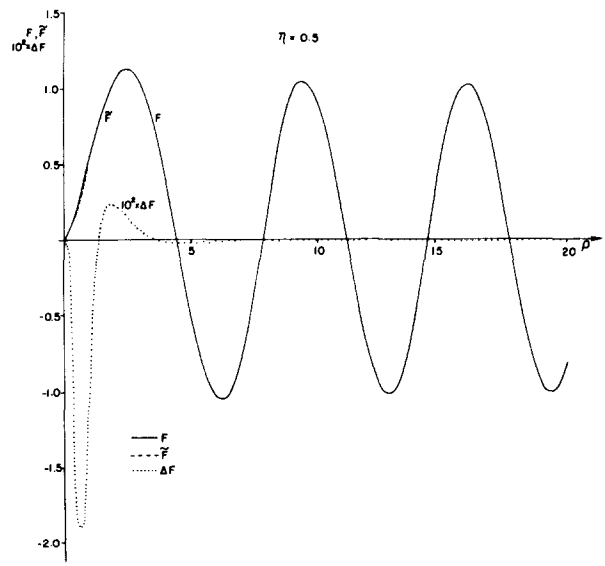


FIG. 6. Fourth degree approximation $\tilde{F}$ for $\eta = 0.5$ (Table III) exact Coulomb function F and absolute error ΔF amplified by a factor of 100.

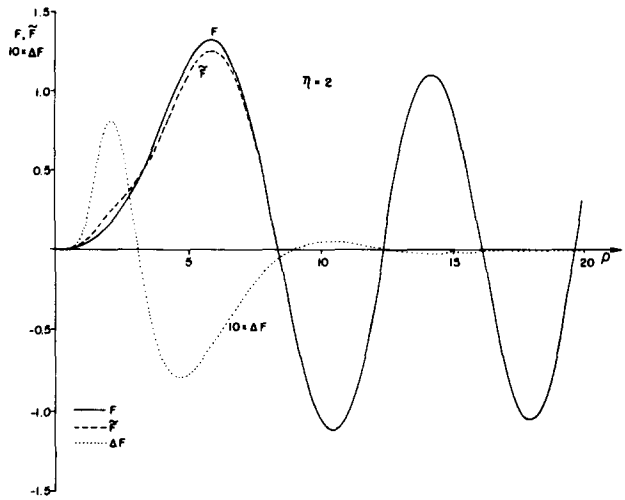


FIG. 7. Fourth degree approximation $\tilde{F}$ for $\eta = 2$ (Table III), exact Coulomb function F and absolute error ΔF amplified by a factor of 10.

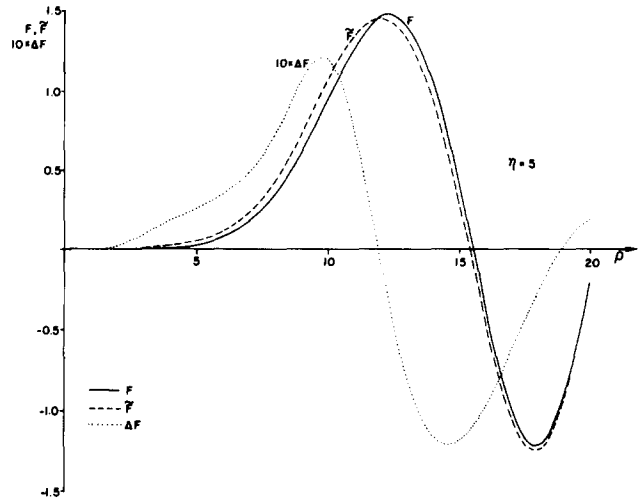


FIG. 8. Fourth degree approximation $\tilde{F}$ for $\eta = 5$ (Table III), exact Coulomb function F and absolute error ΔF amplified by a factor of 10.

small ρ and near the point of inflection $\rho = 99$. We should note that the maxima and minima are well reproduced and also the zeros of the exact function.

The fourth degree approximation for $\eta = 5$ (Fig. 8) is typical of the large η cases, where the approximations get less good with increasing values of η. For $\eta = 5$ the maximum error is 0.14 and occurs near the first maximum and again near the first minimum. The first maximum, in the approximation, is shifted to the left by $\Delta\rho = 0.30$. The other maxima and minima are well reproduced, and the approximation gets, as is the case with all these approximations, much better with increasing ρ. For this η, the value of the phase δ is positive. If one wants to increase the accuracy, fractions of higher degree should be used.

VI. CONCLUSIONS AND DISCUSSION

We have presented improvements to the method of obtaining fractional approximations previously published which permit a quicker and simpler way of obtaining the parameters of the fractional approximation. In the present method, to eliminate the singularity at the origin, instead of substituting $\alpha + x$ for x we use $1 + x$, and this simplifies the subsequent equations. We also eliminate nonlinear equations in the fractional parameters by first determining the power series and the asymptotic expansion of the transformed equation, and from those we determine the fractional parameters. This leads to a set of linear equations that are simple to solve. Another improvement in the method is the inclusion of an imaginary part in the first coefficient of the power series, a_0, which yields, for a given degree, approximations which are an order of magnitude or so better than if this imaginary term were zero. We have given first, second, third, and fourth degree approximations for $\eta = 1$ and shown how the accuracy increases with the degree. We have also shown that the value of δ is almost independent of the degree. We have presented fourth degree approximations for the Coulomb function with maximum errors of 0.02, 0.015, 0.08, and 0.14 for $\eta = 0.5$, 1, 2, and 5, respectively. This maximum error is always near the inflection point, $\rho = 2\eta$ and the first maximum of the function, and decreases substantially as ρ increases therefrom. Given the great precision of these approximations and the fact that the error decreases with increasing ρ, in such computations where the function has to be integrated, either in a finite or infinite range, these approximations can be used instead of the real function with great confidence in the accuracy of the results.

[1] J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972), pp. 259–273.
[2] J. M. Eisenberg and W. Greiner, *Excitation Mechanisms of the Nucleus* (North-Holland, Amsterdam, 1970), pp. 141–239.
[3] Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. I, pp. 211–212.
[4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), pp. 504–554.
[5] P. Martin and J. Zamudio-Cristi, J. Math. Phys. **23**, 2276 (1982).
[6] P. Martin and A. L. Guerrero, Scientia (Valparaiso, Chile) **XLVII(159)**, 123 (1982); Proc. Int. Conf. Plasma Phys. **1**, 325 (1982).

Projection operator techniques in physics

Wesley E. Brittin
Department of Physics, University of Colorado, Boulder, Colorado 80309
Arthur Y. Sakakura
Department of Physics, Colorado School of Mines, Golden, Colorado 80401

(Received 24 February 1983; accepted for publication 29 July 1983)

A systematic account of projection operators (projectors) and orthogonalization techniques together with applications to selected areas of physics is presented. This unified approach is shown to have advantages over other approaches in that the mathematical statements are more precise. The mathematical level, however, is aimed at the practicing physicist and lies between rigorous mathematics and current use in physics. Further, the techniques presented have practical applications as is demonstrated by examples in the quantum theory of measurement, in the relationship between second quantization and configuration space techniques, and in an account of generalized Wannier and Bloch functions. Attention is paid to the problem of construction of orthogonal projection operators (orthogonal projectors). The construction of orthogonal projectors even in approximate form would allow the solution of many practical problems ranging from the eigenvalue spectrum problem to the construction of states for many-body systems. One can almost say that *any* problem in quantum mechanics can be formulated as a problem involving the construction of projectors.

PACS numbers: 02.30.Tb, 03.65.Bz

I. INTRODUCTION

Projection operators (projectors) are useful in vast areas of physics—in quantum theory, many-body physics, applications of group theory, and in geometric theories—to name a few. We present what is hoped to be a unified view of projectors with selected applications. We first discuss projectors in finite-dimensional vector spaces and then orthogonal projectors in separable Hilbert spaces. The level of presentation lies somewhere between rigorous mathematics and customary practice in physics. Next we discuss methods of construction of projectors and of orthogonal sets of vectors. This subject is very large and has been studied extensively, especially by those interested in nuclear physics and in quantum chemistry. Our treatment unifies many of the approaches and also makes an attempt to present arguments which are clear and precise from the mathematical point of view. For a good review of background material for this section we refer to Jørgensen.[1]

The next three sections are given to applications, first to the quantum theory of measurement, then to the relation between second quantization and configuration space, and finally to generalized Wannier and Bloch functions.

Although the examples presented here cover only a small selection of the possible uses of projectors in physics, it is felt that the selection provides a fair coverge of their use and of general techniques in the field.

II. PROJECTORS IN FINITE-DIMENSIONAL VECTOR SPACES

Let V_n be a finite-dimensional vector space over the field F of real or complex numbers. The index n is the dimension of V_n. Since in most physical applications one has a metric or a pseudometric (relativity theory), we assume that V_n is so equipped. We also assume a norm $\|\cdot\|$ on V_n (for finite n, all norms are equivalent). That is, for all x in V_n, $\|x\|$ is the norm of x. Linear operators $\hat{A}$ acting on vectors in V_n are then given a norm, also denoted by $\|\cdot\|$, defined by

$$\|\hat{A}\| \equiv \max_{\|x\|\leq 1} \|\hat{A}x\|. \tag{2.1}$$

A linear operator $\hat{P}$ acting in V_n for which

$$\hat{P}^2 = \hat{P} \tag{2.2}$$

is called a projection operator on V_n, or simply a projector. Since[2,3] for any pair $\hat{A}$, $\hat{B}$ of linear operators on V_n, $\|\hat{A}\hat{B}\| \leq \|\hat{A}\|\|\hat{B}\|$,

$$\|\hat{P}\| = \|\hat{P}^2\| \leq \|\hat{P}\|^2, \tag{2.3}$$

so if $\hat{P} \neq 0$, $1 \leq \|\hat{P}\|$. Thus the norm of a nonzero projector is never less than unity.

If $\hat{P}$ is a projector on V_n so is $\hat{P}' \equiv \hat{1} - \hat{P}$, where $\hat{1}$ is the unit operator on V_n. For all $x \in V_n$, we have the unique decomposition

$$x = \hat{P}x + (\hat{1} - \hat{P})x \tag{2.4}$$

$$= x_{\hat{P}} + x_{\hat{P}'}, \tag{2.5}$$

where

$$x_{\hat{P}} \equiv \hat{P}x \tag{2.6}$$

and

$$x_{\hat{P}'} \equiv \hat{P}'x = (\hat{1} - \hat{P})x. \tag{2.7}$$

We note that $\hat{P}x_{\hat{P}} = x_{\hat{P}}$ and $\hat{P}x_{\hat{P}'} = \hat{P}(\hat{1} - \hat{P})x = 0$. The vector space V_n is thereby decomposed into the direct sum

$$V_n = V_{\hat{P}} \oplus V_{\hat{P}'}, \tag{2.8}$$

where $V_{\hat{P}} = \hat{P}V_n$ and $V_{\hat{P}'} = \hat{P}'V_n = (\hat{1} - \hat{P})V_n$. The subspace $V_{\hat{P}}$ is called the *range* of $\hat{P}$, and the subspace $V_{\hat{P}'}$ is called the *null space* of $\hat{P}$.

 0022-2488/84/051274-22$02.50

Conversely, if V_n is the direct sum of the subspace V_1 and V_2,

$$V_n = V_1 \oplus V_2, \tag{2.9}$$

then we may introduce unique projectors $\hat{P}_1, \hat{P}_2$ with properties

$$V_1 = \hat{P}_1 V_n, \quad V_2 = \hat{P}_2 V_n,$$

and

$$\hat{1} = \hat{P}_1 + \hat{P}_2. \tag{2.10}$$

We point out that the norm $\|\hat{P}\|$ for a nonzero projector $\hat{P}$ may be arbitrarily large for spaces having dimension larger than unity. As an example, we consider a two-dimensional unitary space with an orthonormal basis $\{e_1, f_1\}$. Define an operator $\hat{P}$ by its action on $\{e_1, f_1\}$:

$$\hat{P}e_1 = e_1, \tag{2.11}$$

$$\hat{P}f_1 = Ae_1, \tag{2.12}$$

where A is a number. Clearly, $\hat{P}^2 e_1 = e_1$ and $\hat{P}^2 f_1 = Ae_1$, so $\hat{P}^2 = \hat{P}$. Hence $\hat{P}$ is a projector. Further, $|A|$ may be chosen as large as we please, and, since $\|\hat{P}\| = |A|$, the norm may be arbitrarily large. If we have an infinite-dimensional vector space, a projector may even be an unbounded operator. As an example of the latter, we consider a separable Hilbert space with a basis $B = \{e_1, f_1, e_2, f_2, \cdots\}$. We define $\hat{P}$ by its action on B:

$$\hat{P}e_l = e_l, \quad l = 1,2,3,\cdots,$$

$$\hat{P}f_l = A_l e_l, \quad l = 1,2,3,\cdots,$$

and extend $\hat{P}$ to its domain of definition by linearity. If $|A_l|$ are chosen to grow sufficiently rapidly with increasing l, $\hat{P}$ will be an unbounded projection operator.

A *pseudometric* g on V_n is a nondegenerate Hermitian sesquilinear form defined for every pair x,y of vectors in V_n with the properties

$$g(x,y) = \overline{g(y,x)}, \quad \forall\, x,y \in V_n \tag{2.13}$$

(the bar indicates complex conjugation if F is the complex number field),

$$g(x,y+z) = g(x,y) + g(x,z), \quad \forall\, x,y,z \in V_n, \tag{2.14}$$

$$g(x,ay) = ag(x,y), \quad \forall a \in F, \quad x,y \in V_n. \tag{2.15}$$

If $g(x,y) = 0$, $\forall y \in V_n$, then $x = 0$. From Eq. (2.13),

$$g(x,x) = \overline{g(x,x)} \tag{2.16}$$

so $g(x,x)$ is real. If in addition $g(x,x) \geqslant 0$, g is called a *metric*. Thus a pseudometric (p-metric for short) is a special type of second-order tensor defined on V_n.

Let $B = \{e_k\}_{k=1}^n$ be a basis for V_n. Then we define the components g_{rs} of g relative to B by

$$g_{rs} = g(e_r, e_s). \tag{2.17}$$

If $B' = \{e_{k'}\}_{k'=1}^n$ is another basis for V_n, we may express $e_{k'}$ in terms of e_k by means of the linear relation

$$e_{k'} = e_k A^k{}_{k'}. \tag{2.18}$$

(We adopt a notation similar to that used by Misner, Thorn, and Wheeler.[4]) Equations (2.18) may be inverted to yield

$$e_k = e_{k'} A^{k'}{}_k, \tag{2.19}$$

where

$$A^k{}_{k'} A^{k'}{}_l = \delta^k{}_l \tag{2.20}$$

and

$$A^{k'}{}_k A^k{}_{l'} = \delta^{k'}{}_{l'}. \tag{2.21}$$

In Eqs. (2.18), (2.19), (2.20), and (2.21) we use the Einstein summation convention—repeated indices indicate summation from 1 to n. The notation adopted here may appear somewhat unusual. However, it is the most foolproof and automatic notation that has appeared to date. One minor point is that matrix elements $A^k{}_{k'}$ for equal k and k' can be a source of some confusion, unless we agree to write, for example, $A^2{}_{2'}$ and never $A^2{}_2$.

The p-metric components $g_{k'l'}$ in the basis B' are given by

$$g_{k'l'} = g(e_{k'}, e_{l'}) = g(e_k A^k{}_{k'}, e_l A^l{}_{l'}) = \overline{A}^k{}_{k'} g_{kl} A^l{}_{l'}. \tag{2.22}$$

We introduce the $n \times n$ matrices

$$(g) = (g_{kl}), \quad (g') = (g_{k'l'}) \tag{2.23}$$

and

$$(A) = (A^k{}_{k'}). \tag{2.24}$$

Then Eq. (2.22) may be written in matrix form as

$$(g') = (A)^\dagger (g)(A), \tag{2.25}$$

where $(A)^\dagger$ is the transposed complex conjugate (Hermitian conjugate) of the matrix (A). Since (g) is Hermitian, there exists a unitary matrix (v) such that

$$(v)^\dagger (g)(v) = \begin{pmatrix} \lambda^2{}_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda^2{}_r & & & & \\ & & & 0 & & & \\ & & & & -\mu^2{}_1 & & \\ & & & & & -\mu^2{}_2 & \\ & & & & & & \ddots \\ & & & & & & -\mu^2{}_s \end{pmatrix}. \tag{2.26}$$

That is $(v)^\dagger (g)(v)$ is diagonal with the λ's and μ's taken to be positive (since (g) is nonsingular). Let (Δ) be defined by

$$(\Delta) \equiv \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & 0 \\ & & \lambda_r & & & \\ & & & \mu_1 & & \\ & & & & \ddots & \\ 0 & & & & & \mu_s \end{pmatrix}. \tag{2.27}$$

Then

$$(g') = (A)^{+}(g)(A)$$

$$= \begin{pmatrix} 1 & & & & & & \\ & 1 & & r & & & \\ & & \ddots & & & & 0 \\ & & & 1 & & & \\ & & & & -1 & & \\ & & & & & -1 & \\ & 0 & & & & & \ddots \\ & & & & & & & -1 \end{pmatrix} \tag{2.28}$$

$$\equiv(\eta),$$

where $(A)\equiv(v)(\Delta)^{-1}$, and $\eta_{k'l'} = \epsilon_{(k')}\delta(k'l')$, (no sum) with $\epsilon_{(k')} = 1$, $k' = 1',...,r'$, and $\epsilon_{(k')} = -1$, $k' = r'+1',...,n'$. The trace of (η), $\mathrm{tr}(\eta) = r - s$ is called the *signature* of g. The form $g(x,y)$ may be regarded as an indefinite inner product on V_n and is written variously as

$$g(x,y) = \langle x,y\rangle = x^{*}\cdot y. \tag{2.30}$$

We will use the notation $\langle x,y\rangle$ which conforms most closely with the widely used Dirac notation.[5] When dealing with linear operators and matrices relating to indefinite metric spaces, care must be exercised. For example, the dual or adjoint A^{*} of a linear operator A in V_n is defined by means of

$$\langle A^{*}x,y\rangle \equiv \langle x,Ay\rangle, \quad \forall x,y\in V_n, \tag{2.31}$$

whereas we always define the matrix elements $A^{l}{}_{m}$, $A^{*l}{}_{m}$ in a given basis $B = \{e_k\}_{k=1}^{n}$ by means of

$$Ae_m \equiv e_l A^{l}{}_{m} \tag{2.32}$$

and

$$A^{*}e_m \equiv e_l A^{*l}{}_{m} \tag{2.33}$$

(l refers to row and m to column). Matrix element calculations are facilitated by use of the dual basis $B^{*}\equiv\{e^{*k}\}_{k=1}^{n}$ to B, where

$$\langle e^{*k}, e_l\rangle = \delta^{k}{}_{l}, \tag{2.34}$$

which always exists and is unique. Thus

$$A^{l}{}_{m} = \langle e^{*l}, Ae_m\rangle \tag{2.35}$$

and

$$A^{*l}{}_{m} = \langle e^{*l}, A^{*}e_m\rangle. \tag{2.36}$$

We define g^{kl} by

$$e^{*k} = e_l g^{lk}, \tag{2.37}$$

from which

$$\delta^{k}{}_{m} = \langle e^{*k}, e_m\rangle = \langle e_l g^{lk}, e_m\rangle = \bar{g}^{lk} g_{lm}, \tag{2.38}$$

or since $\bar{g}_{lm} = g_{ml}$,

$$\delta^{k}{}_{m} = g^{lk}\bar{g}_{lm} = g^{lk}g_{ml} = g_{ml}g^{lk}, \tag{2.39}$$

or as matrices

$$(g_{ml})(g^{ml}) = (g^{ml})(g_{ml}) = 1, \tag{2.40}$$

i.e.,

$$g^{kl}g_{lm} = g_{ml}g^{lk} = \delta^{k}{}_{m}. \tag{2.41}$$

Further, $\bar{g}^{lm} = g^{ml}$, so (g^{lm}) is a Hermitian matrix. The above results may now be used to find the relation between the matrix elements of an operator and those of its dual. From Eqs. (2.36)

$$\begin{aligned} A^{*l}{}_{m} &= \langle e^{*l}, A^{*}e_m\rangle \\ &= \langle e_s g^{sl}, A^{*}e_m\rangle \\ &= \bar{g}^{sl}\langle Ae_s, e_m\rangle \\ &= g^{ls}\langle e_t\bar{A}^{t}{}_{s}, e_m\rangle \\ &= g^{ls}\bar{A}^{t}{}_{s}g_{tm}, \end{aligned} \tag{2.42}$$

or in matrix form

$$(A^{*}) = (g)^{-1}(A)^{\dagger}(g). \tag{2.43}$$

It follows that $(A^{*}) = (A)^{\dagger}$, in general, if and only if $(g) = \lambda(1)$ where λ is a scalar and (1) is the unit matrix.

Equation (2.29) shows that there exists a basis $B_L = \{e_k\}_{k=1}^{n}$, for which $\langle e_k, e_l\rangle = \eta_{kl}$. We may call B_L a Lorentz basis. If L is a linear operator which maps one Lorentz basis B_L onto another $B_{L'}$,

$$e_{k'} = Le_k = e_r L^{r}{}_{k'}, \tag{2.44}$$

then

$$\begin{aligned} \eta_{k'l'} &= \langle e_{k'}, e_{l'}\rangle \\ &= \langle Le_k, Le_l\rangle \\ &= \langle e_k, L^{*}Le_l\rangle = \langle e_k, e_l\rangle = \eta_{kl}, \end{aligned} \tag{2.45}$$

or

$$L^{*}L = LL^{*} = \hat{1}, \tag{2.46}$$

where $\hat{1}$ is the unit operator on V_n.

The above equation may not appear to be familiar. It says that L is *pseudounitary*. In terms of matrix elements in a Lorentz basis

$$(L^{*})(L) = (\eta)^{-1}(L)^{\dagger}(\eta)(L) = 1 \tag{2.47}$$

or, since $(\eta)^{-1} = (\eta)$, by Eq. (2.43),

$$(\eta)(L)^{\dagger}(\eta) = (L)^{-1}, \tag{2.48}$$

or, since $(\eta)^2 = (\eta)$,

$$(L)^{\dagger}(\eta)(L) = (\eta), \tag{2.49}$$

which is familiar from relativity where $(L)^{\dagger}$ is the transpose of the real matrix (L). Before returning to projectors, we show the relationship between matrix elements of an operator A in a given basis B and those of the dual A^{*} in the dual basis B^{*}. Let us apply A^{*} to elements e^{*m} of B^{*}, and define new matrix elements $A^{*}{}_{l}{}^{m}$ in the basis B^{*} by means of

$$A^{*}e^{*m} \equiv e^{*l}A^{*}{}_{l}{}^{m}. \tag{2.50}$$

Now

$$\langle A^{*}e^{*m}, e_s\rangle = \langle e^{*m}, e_t A^{t}{}_{s}\rangle,$$

so

$$\langle e^{*l}A^{*}{}_{l}{}^{m}, e_s\rangle = \langle e^{*m}, e_t A^{t}{}_{s}\rangle, \tag{2.51}$$

and therefore

$$\bar{A}^{*m}_{s} = A^{m}{}_{s}$$

or

$$A^{*m}_{s} = \bar{A}^{m}{}_{s}. \tag{2.52}$$

Equation (2.52) states that the matrix of A^{*} in the dual basis

B^* is just the Hermitian conjugate of the matrix of A in the basis B. (Recall that indices to the left refer to *rows* of a matrix and indices to the right to *columns* of a matrix).

If $\hat{P}$ is a projector on V_n, so is the dual $\hat{P}^*$, since $\hat{P}^2 = \hat{P}$, and

$$\langle x,\hat{P}^2 y\rangle = \langle x,\hat{P}y\rangle = \langle \hat{P}^* x,y\rangle$$
$$= \langle \hat{P}^* x,\hat{P}y\rangle = \langle (\hat{P}^*)^2 x,y\rangle, \quad \forall x,y\in V_n.$$

Therefore $(\hat{P}^*)^2 = \hat{P}^*$ is a projector.

Given a projector $\hat{P}$ on V_n, we may choose a basis B adopted to the decomposition

$$V_n = \hat{P}V_n \oplus (\hat{1} - \hat{P})V_n. \tag{2.53}$$

That is, we let $e_1, e_2,...,e_p$, $p\leqslant n$, span $\hat{P}V_n$ and $e_{p+1},...,e_n$ span $(1-\hat{P})V_n$. This means that

$$\hat{P}e_k = \begin{cases} e_k, & k = 1,...,p, \\ 0, & k = p+1,...,n, \end{cases} \tag{2.54}$$

or in the matrix form

$$(\hat{P}) = (P^l{}_m) = \begin{pmatrix} 1_p & 0 \\ 0 & 0_{n-p} \end{pmatrix}, \tag{2.55}$$

where l_p is the diagonal matrix having p ones on the diagonal and 0_{n-p} is $(n-p)\times(n-p)$ zero matrix. A basis B for which $\hat{P}$ is a diagonal with ones or zeros on the diagonal is called a *canonical* basis.

If $x\in V_n$, then in a canonical basis

$$\hat{P}x = \hat{P}x^l e_l = x^l \hat{P}e_l$$
$$= \sum_{l=1}^{p} x^l e_l \tag{2.56}$$
$$= \sum_{l=1}^{p} e_l \langle e^{*l},x\rangle, \tag{2.57}$$

where $\{e^{*l}\}_{l=1}^n = B^*$ is the dual to B.

If x,y are any two vectors in V_n we define a linear operator $x\otimes y$ by means of the definition:

$$(x\otimes y)z \equiv x\langle y,z\rangle. \tag{2.58}$$

With the latter definition Eq. (2.57) may be written

$$\hat{P}x = \sum_{l=1}^{p} (e_l \otimes e^{*l})x \tag{2.59}$$

and since x is arbitrary we have a canonical form for $\hat{P}$

$$\hat{P} = \sum_{l=1}^{p} (e_l \otimes e^{*l}). \tag{2.60}$$

The adjoint (dual) $(x\otimes y)^*$ of the operator $(x\otimes y)$ is established directly to be given by

$$(x\otimes y)^* = y\otimes x, \tag{2.61}$$

so that Eq. (2.60) yields for $\hat{P}^*$:

$$\hat{P}^* = \sum_{l=1}^{p} e^{*l}\otimes e_l. \tag{2.62}$$

A canonical representation [Eq. (2.60)] is clearly not unique, since we may choose other bases for $\hat{P}V_n$ and $(\hat{1}-\hat{P})V_n$, and these give the same $\hat{P}$ via Eq. (2.60).

A canonical representation via the basis $B = \{e_k\}_{k=1}^n$ may be used to form another representation of $\hat{P}$. Introduce a square root $\hat{E}$ of unity by means of

$$\hat{E}e_k = \begin{cases} e_k, & k = 1,...,p, \\ -e_k, & k = p+1,...,n, \end{cases} \tag{2.63}$$

that is, $\hat{E} = 2\hat{P} - \hat{1}$, and

$$\hat{P} = \tfrac{1}{2}(\hat{1} + \hat{E}), \tag{2.64}$$

$\hat{E}^2 = \hat{1}$. Conversely, if $\hat{E}^2 = \hat{1}$, then $\hat{P}$ given by Eq. (2.64) is a projector. A widely used version of Eq. (2.64) occurs in two dimensions. If $\sigma = (\sigma_1,\sigma_2,\sigma_3)$, where σ_k are the 2×2 Pauli matrices, then

$$\hat{P} = \tfrac{1}{2}(\hat{1} + \mathbf{a}\cdot\sigma), \tag{2.65}$$

where $\mathbf{a}^2 \equiv a_1^2 + a_2^2 + a_3^2 = 1$, $a_k\in C$, is a complex 2×2 projection matrix. Furthermore, every 2×2 complex projection matrix except the unit and zero 2×2 projectors can be so represented. Equation (2.65) can easily be generalized to n-dimensional matrices or operators. Let $\hat{\gamma}_k$, $k = 1, ..., n$, be n Dirac type operators in V_n satisfying only the requirement

$$\{\hat{\gamma}_k,\hat{\gamma}_l\} \equiv \hat{\gamma}_k\hat{\gamma}_l + \hat{\gamma}_l\hat{\gamma}_k = 2\delta_{kl}. \tag{2.66}$$

Then $\hat{P}$ defined by

$$\hat{P} = \tfrac{1}{2}(\hat{1} + a^k\hat{\gamma}_k) \tag{2.67}$$

with $\Sigma_{k=1}^n (a^k)^2 = 1$, $a^k\in C$, is a projector.

We consider now the possibility of comparing two projectors $\hat{P}_1$, $\hat{P}_2$ in V_n. First, if the range of $\hat{P}_1$ has the same dimension as the range of $\hat{P}_2$, we say that $\hat{P}_1$ and $\hat{P}_2$ are *equivalent* or *similar*.

Let B_1 be a canonical basis for $\hat{P}_1$ and B_2 a canonical basis for $\hat{P}_2$. The bases $B_1 = \{e_{1l}\}_{l=1}^n$ and $B_2 = \{e_{2l}\}_{l=1}^n$ are related by the nonsingular mapping $\hat{\Omega}$:

$$e_{1l} \to e_{2l} = \hat{\Omega}e_{1l} = e_{1m}\Omega^m{}_l, \tag{2.68}$$

and the bases B_1^*, B_2^* dual to B_1 and B_2 are related by

$$e_2^{*l} = \hat{\Omega}^{*-1} e_1^{*l} \equiv e_1^{*m}\Omega^{*-1}{}_m{}^l. \tag{2.69}$$

The relationship between $\hat{P}_1$ and $\hat{P}_2$ follows:

$$\hat{P}_2 = \sum_{l=1}^{p} e_{2l}\otimes e_2^{*l} = \sum_{l=1}^{p} (\hat{\Omega}e_{1l})\otimes(\hat{\Omega}^{*-1}e_1^{*l})$$
$$= \hat{\Omega}\left(\sum_{l=1}^{p} e_{1l}\otimes e_1^{*l}\right)\hat{\Omega}^{-1}, \tag{2.70}$$

or

$$\hat{P}_2 = \hat{\Omega}\hat{P}_1\hat{\Omega}^{-1}, \tag{2.71}$$

which is the usual relationship between *similar* operators. We now pre- and postmultiply Eq. (2.71) by $\hat{P}_2$ to find

$$\hat{P}_2 = \hat{\Omega}_{21}\hat{\Omega}_{12} \tag{2.72}$$

where $\hat{\Omega}_{12}$ and $\hat{\Omega}_{21}$ are defined by

$$\hat{\Omega}_{12} = \hat{P}_1\hat{\Omega}^{-1}\hat{P}_2 \tag{2.73}$$

and

$$\Omega_{21} = \hat{P}_2\hat{\Omega}\hat{P}_1. \tag{2.74}$$

Similarly, we find that

$$\hat{P}_1 = \hat{\Omega}_{12}\hat{\Omega}_{21}. \tag{2.75}$$

The operator Ω_{12} maps the range of $\hat{P}_2$ onto the range of $\hat{P}_1$ and the null space of $\hat{P}_2$ into zero, while $\hat{\Omega}_{21}$ maps the range of $\hat{P}_1$ onto the range of $\hat{P}_2$ and the null space of $\hat{P}_1$ into zero. The *factorizations* (2.72) and (2.75) are very useful in applica-

tions, especially in Hilbert space, when dealing with orthogonal projectors.

In general, if one has a nontrivial factorization of a projector

$$\hat{P} = \hat{A}\hat{B}, \tag{2.76}$$

then $\hat{B}\hat{A}$ is not a projector, as the following counterexample illustrates. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} \end{pmatrix}$. Then $AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, which is a projection matrix. However, $BA = \begin{pmatrix} -\frac{2}{4} & -\frac{2}{4} \\ -\frac{2}{4} & -\frac{2}{4} \end{pmatrix}$, which is *not* a projection matrix.

If

$$\hat{P}_1 = \hat{\Omega}^*\hat{\Omega}, \tag{2.77}$$

so that $\hat{P}_1$ is pseudo-self-adjoint: $\hat{P}_1^* = \hat{\Omega}^*(\hat{\Omega}^*)^* = \hat{\Omega}^*\hat{\Omega} = \hat{P}_1$, then $\hat{P}_2 = \hat{\Omega}\hat{\Omega}^*$ is not necessarily a projector. We prove this by exhibiting a counterexample. Let g be nondefinite and choose a basis for which g is diagonal, so Eq. (2.72) in matrix form is

$$(P_1) = (g)(\Omega)^\dagger(g)(\Omega) = (\Omega^*)(\Omega). \tag{2.78}$$

Let the nondefinite $(g) = \begin{pmatrix} 1 & \\ & -1' \end{pmatrix}$, where 1 is the $s\times s$ unit matrix and 1′ is the $t\times t$ unit matrix ($s+t=n$). Write $(\Omega) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ where A, B, C, D are matrices adopted to the $\begin{pmatrix} 1 & \\ & -1' \end{pmatrix}$ decomposition of (g). Choose the matrices B, C, D such that

$$A^\dagger A = C^\dagger C, \quad A^\dagger B = C^\dagger D, \quad D^\dagger D = B^\dagger B, \tag{2.79}$$

then

$$(\Omega^*)(\Omega) = \begin{pmatrix} A^\dagger A - C^\dagger C & A^\dagger B - C^\dagger D \\ -BA^\dagger + D^\dagger C & D^\dagger D - B^\dagger B \end{pmatrix} = 0. \tag{2.80}$$

Next

$$(P_2) \equiv (\Omega)(\Omega^*) = \begin{pmatrix} I & J \\ -J^\dagger & K \end{pmatrix}, \tag{2.81}$$

where

$$I = AA^\dagger - BB^\dagger, \quad J = BD^\dagger - AC^\dagger, \quad K = DD^\dagger - CC^\dagger. \tag{2.82}$$

We choose A such that $AA^\dagger - BB^\dagger \neq 0$; then $(P_2) \neq 0$. But

$$(P_2)^2 = \begin{pmatrix} I^2 - JJ^\dagger & IJ + JK \\ -J^\dagger I - KJ^\dagger & K^2 - J^\dagger J \end{pmatrix} = 0 \neq (P_2), \tag{2.83}$$

as can be readily shown by using Eqs. (2.82). Hence $\hat{P}_2$ is not a projector.

Suppose now that $\hat{\Omega}^*\hat{\Omega} = \hat{P}_1$ is a *nonzero* projector. What can be said about $\hat{P}_2 \equiv \hat{\Omega}\hat{\Omega}^*$? Certainly, $\hat{P}_2^2 = \hat{P}_2^3 = \hat{P}_2^4 = \ldots$, so $\hat{P}_2^2$ is a projector. Further $(\hat{P}_2 - \hat{P}_2^2)^2 = 0$ and $(\hat{P}_2 - \hat{P}_2^2)^* = (\hat{P}_2 - \hat{P}_2^2)$, so

$$g((\hat{P}_2 - \hat{P}_2^2)x, \quad (\hat{P}_2 - \hat{P}_2^2)x) \equiv 0, \tag{2.84}$$

and, therefore, if g is definite, $(\hat{P}_2 - \hat{P}_2^2)x = 0$ all $x \in V_n$ so $\hat{P}_2 = \hat{P}_2^2$. This holds also for the case $\hat{P}_1 = 0$ for definite g.

We have not investigated this question beyond dim $V_n = 2$. For 2 dimensions, g-nondefinite, we have the theorem that $\hat{\Omega}^*\hat{\Omega} = \hat{P}_1 = \hat{P}_1^2 \neq 0$ implies $\hat{\Omega}\hat{\Omega}^* \equiv \hat{P}_2$ is a projector. We prove this result using 2×2 matrices and choosing a basis such that $(g) = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} = \sigma_3$. We then use the fact that any nonzero matrix projector except unity (where the theorem is trivially valid) can be expressed as

$$\hat{P}_1 = (1 + \mathbf{a}\cdot\boldsymbol{\sigma})/2, \tag{2.85}$$

where a_1, a_2, a_3 are complex and $a_1^2 + a_2^2 + a_3^2 = 1$. Write

$$\hat{\Omega} = \Omega_0 + \boldsymbol{\Omega}\cdot\boldsymbol{\sigma},$$

where Ω_0, Ω_1, Ω_2, Ω_3 are complex numbers. Then

$$\hat{\Omega}^* = \sigma_3\hat{\Omega}^\dagger\sigma_3 = \bar{\Omega}_0 + \tilde{\boldsymbol{\Omega}}\cdot\boldsymbol{\sigma}, \tag{2.86}$$

where $\tilde{\Omega}_1 = -\bar{\Omega}_1$, $\tilde{\Omega}_2 = -\bar{\Omega}_2$, $\tilde{\Omega}_3 = \bar{\Omega}_3$, so that

$$\hat{P}_1 = \hat{\Omega}^*\hat{\Omega} = \tfrac{1}{2} + \tfrac{1}{2}\mathbf{a}\cdot\boldsymbol{\sigma} \tag{2.87}$$

yields

$$\bar{\Omega}_0\Omega_0 + \tilde{\boldsymbol{\Omega}}\cdot\boldsymbol{\Omega} = \tfrac{1}{2},$$
$$\tilde{\boldsymbol{\Omega}}\Omega_0 + \Omega_0\boldsymbol{\Omega} + i\tilde{\boldsymbol{\Omega}}\times\boldsymbol{\Omega} \equiv \mathbf{A} + i\mathbf{B} = \mathbf{a}/2, \tag{2.88}$$

where $\mathbf{A} = \tilde{\boldsymbol{\Omega}}\Omega_0 + \Omega_0\boldsymbol{\Omega}$, $\mathbf{B} = \tilde{\boldsymbol{\Omega}}\times\boldsymbol{\Omega}$.

The condition $\hat{P}_1^2 = \hat{P}_1$ then is expressed as

$$\mathbf{A}\cdot\mathbf{A} - \mathbf{B}\cdot\mathbf{B} = \tfrac{1}{4} \tag{2.89}$$

($\mathbf{A}\cdot\mathbf{B} \equiv 0$). The expression for $\hat{P}_2 = \hat{\Omega}\hat{\Omega}^*$ is then calculated and found to be

$$\hat{P}_2 = \hat{\Omega}\hat{\Omega}^* \equiv \tfrac{1}{2} + (\mathbf{b}\cdot\boldsymbol{\sigma})/2, \tag{2.90}$$

where $\mathbf{b}/2 \equiv \mathbf{A} - i\mathbf{B}$. We find

$$\mathbf{b}^2/4 = \mathbf{A}\cdot\mathbf{A} - \mathbf{B}\cdot\mathbf{B} = \tfrac{1}{4} \tag{2.91}$$

or $\mathbf{b}^2 = 1$, and hence $\hat{P}_2 = \hat{\Omega}\hat{\Omega}^*$ is a projector. We conjecture that this result holds for V_n, $n \geqslant 2$.

A sufficient condition that $\hat{P}_1$ and $\hat{P}_2$ be equivalent is given by a theorem due to Kato,[6] which we state without proof: If $(\hat{1} - \hat{R})^{-1/2}$ exists, where $\hat{R} \equiv (\hat{P}_1 - \hat{P}_2)^2$, then $\hat{P}_1$ and $\hat{P}_2$ are similar. The inverse square root $(\hat{1} - \hat{R})^{-1/2}$ exists if $\|\hat{R}\| < 1$, which in turn is valid if $\|\hat{P}_1 - \hat{P}_2\| < 1$, since $\|\hat{R}\| = \|(\hat{P}_1 - \hat{P}_2)^2\| \leqslant \|\hat{P}_1 - \hat{P}_2\|^2$. Kato's theorem holds in infinite-dimensional Banach spaces, so in particular it is valid for Hilbert spaces. For orthogonal projectors in Hilbert space H the norm $\|\hat{P}_1 - \hat{P}_2\|$ is called the *aperture*[7] of the closed subspaces $\hat{P}_1 H$ and $\hat{P}_2 H$. In Ref. 7 it is shown that $\|\hat{P}_1 - \hat{P}_2\| < 1$ is a sufficient condition that $\dim(\hat{P}_1 H) = \dim(\hat{P}_2 H)$. That is, if the aperture is less than unity, the two projectors are equivalent, which is a special case of Kato's theorem.

We conclude this general introduction to projectors with some comments concerning certain combinations of projectors which are themselves projectors. We work in a general n-dimensional, $n < \infty$, space V_n, although most of the results extend to infinite-dimensional vector space.

If $\hat{P}_1$ and $\hat{P}_2$ are projectors and $\hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1$, then $\hat{P}_1\hat{P}_2$ is a projector.

If $\hat{P}_1$ and $\hat{P}_2$ are projectors and $\hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1 = \hat{0}$, then $\hat{P}_1 + \hat{P}_2$ is a projector.

If $\hat{P}_1$ and $\hat{P}_2$ are projectors and $\hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1 = \hat{P}_1$, then $\hat{P}_2 - \hat{P}_1$ is a projector, and we say that $\hat{P}_1$ is a subprojector of $\hat{P}_2$ and write $\hat{P}_1 < \hat{P}_2$. Note that $\hat{P} < \hat{P}$ for any projector. If $\hat{P}_1 < \hat{P}_2$ we also state that $\hat{P}_1$ is *less than or equal to* $\hat{P}_2$. The range of $\hat{P}_1$ is a subspace of the range of $\hat{P}_2$, and "$<$" is a partial ordering on the set of all projectors on V_n.

Given $\hat{P}_1 < \hat{P}$, then we may write

$$\hat{P} = \hat{P}_1 + \hat{P}_2,$$

where $\hat{P}_2$ is the projector $\hat{P} - \hat{P}_1$. We see that $\hat{P}_1 \hat{P}_2 = \hat{P}_1 (\hat{P} - \hat{P}_1) = \hat{P}_1 - \hat{P}_1 = \hat{0}$ and similarly $\hat{P}_2 \hat{P}_1 = \hat{0}$.

The set of all projectors on V_n together with the partial order "$<$" forms a lattice with the lattice operations "$\vee$" and "$\wedge$" defined for every pair $\hat{P}_1, \hat{P}_2$ of projectors as

$$\hat{P}_1 \vee \hat{P}_2 \equiv \inf\{\hat{P},\ \hat{P}^2 = \hat{P},\ \hat{P} > \hat{P}_1,\ \hat{P} > \hat{P}_2\}, \tag{2.92}$$

$$\hat{P}_1 \wedge \hat{P}_2 \equiv \sup\{\hat{P},\ \hat{P}^2 = \hat{P},\ \hat{P} < \hat{P}_1,\ \hat{P} < \hat{P}_2\}. \tag{2.93}$$

The existence of $\hat{P}_1 \vee \hat{P}_2$ and $\hat{P}_1 \wedge \hat{P}_2$ is assured since $\hat{0} < \hat{P} < \hat{1}$ for every projector. The terms *least upper bound* (lub), resp., *greatest lower bound* (glb), are given to $\hat{P}_1 \vee \hat{P}_2$, resp. $\hat{P}_1 \wedge \hat{P}_2$. The lub and glb operations apply to arbitrary collections of projectors on V_n. That is, if $\{\hat{P}_\alpha\}_{\alpha \in I}$ is a family of projectors on V_n, where I is some index set, then $\vee_\alpha \hat{P}_\alpha, \wedge_\alpha \hat{P}_\alpha$ are defined in a manner analogously to the definitions given by Eqs. (2.92) and (2.93). Such compound projectors have many applications such as to the theory of von Neumann algebras, and for I a countable set, to many-body physics. We shall be particularly interested in the construction of $\vee_{n \in I} \hat{P}_n$, $\wedge_{n \in I} \hat{P}_n$, where I is countable.

III. ORTHOGONAL PROJECTORS

We single out orthogonal projectors under their own heading because of their central use in physics, theoretical chemistry and in other areas. We deal here with a unitary vector space H which is complete, separable, and (usually) infinite-dimensional. That is H is a separable Hilbert space. In order to distinguish H from other spaces we have mentioned, we use the Dirac notation $\langle x|y\rangle$ for the inner product of two vectors x and y in H. We state the axioms for the inner product:

$$\langle x|y\rangle = \overline{\langle y|x\rangle}, \text{ (Hermitian symmetry)} \tag{3.1}$$

all $x, y \in H$;

$$\langle x|y_1 + y_2\rangle = \langle x|y_1\rangle + \langle x|y_2\rangle, \tag{3.2}$$

all $x_1, y_1, y_2 \in H$;

$$\langle x|ay\rangle = a\langle x|y\rangle, \tag{3.3}$$

all $x, y \in H$, $a \in C$;

$$\langle x|x\rangle \geqslant 0, \tag{3.4}$$

all $x \in H$; and

$$\langle x|x\rangle = 0 \text{ if and only if } x = 0. \tag{3.5}$$

The norm $\|x\|$ of a vector in H is given by $\|x\| = +\sqrt{\langle x|x\rangle}$. Separability means that there exists in H a dense countable set of vectors. This in turn is equivalent to the statement that there exists in H a countable orthonormal basis. If $B = \{e_i\}_{i=1}^{\infty}$ is such an orthonormal basis, we have

$$\langle e_i|e_j\rangle = \delta_{ij}, \tag{3.6}$$

and for every $x \in H$,

$$x = \sum_{i=1}^{\infty} x_i e_i, \tag{3.7}$$

with $x_i \equiv \langle e_i|x\rangle$, and with the sum converging in vector norm:

$$\lim_{N \to \infty} \left\| x - \sum_{i=1}^{N} x_i e_i \right\| = 0. \tag{3.8}$$

The expansion equation (3.7) is also written as a resolution of the unit operator $\hat{1}$ on H:

$$x = \hat{1}x = \sum_{i=1}^{\infty} (e_i \otimes e_i)x \equiv \sum_{i=1}^{\infty} |e_i\rangle\langle e_i|x, \tag{3.9}$$

where we have used the Dirac notation $|e_i\rangle\langle e_i|$ for the one-dimensional projector $e_i \otimes e_i$ defined earlier. The operator expansion

$$\hat{1} = \sum_{i=1}^{\infty} |e_i\rangle\langle e_i| \tag{3.10}$$

is understood as a *strong* operator limit which means that $\Sigma_{i=1}^{\infty} |e_i\rangle\langle e_i|x\rangle$ converges in vector norm as $N \to \infty$ to x, for all $x \in H$.

If S is any set of vectors in H, then $S^\perp$, the set of vectors orthogonal to all the vectors of S, is a *closed* subspace of H. In particular, if P is itself a closed subspace of H, then the set $P^\perp$ of all vectors orthogonal to those of P is a closed subspace of H, and H may be decomposed into the direct sum

$$H = P \oplus P^\perp. \tag{3.11}$$

Equation (3.11) means simply that every $x \in H$ may be expressed uniquely as

$$x = x_P + x_{P^\perp}, \tag{3.12}$$

where $x_P \in P$ and $x_{P^\perp} \in P^\perp$. Clearly $\langle x_P|x_{P^\perp}\rangle = 0$. Orthogonal projectors $\hat{P}$ and $\hat{P}^\perp$ are defined by

$$\hat{P}x = x_P,\ \hat{P}^\perp x = x_{P^\perp}, \tag{3.13}$$

and we have $\hat{1} = \hat{P} + \hat{P}^\perp$, $\hat{P}^2 = \hat{P} = \hat{P}^*$, and $(\hat{P}^\perp)^2 = (\hat{P}^\perp) = (\hat{P}^\perp)^*$, where $\hat{P}^*$ is the operator adjoint (dual) to $\hat{P}$:

$$\langle \hat{P}x|y\rangle \equiv \langle x|\hat{P}^* y\rangle. \tag{3.14}$$

That is orthogonal projectors (from now on we drop the adjective *orthogonal*, it being understood unless we explicitly state otherwise, i.e., *oblique* projector $\neq$ orthogonal projector) are characterized as being the self-adjoint idempotent operators on H.

Since

$$\|x\|^2 = \|\hat{P}x\|^2 + \|\hat{P}^\perp x\|^2, \tag{3.15}$$

it follows that

$$\|\hat{P}x\| \leqslant \|x\| \tag{3.16}$$

and, therefore, that $\|\hat{P}\| \leqslant 1$. However we proved quite generally in Sec. II that if $\hat{P} \neq 0$, $\|\hat{P}\| \geqslant 1$: therefore, a nonzero projector has norm unity, $\hat{P} \neq 0$, $\|\hat{P}\| = 1$.

Since the *orthogonal complement* $P^\perp$ to a given subspace P of H is uniquely determined by P, we have a one-to-one correspondence between projectors and closed subspaces of H. (For the general oblique case the range subspace does not determine a unique complementary subspace, and, hence, the given range subspace does not determine a unique complementary oblique projector.) If the closed subspace P_1 is contained in the closed subspace P_2 ($P_1 \subset P_2$), then the corresponding projectors $\hat{P}_1$, $\hat{P}_2$ stand in relation $\hat{P}_1 < \hat{P}_2$.

If $\hat{P}_1$ and $\hat{P}_2$ are projectors in H, then the necessary and sufficient condition that $\hat{P}_2 - \hat{P}_1$ be a projector is $\hat{P}_1 \hat{P}_2 = \hat{P}_1$.

Proof: If $\hat{P}_1\hat{P}_2 = \hat{P}_1$, then the adjoint gives $\hat{P}_2\hat{P}_1 = \hat{P}_1$, and $(\hat{P}_2 - \hat{P}_1)^2 = \hat{P}_2 + \hat{P}_1 - \hat{P}_2\hat{P}_1 - \hat{P}_1\hat{P}_2 = \hat{P}_2 - \hat{P}_1$.

Conversely, if $\hat{P}_2 - \hat{P}_1$ is a projector, then $(\hat{P}_2 - \hat{P}_1)^2 = (\hat{P}_2 - \hat{P}_1)$ yields

$$2\hat{P}_1 = \hat{P}_1\hat{P}_2 + \hat{P}_2\hat{P}_1.$$

If we multiply the above equation first on the right, then on the left, with $\hat{P}_2$ we find that

$$\hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1.$$

Further $2\hat{P}_1 = 2\hat{P}_1\hat{P}_2$ and, therefore, $\hat{P}_1 = \hat{P}_1\hat{P}_2 = \hat{P}_2\hat{P}_1$.

If we are given two projectors $\hat{P}_1$, $\hat{P}_2$ on H, then $\hat{P}_1\hat{P}_2$ is in general not a projector. However, $\hat{P}_1 \wedge \hat{P}_2 = \hat{P}_2 \wedge \hat{P}_1$ is a projector; the projector on the subspace (we now drop the adjective *closed* for subspaces, it being understood unless otherwise stated) $P_1 \cap P_2$, where P_1 and P_2 are the subspaces $\hat{P}_1 H$ and $\hat{P}_2 H$.

Similarly, if $\hat{P}_1$ and $\hat{P}_2$ are two projectors on H, $\hat{P}_1 + \hat{P}_2$ is not in general a projector. However, $\hat{P}_1 \vee \hat{P}_2$ is a projector; the projector on the closure of the set $P_1 + P_2$ ($P_1 + P_2$ is the set of all vectors in H of the form $x_1 + x_2$ with $x_1 \in P_1$ and $x_2 \in P_2$). The projectors $\hat{P}_1 \wedge \hat{P}_2$ and $\hat{P}_1 \vee \hat{P}_2$ are related by the relationship $\hat{P}_1 \wedge \hat{P}_2 = (\hat{P}_1^{\perp} \vee \hat{P}_2^{\perp})^{\perp}$.

If we are given two projectors $\hat{P}_1$ and $\hat{P}_2$ on H, one may be interested in the result of first applying $\hat{P}_2$ and then $\hat{P}_1$. We may ask for the projector on the closure of the set $\hat{P}_1\hat{P}_2 H$. That $\hat{P}_1\hat{P}_2 H$ is not in general a closed subspace is shown by the following example:

Let $\{e_n\}_{n=1}^{\infty}$ and $\{V_n\}_{n=1}^{\infty}$ be two infinite orthonormal sets in H which are orthogonal to each other, $\langle e_n | V_m \rangle = 0$. Define the orthonormal set $\{W_n\}_{n=1}^{\infty}$ by

$$W_n = (1 + 1/n^2)^{-1/2}[V_n + (1/n)e_n],$$

n = 1,2,3,···, and projectors $\hat{P}_1$ and $\hat{P}_2$ by

$$\hat{P}_1 = \sum_n |e_n\rangle\langle e_n|$$

and

$$\hat{P}_2 = \sum_n |W_n\rangle\langle W_n|.$$

Then

$$\hat{P}_1\hat{P}_2 = \sum_n \left(n + \frac{1}{n}\right)^{-1}\left(|e_n\rangle\langle V_n| + \frac{1}{n}|e_n\rangle\langle e_n|\right).$$

If now we set

$$Z_N \equiv \sum_{n=1}^{N} V_n,$$

then $\langle e_n | Z_N \rangle = 0$ and

$$\langle V_n | Z_N \rangle = \begin{cases} 1, & n < N, \\ 0, & n > N. \end{cases}$$

We now construct a Cauchy sequency $\{y_N\}_{N=1}^{\infty}$ of vectors in $\hat{P}_1\hat{P}_2 H$:

$$y_N \equiv \hat{P}_1\hat{P}_2 Z_N = \sum_{n=1}^{N}\left(n + \frac{1}{n}\right)^{-1} e_n.$$

Further, the limit of y of y_N as $N \to \infty$ is just

$$y = \sum_{n=1}^{\infty}\left(n + \frac{1}{n}\right)^{-1} e_n.$$

We next show that $y \notin \hat{P}_1\hat{P}_2 H$. Assume otherwise; $y = \hat{P}_1\hat{P}_2 z$ for some $z \in H$. Then,

$$y = \sum_n\left(n + \frac{1}{n}\right)^{-1} e_n = \hat{P}_1\hat{P}_2 z$$
$$= \sum_n\left(n + \frac{1}{n}\right)^{-1} e_n\left(\langle V_n | z\rangle + \frac{1}{n}\langle e_n | z\rangle\right),$$

and hence

$$\langle V_n | z\rangle + \frac{1}{n}\langle e_n | z\rangle = 1, \quad n = 1,2,3,\cdots.$$

However, the above relation cannot be satisfied by any $z \in H$, since for all $z \in H$ $\langle V_n | z\rangle$ and $\langle e_n | z\rangle$ approach zero as n gets large. Therefore, $\hat{P}_1\hat{P}_2 H$ is not a *closed* subspace of H. Of course, the fact that H is infinite-dimensional was crucial in the above example. Since $W_n \approx V_n$ as n gets large, we can say that $\hat{P}_1$ and $\hat{P}_2$ are *nearly* orthogonal to each other.

The projector $\hat{L}(\hat{P}_1\hat{P}_2)$ onto $\overline{\hat{P}_1\hat{P}_2 H}$ is called the *left projector* or *closed range*[8] of the operator $\hat{P}_1\hat{P}_2$. The projector $\hat{L}(\hat{P}_1\hat{P}_2)$ is the smallest projector $\hat{L}$ for which $\hat{L}\hat{P}_1\hat{P}_2 = \hat{P}_1\hat{P}_2$, and it may be expressed as[9]

$$\hat{L}(\hat{P}_1\hat{P}_2) = \hat{P}_1 - \hat{P}_1 \wedge \hat{P}_2^{\perp},$$

or equivalently by

$$\hat{L}(\hat{P}_1\hat{P}_2) = \hat{P}_1 \wedge (\hat{P}_1^{\perp} \vee \hat{P}_2)^{\perp}. \tag{3.17}$$

The right projector $\hat{R}(\hat{P}_1\hat{P}_2)$ (or *support*) of $\hat{P}_1\hat{P}_2$ is the smallest projector $\hat{R}$ such that $\hat{P}_1\hat{P}_2\hat{R} = \hat{P}_1\hat{P}_2$, and is equal to the left projector $\hat{L}(\hat{P}_2\hat{P}_1)$ of $\hat{P}_2\hat{P}_1$;

$$\hat{R}(\hat{P}_1\hat{P}_2) = \hat{L}(\hat{P}_2\hat{P}_1) = \hat{P}_2 - (\hat{P}_2 \wedge \hat{P}_1^{\perp}). \tag{3.18}$$

We note that the support $\hat{R}(\hat{P}_1\hat{P}_2)$ is equal to the orthogonal complement $\hat{N}(\hat{P}_1\hat{P}_2)^{\perp}$ of the projector $\hat{N}(\hat{P}_1\hat{P}_2)$ onto the null space of $\hat{P}_1\hat{P}_2$, $\hat{R} + \hat{N} = \hat{1}$.

For applications we are interested in the actual construction of various projectors, such as $\hat{P}_1 \wedge \hat{P}_2$, etc. It is well known[10] that $\hat{P}_1 \wedge \hat{P}_2$ can be expressed as a strong operator limit

$$\hat{P}_1 \wedge \hat{P}_2 = \underset{N\to\infty}{\text{s-lim}}(\hat{P}_1\hat{P}_2\hat{P}_1)^N$$

which means that for all $x \in H$

$$\hat{P}_1 \wedge \hat{P}_2 x = \lim(\hat{P}_1\hat{P}_2\hat{P}_1)^N x.$$

Less well known is an interesting expression for $\hat{P}_1 \vee \hat{P}_2$, given by Hill[11] for the case that $(\hat{P}_1\hat{P}_2\hat{P}_1)$, $(\hat{P}_2\hat{P}_1\hat{P}_2)$ are Hilbert–Schmidt class operators:

$$\hat{P}_1 \vee \hat{P}_2 = \frac{1}{2}(\hat{P}_1 + \hat{P}_2) + \frac{1}{2}\sum_{l=0}^{\infty}\hat{P}_1^{\perp}\hat{P}_2(\hat{P}_2\hat{P}_1\hat{P}_2)^l\hat{P}_2\hat{P}_1^{\perp}$$
$$+ \frac{1}{2}\sum_{l=0}^{\infty}\hat{P}_2^{\perp}\hat{P}_1(\hat{P}_1\hat{P}_2\hat{P}_1)^l\hat{P}_1\hat{P}_2^{\perp}.$$

For $\hat{P}_1\hat{P}_2\hat{P}_1$ to be a Hilbert–Schmidt operator means that $\text{tr}[(\hat{P}_1\hat{P}_2\hat{P}_1)^2] < \infty$, or if $\hat{P}_1 = \Sigma_{m=1}^{\infty}|e_m\rangle\langle e_m|$, that $\Sigma_{m,n}|\langle e_m|\hat{P}_2|e_n\rangle|^2 < \infty$. So $\hat{P}_1\hat{P}_2\hat{P}_1$ Hilbert–Schmidt is rather special. On the other hand, Hill's formula can apply in other cases.

We would like to have procedures which lend themselves to computation not only for $\hat{P}_1 \wedge \hat{P}_2$, $\hat{P}_1 \vee \hat{P}_2$, etc., but also to more complicated projectors such as $\wedge_i \hat{P}_i$, where i

ranges over a countable set. Before taking up these and other related questions, we discuss the important concept of (partial) isometries.

If P_1 and P_2 are subspaces of H having the same dimensions (i.e., both have finite and equal dimensions or both have countably infinite dimensions), then there exist *isometries* $\hat{\Omega}_{21}$ and $\hat{\Omega}_{12} = \hat{\Omega}^*_{21}$ such that

$$P_1 = \hat{\Omega}_{12} P_2 \tag{3.19}$$

and

$$P_2 = \hat{\Omega}_{21} P_1. \tag{3.20}$$

If $\hat{P}_1$ and $\hat{P}_2$ are projectors such that $P_1 = \hat{P}_1 H$ and $P_2 = \hat{P}_2 H$, then

$$\hat{\Omega}^*_{12}\hat{\Omega}_{12} = \hat{\Omega}_{21}\hat{\Omega}_{12} = \hat{\Omega}_{21}\hat{\Omega}^*_{21} = \hat{P}_2 \tag{3.21}$$

and

$$\hat{\Omega}^*_{21}\hat{\Omega}_{21} = \hat{\Omega}_{12}\hat{\Omega}_{21} = \hat{\Omega}_{12}\hat{\Omega}^*_{12} = \hat{P}_1. \tag{3.22}$$

That is, $\hat{\Omega}_{21}$, resp. $\hat{\Omega}_{12}$, acts as a unitary map when restricted to P_1, resp. P_2. The simplest way to establish the validity of the above statements is through the introduction of orthonormal bases $\{e_l(1)\}$ and $\{e_l(2)\}$ for P_1 and P_2. Then we define $\hat{\Omega}_{21}$ by

$$e_l(2) \equiv \hat{\Omega}_{21} e_l(1), \quad l = 1,2,3,\cdots, \tag{3.23}$$

and

$$\hat{\Omega}_{21} x = 0, \quad \text{for } x \in P_1^{\perp}. \tag{3.24}$$

It follows that

$$e_l(1) = \hat{\Omega}^*_{21} e_l(2), \quad l = 1,2,3,\cdots, \tag{3.25}$$

and

$$\hat{\Omega}^*_{21} y = 0, \quad \text{for } y \in P_2^{\perp}. \tag{3.26}$$

Therefore, $\hat{\Omega}_{12}$ satisfies the relations expressed by Eqs. (3.19)–(3.22) with $\hat{\Omega}_{12} = \hat{\Omega}^*_{21}$.

The isometry connecting P_1 and P_2 is not unique since any other orthonormal bases could be employed for P_1 and P_2. This lack of uniqueness may be expressed in terms of isometries $\hat{\Omega}_1$, $\hat{\Omega}_2$ defined by

$$\hat{\Omega}^*_1\hat{\Omega}_1 = \hat{\Omega}_1\hat{\Omega}^*_1 = \hat{P}_1 \tag{3.27}$$

and

$$\hat{\Omega}^*_2\hat{\Omega}_2 = \hat{\Omega}_2\hat{\Omega}^*_2 = \hat{P}_2. \tag{3.28}$$

If $\hat{\Omega}_{21}$ is an isometry connecting P_1 and P_2, then the isometry

$$\hat{\Omega}'_{21} \equiv \hat{\Omega}_2\hat{\Omega}_{21}\hat{\Omega}_1 \tag{3.29}$$

also connects P_1 and P_2.

If we are given a bounded operator $\hat{\Omega}$ on H for which $\hat{\Omega}^*\hat{\Omega} = \hat{P}$ is a projector, then $(\hat{\Omega}^*)^*\hat{\Omega}^* = \hat{\Omega}\hat{\Omega}^* \equiv \hat{P}'$ is also a projector. That is, if $\hat{\Omega}$ is any isometry, so is $\hat{\Omega}^*$. We prove this by first showing that $\hat{\Omega}^*\hat{\Omega} = \hat{P}$ implies that $\hat{\Omega}\hat{P} = \hat{\Omega}$, i.e., $\hat{P}$ is the right projector $\hat{R}(\hat{\Omega})$ of $\hat{\Omega}$. Let x be any vector in H, and write $x = \hat{P}x + \hat{P}^{\perp}x$, where $\hat{P}^{\perp} = \hat{1} - \hat{P}$. Then

$$\begin{aligned}\|\hat{\Omega}\hat{P}^{\perp}x\|^2 &= \langle \hat{\Omega}\hat{P}^{\perp}x | \hat{\Omega}\hat{P}^{\perp}x\rangle \\ &= \langle \hat{P}^{\perp}x | \hat{\Omega}^*\hat{\Omega}\hat{P}^{\perp}x\rangle = \langle \hat{P}^{\perp}x | \hat{P}\hat{P}^{\perp}x\rangle = 0.\end{aligned} \tag{3.30}$$

Hence $\hat{\Omega}\hat{P}^{\perp}x = \hat{\Omega}(\hat{1} - \hat{P})x \equiv 0$ and $\hat{\Omega}x = \hat{\Omega}\hat{P}x$, $\forall x \in H$, therefore, $\hat{\Omega} = \hat{\Omega}\hat{P}$. From the latter relation it follows that $\hat{P}'^2 = (\hat{\Omega}\hat{\Omega}^*)^2 = \hat{\Omega}\hat{\Omega}^*\hat{\Omega}\hat{\Omega}^* = \hat{\Omega}\hat{P}\hat{\Omega}^* = \hat{\Omega}\hat{\Omega}^* = \hat{P}'$, and therefore $\hat{P}'$ is also a projector. Further, since $(\hat{\Omega}^*)^*\hat{\Omega}^* = \hat{P}', \hat{P}'$ is the right projector $\hat{R}(\hat{\Omega}^*)$ of $\hat{\Omega}^*, \hat{\Omega}^*\hat{P}' = \hat{\Omega}^*$ and $\hat{P}'\hat{\Omega} = \hat{\Omega}$, or $\hat{P}' = \hat{L}(\hat{\Omega})$. The projector $\hat{P}$ is also called the *initial* projector of $\hat{\Omega}$ and $\hat{P}'$ the *final* projector of $\hat{\Omega}$. We have $\hat{\Omega} = \hat{P}'\hat{\Omega}\hat{P}$ and $\hat{\Omega}^* = \hat{P}\hat{\Omega}^*\hat{P}'$, and so the roles of $\hat{P}$ and $\hat{P}'$ are interchanged for $\hat{\Omega}^*$.

In general, the product $\hat{\Omega}_a\hat{\Omega}_b$ of two isometries $\hat{\Omega}_a$ and $\hat{\Omega}_b$ is not an isometry. However, if the final projector of $\hat{\Omega}_b$ is equal to the initial projector of $\hat{\Omega}_a$, then the product $\hat{\Omega}_a\hat{\Omega}_b$ is an isometry. This statement follows from $(\hat{\Omega}_a\hat{\Omega}_b)^*\hat{\Omega}_a\hat{\Omega}_b = \hat{\Omega}^*_b\hat{\Omega}^*_a\hat{\Omega}_a\hat{\Omega}_b = \hat{\Omega}^*_b\hat{R}(\hat{\Omega}_a)\hat{\Omega}_b = \hat{\Omega}^*_b\hat{L}(\hat{\Omega}_b)\hat{\Omega}_b = \hat{\Omega}^*_b\hat{\Omega}_b = \hat{R}(\hat{\Omega}_b)$, where $\hat{L}(\hat{\Omega})$, resp. $\hat{R}(\hat{\Omega})$, stands for the left, resp. right, projector of $\hat{\Omega}$.

An important problem in applications is the construction of an orthonormal set of vectors from a given set of vectors. The Schmidt construction is well known, but new and more useful constructions have been developed during the past few decades or so. We will discuss some of these in the next section. Before turning to these constructions, we present an account of a generalized Schmidt process[12] in which an orthogonal set of projectors is constructed from a given arbitrary countable set of projectors.

Let $\{\hat{P}_i\}_{i=1}^{\infty}$ be a given set of projectors. We define for each n, $n = 1,2,3,\cdots$, new projectors $\overline{P}_n$:

$$\overline{P}_n \equiv \bigvee_{i=1}^{n} \hat{P}_i. \tag{3.31}$$

That is, $\overline{P}_1 = \hat{P}_1$, $\overline{P}_2 = \hat{P}_1 \vee \hat{P}_2$, $\overline{P}_3 = \hat{P}_1 \vee \hat{P}_2 \vee \hat{P}_3, \cdots$. It is clear that

$$\overline{P}_n < \overline{P}_{n+1}, \quad n = 1,2,3,\cdots, \tag{3.32}$$

and therefore $\tilde{P}_n$ defined by

$$\tilde{P}_n \equiv \overline{P}_n - \overline{P}_{n-1} \tag{3.33}$$

is a projector ($\overline{P}_0 \equiv 0$). The $\tilde{P}_n$ are orthonormal:

$$\tilde{P}_n\tilde{P}_m = \delta_{nm}\tilde{P}_n. \tag{3.34}$$

Further, $\tilde{P}_n$ span the same subspace of H that is spanned by the original $\hat{P}_i$. If $\tilde{P} \equiv \Sigma_n \tilde{P}_n$, then $\tilde{P}H$ is the same as the closure of the set of all vectors x_P of the form

$$x_P \equiv \sum_{i=1}^{N} \hat{P}_i x, \quad x \in H, \quad N = 1,2,3,\cdots. \tag{3.35}$$

We may discard all the $\tilde{P}_n$ which are zero. Then the remaining set $\{\tilde{P}_{n'}\}$, say, are still orthogonal

$$\tilde{P}_{n'}\tilde{P}_{m'} = \delta_{n'm'}\tilde{P}_{n'} \tag{3.36}$$

and

$$\tilde{P} = \sum_{n'} \tilde{P}_{n'}. \tag{3.37}$$

We note again that it would be desirable and practical to have explicit methods to compute $\vee_i \hat{P}_i$ and $\wedge_i \hat{P}_i$. We address this problem in the next section.

IV. CONSTRUCTION OF PROJECTORS

Let $\{\psi_n\}_{n=1}^{\infty}$ be a countable collection of vectors in a separable Hilbert space H. We wish to find the projector $\hat{P}$ corresponding to the (closed) subspace P spanned by the vectors ψ_n. It may also be desirable to construct an orthonormal

set of vectors spanning P. Of course, the Schmidt process may be applied to the latter problem, but this procedure requires the selection of the first, second, ..., vector of the given collection. We would like a more democratic process in which the given vectors are all treated on the same basis.

First, assume that the set $\{\psi_n\}_{n=1}^{\infty}$ is linearly independent. We then know that there exists a set $\{\psi^n\}_{n=1}^{\infty}$ of vectors which are biorthogonal to $\{\psi_n\}_{n=1}^{\infty}$. That is there are vectors ψ^n which satisfy the relations

$$\langle\psi^n|\psi_m\rangle = \langle\psi_m|\psi^n\rangle = \delta^n{}_m, \text{ all } n,m. \tag{4.1}$$

Further, the projector $\hat{P}$ on the span of $\{\psi_n\}$ is just

$$\hat{P} = \sum_n |\psi^n\rangle\langle\psi_n|, \tag{4.2}$$

so if $|\psi^n\rangle$ can be found, we have a construction for $\hat{P}$. Let $P = \hat{P}H$; then, if $\phi \in P^{\perp}$, $\hat{P}\phi = 0$ or, equivalently, $\langle\psi_n|\phi\rangle = 0$, all n, which is equivalent to

$$\hat{g}\phi = 0, \tag{4.3}$$

where

$$\hat{g} \equiv \sum_n |\psi_n\rangle\langle\psi_n|. \tag{4.4}$$

The sum in Eq. (4.4) is the strong limit of the sum $\Sigma_{n=1}^{N}$ as $N\to\infty$. In order for $\hat{g}$ to exist, some growth property of ψ_n as $n\to\infty$ must be assumed. It is clear that the vectors ψ_n can always be renormalized to satisfy any desired features of growth and still span the same subspace P. For mathematical convenience we make the mild technical assumption that

$$\|\psi_n\| \leqslant A/n, \quad n = 1,2,3,\cdots, \tag{4.5}$$

where A is a positive constant. This assumption is not necessary for certain of our results and can be dispensed with in many instances if one is careful. It follows from Eq. (4.5) that

$$\left\| \sum_{n=1}^{N} |\psi_n\rangle\langle\psi_n| \right\| \leqslant \sum_{n=1}^{N} \| |\psi_n\rangle\langle\psi_n| \| = \sum_{n=1}^{N} \| |\psi_n\rangle \|^2 \leqslant \sum_{n=1}^{N} \frac{A^2}{n^2}, \tag{4.6}$$

so that $\hat{g}$ defined by Eq. (4.4) is the uniform limit as $N\to\infty$ of $\Sigma_{n=1}^{N}|\psi_n\rangle\langle\psi_n|$. The latter sum is a positive finite rank operator, and, therefore, $\hat{g}$ exists as a positive compact operator. This means in turn that $\hat{g}$ has a discrete spectrum $\lambda_1 > \lambda_2 > \cdots > 0$, with no accumulation points, except possibly $\lambda = 0$. Further, the eigenspace P_λ for a given $\lambda \neq 0$ is finite-dimensional. The spectral resolution of $\hat{g}$ is given by

$$\hat{g} = \sum_r \lambda_r \hat{P}_r, \tag{4.7}$$

where $\hat{P}_r$ is the projector onto the eigenspace of $\hat{g}$ spanned by the eigenfunctions of $\hat{g}$ having eigenvalue λ_r ($\lambda_r \neq \lambda_s$, $r \neq s$). We note that $\lambda = 0$ does not appear in Eq. (4.7).

Let $\hat{g}_\epsilon = \hat{g} + \epsilon\hat{1}$, $\epsilon > 0$, so that $\hat{g}_\epsilon$ is a strictly positive operator. The product $\hat{g}_\epsilon^{-1}\hat{g}$ may be expressed by

$$\hat{g}\hat{g}_\epsilon^{-1} = \hat{g}_\epsilon^{-1}\hat{g} = \sum_r \frac{\lambda_r}{\lambda_r + \epsilon}\hat{P}_r. \tag{4.8}$$

Since $\lambda_r = 0$ does not contribute in Eq. (4.8), we may safely allow $\epsilon \downarrow 0$ to find

$$\text{s-lim}_{\epsilon\downarrow 0}\, \hat{g}\hat{g}_\epsilon^{-1} = \text{s-lim}_{\epsilon\downarrow 0}\, \hat{g}_\epsilon^{-1}\hat{g} = \sum_{\lambda_r > 0} \hat{P}_r = \hat{P}, \tag{4.9}$$

the sum being the strong limit of a sum of orthogonal projectors is itself an orthogonal projector $\hat{P}$. The projector $\hat{P}$ is the projector on the subspace of H spanned by the vectors $\{\psi_n\}_{n=1}^{\infty}$.

The prescription given by Eq. (4.9) is valid even if the vectors ψ_n are dependent, so long as $\hat{g}$ exists as a suitable limit. In fact,[13] it can be established rigorously that if $\hat{A}$ is any bounded nonnegative operator, then

$$\hat{P} = \text{s-lim}_{\epsilon\downarrow 0}(\hat{A}_\epsilon^{-1}\hat{A}) \tag{4.10}$$

exists and is the projector on the support of $\hat{A}$. Similarly

$$\hat{Q} = \text{s-lim}_{\epsilon\downarrow 0}(\epsilon\hat{A}_\epsilon^{-1}) \tag{4.11}$$

exists and is the projector on the null space of $\hat{A}$. The sum of the projectors, Eq. (4.10) and (4.11), is unity, $\hat{P} + \hat{Q} = \hat{1}_H$. In addition $\hat{A}\hat{P} = \hat{P}\hat{A} = \hat{A}$ and $\hat{A}\hat{Q} = \hat{Q}\hat{A} = \hat{0}$, where $\hat{0}$ is the zero operator on H.

Now define[14]

$$\psi^n \equiv \lim_{\epsilon\downarrow 0} \hat{g}_\epsilon^{-1}\psi_n. \tag{4.12}$$

Lemma: $\{\psi^m\}$ forms a biorthogonal set to $\{\psi_n\}$, $\langle\psi^m|\psi_n\rangle = \delta^m{}_n$, $m,n = 1,2,3,\cdots$.

Proof:

$$\hat{P} = \text{s-lim}_{\epsilon\downarrow 0}\, \hat{g}\hat{g}_\epsilon^{-1} = \text{s-lim}_{\epsilon\downarrow 0} \sum_n |\psi_n\rangle\langle\psi_n|\hat{g}_\epsilon^{-1} = \sum_n |\psi_n\rangle\langle\psi^n|. \tag{4.13}$$

Since $\psi_r \in P$, $\hat{P}\psi_r = \psi_r$, and

$$\psi_r = \sum_n \psi_n\langle\psi^n|\psi_r\rangle = \sum_n \delta^n{}_r\psi_n. \tag{4.14}$$

Therefore,

$$\sum_n \psi_n\{\langle\psi^n|\psi_r\rangle - \delta^n_r\} = 0, \tag{4.15}$$

and, since the ψ_n are linearly independent,

$$\langle\psi^n|\psi_r\rangle = \delta^n{}_r. \tag{4.16}$$

We may write Eqs. (4.16) as the limit as $\epsilon \downarrow 0$ of

$$\langle\psi^n|\psi_m\rangle = \langle\psi_n|\hat{g}_\epsilon^{-1}|\psi_m\rangle = \langle\psi_n|\hat{g}_\epsilon^{-1/2}|\hat{g}_\epsilon^{-1/2}\psi_m\rangle = \langle\tilde{\psi}_n|\tilde{\psi}_m\rangle = \delta^n{}_r, \tag{4.17}$$

where

$$\tilde{\psi}_m = \hat{g}_\epsilon^{-1/2}\psi_m, \quad \epsilon \downarrow 0. \tag{4.18}$$

Clearly,

$$\sum_m |\tilde{\psi}_m\rangle\langle\tilde{\psi}_m| = \text{s-lim}_{\epsilon\downarrow 0} \sum_m \hat{g}_\epsilon^{-1/2}|\psi_m\rangle\langle\psi_m|\hat{g}_\epsilon^{-1/2} = \text{s-lim}_{\epsilon\downarrow 0}\, \hat{g}_\epsilon^{-1/2}\hat{g}\hat{g}_\epsilon^{-1/2} = \hat{P}. \tag{4.19}$$

Operators such as $\hat{g}_\epsilon^{-1/2}$ are somewhat unfamiliar objects in physics although operators such as $\hat{g}_\epsilon^{-1}$ appear fre-

quently. There are known expressions for $\hat{g}_\epsilon^{-1/2}$ and we mention two. The first, given by Kato,[15] is expressed by the integral,

$$\hat{g}_\epsilon^{-1/2} = \frac{1}{\pi} e^{i\theta/2} \int_0^\infty \frac{d\lambda}{\lambda^{1/2}} (\hat{g}_\epsilon + \lambda e^{i\theta})^{-1}, \tag{4.20}$$

where θ is fixed and $|\theta| < \pi/2$. The integral in Eq. (4.20) is absolutely convergent. The second is based on a suggestion by Montroll.[16] If $0 < \alpha$, then $(\hat{g}_\epsilon)^{-\alpha}$ may be expressed by the integral,

$$(\hat{g}_\epsilon)^{-\alpha} = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda \hat{g}_\epsilon} \lambda^{\alpha-1} d\lambda. \tag{4.21}$$

Since our $\hat{g}_\epsilon$ is a positive self-adjoint operator, so is $(\hat{g}_\epsilon)^{-\alpha}$, $0 < \alpha$.

If $\{\psi_n\}$ is not a linearly independent set, the contstruction of $\hat{P}$ as $\Sigma |\tilde{\psi}_n\rangle\langle\tilde{\psi}_n|$ is still valid; however, the set $\{\tilde{\psi}_n\}$ is no longer orthonormal. Similarly, $\hat{P} = \Sigma |\psi^n\rangle\langle\psi_n|$; however, $\{\psi^n\}$ and $\{\psi_n\}$ are no longer biorthogonal sets.

We now let $\{\psi_n\}$ be a linearly dependent set and for convenience assume the same type of growth as before so that

$$\hat{g} \equiv \sum_n |\psi_n\rangle\langle\psi_n| \tag{4.22}$$

is a compact positive (nonnegative) operator. First note that any eigenvector of $\hat{g}$ belonging to the eigenvalue zero lies in $P^\perp$ the subspace of H orthogonal to all vectors in P (spanned by the ψ_n). Then if $\{\phi_n\}$ is the complete orthornormal set of eigenvectors of $\hat{g}$,

$$\hat{g}\phi_n = \lambda_n \phi_n. \tag{4.23}$$

We have as before

$$\hat{P} = \sum_{\lambda_n > 0} |\phi_n\rangle\langle\phi_n| \tag{4.24}$$

with the sum including multiplicity of eigenvalues. Since ψ_n span P, $\phi \in P$, may be expanded (nonuniquely)

$$\phi = \sum a_n \psi_n, \tag{4.25}$$

with

$$\|\phi\|^2 = \langle\phi|\phi\rangle = \sum_{m,n} \bar{a}_m g_{mn} a_n, \tag{4.26}$$

where $g_{mn} = \langle\psi_m|\psi_n\rangle = \bar{g}_{nm}$. The matrix $(g) \equiv (g_{mn})$ is called the *overlap matrix*. If we require the a_n in Eq. (4.25) to satisfy the eigenvalue equation, in l_2,

$$\sum_n g_{mn} a_n = \lambda a_m, \quad m = 1, 2, \ldots, \tag{4.27}$$

then

$$\hat{g}\phi = \lambda\phi. \tag{4.28}$$

Hence solutions of Eq. (4.27) yield eigenfunctions of $\hat{g}$ via Eq. (4.25), *except* when $\lambda = 0$, $\phi \in P$, and also $\phi \in P^\perp$, so $\phi \equiv 0$.

The construction of $\hat{P}$ may be computed from the nonzero eigenvalue solutions of Eq. (4.28). Let $a_n(\lambda_k)$, $\lambda_k \neq 0$, be solutions of

$$\sum_n g_{mn} a_n(\lambda_k) = \lambda_k a_m(\lambda_k), \quad m = 1, 2, \ldots \text{ and } k = 1, 2, \ldots, \tag{4.29}$$

normalized to unity,

$$\sum_n \bar{a}_n(\lambda_k) a_n(\lambda_{k'}) = \delta(\lambda_k, \lambda_{k'}). \tag{4.30}$$

Then

$$\phi_{\lambda_k} = N_k \sum_n a_n(\lambda_k)\psi_n \tag{4.31}$$

are eigenvectors of $\hat{g}$. We choose N_k so that

$$\langle\phi_{\lambda_k}|\phi_{\lambda_{k'}}\rangle = \delta(\lambda_k, \lambda_{k'}), \tag{4.32}$$

which means that

$$1 = \|\phi_{\lambda_k}\|^2 = |N_k|^2 \sum_{mn} \bar{a}_m(\lambda_k) g_{mn} a_n(\lambda_k) \tag{4.33}$$

$$= \lambda_k |N_k|^2 \sum_n |a_n(\lambda_k)|^2 = \lambda_k |N_k|^2, \tag{4.34}$$

or

$$\phi_{\lambda_k} = \frac{1}{\sqrt{\lambda_k}} \sum_n a_n(\lambda_k)\psi_n. \tag{4.35}$$

Thus

$$\hat{P} = \sum_k |\phi_{\lambda_k}\rangle\langle\phi_{\lambda_k}|. \tag{4.36}$$

The sum in Eq. (4.36) can be extended over all λ since $\phi_{\lambda=0} \equiv 0$. The construction of the orthornormal set $\{\phi_\lambda\}$ of eigenvectors of $\hat{g}$ is known as *canonical* orthogonalization while the construction of the orthonormal set $\tilde{\psi}_n$, when they exist (i.e., when ψ_n are linearly independent), is known as *symmetric* orthogonalization.[17]

We may exploit the overlap matrix $g = (g_{mn})$ to compute $\hat{P}$ in another way, even when the ψ_n are linearly dependent. We have from the definition of $\hat{g}$,

$$\hat{g}\psi_n = \sum_m |\psi_m\rangle\langle\psi_m|\psi_n\rangle = \sum_m \psi_m g_{mn},$$

$$\hat{g}^2\psi_n = \sum_{r,m} \psi_r g_{rm} g_{mn} = \sum_m \psi_m g^2_{mn}, \ldots, \tag{4.37}$$

where g^2_{mn} is the m, nth matrix element of the square of the matrix g, and, if $F(\hat{g})$ is any well-defined function of $\hat{g}$,

$$F(\hat{g})\psi_n = \sum_m \psi_m F(g)_{mn}. \tag{4.38}$$

Now let $F(z) = (z + \epsilon I)^{-1}$ in Eq. (4.38); then

$$\psi^n = \lim_{\epsilon \downarrow 0} \hat{g}_\epsilon^{-1}\psi_n = \lim_{\epsilon \downarrow 0} \sum_m \psi_m (g_\epsilon^{-1})_{mn}, \tag{4.39}$$

where (g_ϵ^{-1}) is the inverse of the matrix $(\langle\psi_m|\psi_n\rangle + \epsilon(\delta_{mn})) \equiv (g_\epsilon)$.

Thus we find the following expression for $\hat{P}$:

$$\hat{P} = \sum_n |\psi^n\rangle\langle\psi_n| = \lim_{\epsilon \downarrow 0} \sum_{m,n} |\psi_m\rangle (g_\epsilon^{-1})_{mn} \langle\psi_n|, \tag{4.40}$$

in terms of the given set $\{\psi_n\}$. Equation (4.40) generalizes the

formula of Feshbach,[18,19] to which Eq. (4.40) reduces when the ψ_n are linearly independent, namely,

$$\hat{P} = \sum_{mn} |\psi_m\rangle (g^{-1})_{mn} \langle\psi_n|, \tag{4.41}$$

since in this case the matrix g is nonsingular. The vectors $\tilde{\psi}_n$ may also be computed from

$$\tilde{\psi}_n = \lim_{\epsilon \downarrow 0} \hat{g}_\epsilon^{-1/2} \psi_n = \lim_{\epsilon \downarrow 0} \sum_m \psi_m (g_\epsilon^{-1/2})_{mn}. \tag{4.42}$$

The latter $\tilde{\psi}_n$ are orthonormal if and only if ψ_n are linearly independent. The projector $\hat{P}$ may in any case be expressed in terms of $\tilde{\psi}_n$,

$$\hat{P} = \sum_n |\tilde{\psi}_n\rangle\langle\tilde{\psi}_n|. \tag{4.43}$$

Since

$$\tilde{\psi}_n = \hat{P}\tilde{\psi}_n,$$

$$\|\tilde{\psi}_n\|^2 = \langle\tilde{\psi}_n|\hat{P}\tilde{\psi}_n\rangle$$

$$= \sum_m |\langle\tilde{\psi}_n|\tilde{\psi}_m\rangle|^2$$

$$= \|\tilde{\psi}_n\|^4 + \sum_{m\neq n} |\langle\tilde{\psi}_n|\tilde{\psi}_m\rangle|^2,$$

we must have for the linearly dependent case ($\langle\tilde{\psi}_n|\tilde{\psi}_m\rangle \neq 0$, some n, m, $n \neq m$)

$$\|\tilde{\psi}_n\| < 1, \quad \text{some } n. \tag{4.44}$$

The closer the $\|\tilde{\psi}_n\|$ are to unity, the closer $\tilde{\psi}_n$ are to being orthonormal. In fact, we can characterize a measure of independence of the set $\{\psi_n\}$ by means of the smallness of $\Sigma_n (1 - \|\tilde{\psi}_n\|^2)$. The dimension of the subspace P is given by trace $(\hat{P})$ (positive integer or $+\infty$). Thus

$$\dim P = \mathrm{tr}\left(\sum_n |\tilde{\psi}_n\rangle\langle\tilde{\psi}_n|\right) = \sum_n \langle\tilde{\psi}_n|\tilde{\psi}_n\rangle = \sum_n \|\tilde{\psi}_n\|^2, \tag{4.45}$$

and $\Sigma_n (1 - \|\tilde{\psi}_n\|^2)$ is the "total number" of ψ_n's minus the dim P (possibly $+\infty$). The ψ_n's are linearly independent if and only if this sum is zero.

We now proceed to construct various combinations of projectors. Given an arbitrary countable set $\{\hat{P}_n\}_{n=1}^{\infty}$ of projectors, we would like to find expressions for $\vee_n \hat{P}_n$ and $\wedge_n \hat{P}_n$. Since $\wedge_n \hat{P}_n = (\vee_n \hat{P}_n^{\perp})^{\perp}$, where $\hat{P}_n^{\perp} = \hat{1} - \hat{P}_n$, we need only to compute $\vee_n \hat{P}_n$. The subspace $P = \hat{P}H$ where $\hat{P} \equiv \vee_n \hat{P}_n$ consists of linear combinations of $\hat{P}_n x$, $x \in H$, together with the closure of the set of such linear combinations. If $\{x_n\}$ is a complete orthonormal set of vectors in H, we define ψ_{nm} by

$$\psi_{nm} = a_n P_n x_m, \quad a_n \neq 0, \tag{4.46}$$

where $a_n \in C$, $\Sigma_n |a_n|^2 < \infty$. Then let $\hat{g}$ be defined by

$$\hat{g} = \sum_{n,m} |\psi_{nm}\rangle\langle\psi_{nm}| = \sum_{n,m} |a_n|^2 \hat{P}_n |x_m\rangle\langle x_m| \hat{P}_n$$

$$= \sum_n |a_n|^2 \hat{P}_n. \tag{4.47}$$

Since $\|\hat{P}_n\| \leq 1$, $\|\hat{g}\| \leq \sum_n |a_n|^2 < \infty$ so $\hat{g}$ is a bounded positive operator. If $\hat{g}\psi = 0$, then $\hat{P}_n\psi = 0$, $n = 1, 2, 3, \ldots$, which is proved as follows: Let

$$\hat{g}\psi = \sum_n |a_n|^2 \hat{P}_n \psi = 0.$$

Then $\|\hat{g}^{1/2}\psi\|^2 = \langle\psi|\hat{g}\psi\rangle = \Sigma_n |a_n|^2 \|\hat{P}_n\psi\|^2 = 0$, and hence $\|\hat{P}_n\psi\| = 0 \ \forall\, n$, so $\hat{P}_n\psi = 0$. Therefore, $\hat{P} = \vee_n \hat{P}_n$ is just the support of $\hat{g}$:

$$\vee_n \hat{P}_n = \text{s-lim}_{\epsilon \downarrow 0} (\hat{g}_\epsilon^{-1} g), \tag{4.48}$$

$$\hat{g} \equiv \Sigma_n |a_n|^2 \hat{P}_n, \quad \Sigma_n |a_n|^2 = 1, \text{ say.}$$

Since $\text{s-lim}_{\lambda \to \infty} \exp(-\lambda\hat{g}) = \hat{P}_0 = \hat{1} - \hat{P}$, the projector on the orthogonal complement of P, we may also write Eq. (4.48) as

$$\vee_n \hat{P}_n = \hat{1} - \text{s-lim}_{\epsilon \downarrow 0} (\epsilon \hat{g}_\epsilon^{-1}) = \hat{1} - \text{s-lim}_{\lambda \to \infty} \exp(-\lambda\hat{g}). \tag{4.49}$$

More general collections of projectors arise in applications. For example, $\{\hat{P}_z\}_{z \in C}$, where $\hat{P}_z \equiv |z\rangle\langle z|$, and $|z\rangle$ is a coherent state. If we are given a collection $\{\hat{P}_\alpha\}_{\alpha \in I}$ (I a *measurable* set) of projectors $\hat{P}_\alpha$, we may use a similar technique to compute $\vee_{\alpha \in I} \hat{P}_\alpha$ to that expressed by Eq. (4.48). Let $d\mu(\alpha)$ be a positive measure on I with $\int_{\alpha \in I} d\mu(\alpha) < \infty$, then we define $\hat{g}$ similarly to that given by Eq. (4.47),

$$\hat{g} \equiv \int_{\alpha \in I} d\mu(\alpha)\, \hat{P}_\alpha, \tag{4.50}$$

and show that $\hat{g} \geq 0$, $\hat{g}\phi = 0 \Rightarrow \hat{P}_\alpha \phi = 0$, etc. Then we use arguments similar to those used for $\vee_n \hat{P}_n$, to show that

$$\vee_{\alpha \in I} \hat{P}_\alpha = \text{s-lim}_{\epsilon \downarrow 0} \hat{g}_\epsilon^{-1} \hat{g}. \tag{4.51}$$

V. MEASUREMENT THEORY AND STATISTICAL MECHANICS

Pure states in quantum theory are represented by unit vectors $|\psi\rangle$ in a separable Hilbert space H. Since $|\psi\rangle$ and $\lambda|\psi\rangle$, $|\lambda| = 1$, represent the same state, it is more accurate to say that a pure state is represented by a unit ray $\{\lambda|\psi\rangle\}_{|\lambda|=1}$ in H, or more concisely by a one-dimensional projector $\rho(\psi) = |\psi\rangle\langle\psi|$. The projector $\rho(\psi)$ is independent of which vector is selected from the ray representation, for

$$\rho(\lambda\psi) = \lambda|\psi\rangle\langle\psi|\bar{\lambda} = |\lambda|^2 \rho(\psi) = \rho(\psi). \tag{5.1}$$

A pure state is regarded[20] as containing information relating to potential or propensity probabilities concerning the outcome of measurements which can be performed on an ensemble of systems represented by $\rho(\psi)$.

More general states involving *incomplete informational* probabilities also occur in quantum theory. Such states are represented by (positive) self-adjoint operators ρ of trace unity. The operators ρ are called statistical operators (or density matrices). If the system is in a state represented by ρ and if A is a bounded self-adjoint operator representing an observable (also denoted by A), then the average $\langle A \rangle$ of A when measurements are made on the ensemble represented by ρ is given by

$$\langle A \rangle = \tau(\rho A) \tag{5.2}$$

where τ represents *trace*.

Not all states can be represented by a statistical operator. For example, an equilibrium system having finite volume V and finite number N of degrees of freedom is represented by the canonical

$$\rho(N, V) = e^{-\beta H_N}/Z(N, V), \tag{5.3}$$

where H_N is the Hamiltonian of the N degrees of freedom in V and $\beta = 1/kT$. In the limit $N, V \to \infty$, $N/V \to$ const (thermodynamic limit), $\lim_{\substack{N\to\infty \\ V\to\infty}} \rho(N, V)$ does not exist. However, for certain observables C it can happen that

$$\lim_{\substack{n\to\infty \\ V\to\infty \\ N/V\to\text{const}}} \tau(C\rho(N, V)) \equiv \langle C \rangle \tag{5.4}$$

exists, and this limiting form defines then a *state* as a linear functional $\pi(C) = \langle C \rangle$ on the observables C.[21] Thus more generally states will be represented by positive $[\pi(C^*C) \geqslant 0]$ linear functionals π on the algebra $A(C)$ of certain observables C. We will not consider further such general states, but will restrict our discussion to states represented by a statistical operator ρ. (This in no way is meant to minimize their importance for certain aspects of modern statistical mechanics.)

The statistical operator ρ satisfies

$$\tau(\rho) = 1, \quad 0 < \rho^2 \leqslant \rho < 1, \tag{5.5}$$

where the latter inequality means that for any $|\psi\rangle \in H$

$$0 \leqslant \langle\psi|\rho^2|\psi\rangle \leqslant \langle\psi|\rho|\psi\rangle \leqslant 1. \tag{5.6}$$

Statistical operators have the nice mathematical property of belonging to the class of compact operators and thus have a discrete spectrum[22] which has at most one accumulation point, zero. That is, if ρ_k are the eigenvalues of ρ, then

$$1 \geqslant \rho_1 > \rho_2 \cdots > \rho_k > \cdots \geqslant 0. \tag{5.7}$$

If $\hat{P}_k$ is the projector onto the k th eigenspace of ρ, then the dimension $d_k \equiv \tau(\hat{P}_k)$ of the k th eigenspace is finite, $d_k < \infty$. Thus

$$\rho = \sum_{k=1}^{\infty} \rho_k \hat{P}_k \tag{5.8}$$

(strong limit), and

$$1 = \tau(\rho) = \sum_{k=1}^{\infty} \rho_k d_k. \tag{5.9}$$

Further, since $\hat{P}_k \rho \hat{P}_k = \hat{P}_k \rho = \rho \hat{P}_k = \rho_k \hat{P}_k$,

$$\rho = \sum_k \hat{P}_k \rho \hat{P}_k. \tag{5.10}$$

Let A be a bounded observable having a discrete spectrum. Further let A commute with ρ. Then $0 = [\rho, A] = \Sigma_a a[\rho, \hat{P}_a]$, where $\hat{P}_a$ is a spectral projector and a the corresponding eigenvalue of A. Since $\hat{P}_a \hat{P}_b = \delta_{ab} \hat{P}_b$,

$$0 = \sum_a a\hat{P}_{a'}, [\rho, \hat{P}_a] \hat{P}_{a'} = (a'' - a')\hat{P}_{a'} \rho \hat{P}_{a'} \tag{5.11}$$

and

$$\rho = \sum_a \hat{P}_a \rho \hat{P}_a. \tag{5.12}$$

We now derive what can be called the canonical theory of the quantum mechanical measurement process. This theory was apparently first formulated by Luders. However, we base it upon the *principle of minimum disturbance* due to Goldberg and Watson[23] (anticipated by Wigner), but follow the presentation of Herbut.[24]

The principle of minimum disturbance states that in an ideal measurement the state immediately after the measurement is as *close* as possible to the state just prior to the measurement, subject to the constraints given by the measurement. For pure states the principle states that the overlap $|\langle\psi'|\psi\rangle|$ is as large as possible, where $|\psi\rangle$ is the state just before the measurement is made and $|\psi'\rangle$ the state immediately after the measurement. The principle of maximum overlap is equivalent to the minimization of the *distance* $d(\psi, \psi')$ between *states*, where d is the positive square root of

$$d(\psi, \psi')^2 \equiv 2(1 - |\langle\psi'|\psi\rangle|^2). \tag{5.13}$$

This distance $d(\psi, \psi')$ has its maximum $\sqrt{2}$ when $|\psi'\rangle$ and $|\psi\rangle$ are orthogonal and is zero when $|\psi'\rangle = \lambda|\psi\rangle$, $|\lambda| = 1$, i.e., when the states are the same.

Since ρ is trace class, $\sqrt{\rho}$ is Hilbert–Schmidt. The set $B_2(H)$ of all Hilbert–Schmidt operators forms a Hilbert space on which the inner product, $\langle A|B\rangle_2$, between any pair A, B of operators in $B_2(H)$ is defined by

$$\langle A|B\rangle_2 \equiv \tau(A^*B). \tag{5.14}$$

Be definition $A \in B_2(H)$ implies that

$$\|A\|_2^2 \equiv \langle A|A\rangle_2 = \tau(A^*A) < \infty. \tag{5.15}$$

The positive square root of $\|A\|_2^2$ is the Hilbert–Schmidt norm of A. For trace-class operators A (such as ρ) the natural norm is the trace norm $\|A\|_1$ which is defined as

$$\|A\|_1 \equiv \tau(|A|), \tag{5.16}$$

where $|A|$ is the unique positive square root of A^*A. It would then be natural to define the distance between states ρ_1 and ρ_2 by means of $\|\rho_1 - \rho_2\|_1$. However, the trace is technically more difficult to handle than the Hilbert–Schmidt norm $\|\rho_1 - \rho_2\|_2$. The operator, Hilbert–Schmidt, and trace norms stand in relation[22] $\|\cdot\| < \|\cdot\|_2 < \|\cdot\|_1$. Nevertheless, we take $\|\rho_1 - \rho_2\|_2$ as the distance between states, realizing that there may be other, and for some purposes, more suitable distance functions.[25] In fact, another candidate for the distance between states ρ and ρ' is $d_1^2(\rho, \rho') \equiv \|\sqrt{\rho} - \sqrt{\rho'}\|_2^2$
$= \tau((\sqrt{\rho} - \sqrt{\rho'})^2) = \tau(\rho + \rho' - 2\sqrt{\rho}\sqrt{\rho'})$
$= 2(1 - \tau(\sqrt{\rho}\sqrt{\rho'}))$. This distance is analogous to the distance given by Eq. (4.11) for pure states with the overlap given by $\tau(\sqrt{\rho}\sqrt{\rho'})$. Since for pure states $\rho^2 = \rho$ and $\rho'^2 = \rho'$, we have in this case $\|\sqrt{\rho} - \sqrt{\rho'}\|_2 = \|\rho - \rho'\|_2$.

We now apply the above concepts to formulate the canonical theory of measurement. If A is an observable (for simplicity assume that A has a discrete spectrum), and if a measurement on the system yields an eigenvalue a_m of A, then if the measurement is immediately repeated the result is again a_m. This statement is a basic assumption of measurement theory. Further the state ρ just after the first measurement must satisfy the relation $\rho = \hat{P}_m \rho \hat{P}_m$, where $\hat{P}_m$ is the projector on the eigenspace of A corresponding to the eigenvalue a_m. We demonstrate the above statement by observing

that the probability, after the first measurement, of finding the value a_n, $n \neq m$, must be zero:

$$p_n = \langle \hat{P}_n \rho \rangle = \tau(\hat{P}_n \rho) = 0, \quad n \neq m. \tag{5.17}$$

Therefore,

$$0 = \tau(\hat{P}_n \rho) = \tau(\hat{P}_n \rho \hat{P}_n) = \tau\{(\rho^{1/2}\hat{P}_n)^*(\rho^{1/2}\hat{P}_n)\} = \| \rho^{1/2}\hat{P}_n \|_2^2. \tag{5.18}$$

Thus

$$\rho^{1/2}\hat{P}_n = 0$$

and

$$\rho \hat{P}_n = 0, \quad n \neq m. \tag{5.19}$$

The adjoint of Eq. (5.19) yields $\hat{P}_n \rho = 0$, $n \neq m$ and since $1 = \Sigma \hat{P}_n$,

$$\rho = \sum_{n,r} \hat{P}_n \rho \hat{P}_r = \hat{P}_m \rho \hat{P}_m. \tag{5.20}$$

Hence, the necessary and sufficient condition that ρ represent a state for which the value of A is with certainty a_m, is expressed by Eq. (5.20) above.

Now let ρ represent a state just before a measurement on A is made. Let p_n be the fraction of the measurements which yield the value a_n (a measurement means determining the values of A on the ensemble of systems represented by ρ). We expect, if the measurements are precise, that $p_n = \tau(\hat{P}_n \rho)$. We now seek ρ' the state of the ensemble just after the measurement. We may decompose ρ'

$$\rho' = \sum_n w_n \rho_n', \quad \sum_n w_n = 1, \tag{5.21}$$

where $w_n \geqslant 0$ and ρ_n' is a statistical operator for which the value of A is a_n with certainty. From Eq. (5.20), $\rho_n' = \hat{P}_n \rho_n' \hat{P}_n$, and we have

$$\rho' = \sum_n w_n \hat{P}_n \rho_n' \hat{P}_n. \tag{5.22}$$

If Eq. (5.21) is pre- and postmultiplied by $\hat{P}_m$, we find

$$\hat{P}_m \rho' \hat{P}_m = \sum w_n \hat{P}_m \rho_n' \hat{P}_m, \tag{5.23}$$

and, since $\rho_n' = \hat{P}_n \rho_n' \hat{P}_n$,

$$\hat{P}_m \rho' \hat{P}_m = w_m P_m \rho_m' \hat{P}_m, \tag{5.24}$$

or

$$\rho' = \sum_n \hat{P}_n \rho' \hat{P}_n. \tag{5.25}$$

We now interpret the principle of minimum disturbance as the requirement that[24] $\|\rho' - \rho\|_2$ should be a minimum subject to Eq. (5.25). Since $\rho = \Sigma_{r,s} \hat{P}_r \rho \hat{P}_s$, and since $\hat{P}_m A \hat{P}_n$ is, with respect to the Hilbert–Schmidt inner product, orthogonal to $\hat{P}_{m'} B \hat{P}_{n'}$, for $m \neq m'$ or $n \neq n'$, i.e.,

$$\langle \hat{P}_m A \hat{P}_n | \hat{P}_{m'} B \hat{P}_{n'} \rangle_2 = (\tau(\hat{P}_n A^* \hat{P}_m \hat{P}_{m'} B \hat{P}_{n'}) = \tau(\hat{P}_{n'} \hat{P}_n A^* \hat{P}_m \hat{P}_{m'} B) = \delta_{nn'} \delta_{mm'} \tau(\hat{P}_n A^* \hat{P}_m B),$$

we have

$$\|\rho - \rho'\|_2^2 = \sum_m \|\hat{P}_m(\rho - \rho')\hat{P}_m\|_2^2 + \sum_{m \neq n} \|\hat{P}_m \rho \hat{P}_n\|_2^2. \tag{5.26}$$

The terms in Eq. (5.26) for $m \neq n$ are fixed, so $\|\rho - \rho'\|_2$ attains its minimum for

$$\|\hat{P}_m(\rho - \rho')\hat{P}_m\|_2 = 0, \quad m = 1,2,3, \ldots, \tag{5.27}$$

and since $\|A\|_2 = 0 \Rightarrow A = 0$,

$$\hat{P}_m(\rho - \rho')\hat{P}_m = 0, \quad m = 1,2,3, \ldots. \tag{5.28}$$

Therefore,

$$\rho' = \sum_m \hat{P}_m \rho \hat{P}_m. \tag{5.29}$$

This is the result achieved by Lüders. It is assumed that no selection is made after the measurement. That is, all elements in the ensemble are measured for A and that the resulting ensemble represented by ρ' consists of all elements of the ensemble (no selection of sub ensemble corresponding to, say, specific values of A). If a selection of the measured ensemble were to be made and the selected elements used to form a new ensemble, then the sum over m in Eq. (5.29) would be restricted to the subset M of $\{1,2,3, \ldots\}$ corresponding to the selected values of A. The resultant ρ' would then have to be normalized [ρ' in Eq. (5.29) is already normalized, $\tau(\rho') = 1$],

$$\rho' = \frac{1}{N} \sum_{n \in M} \hat{P}_m \rho \hat{P}_m, \tag{5.30}$$

where

$$N = \sum_{m \in M} \tau(\hat{P}_m \rho \hat{P}_m) = \sum_{m \in M} \|\sqrt{\rho}\hat{P}_m\|_2^2. \tag{5.31}$$

We may regard Eq. (5.29) as a special case of Eq. (5.30) by selecting M to be all $\{1,2,3, \ldots\}$. The case of selection of a single eigenvalue m in Eq. (5.30) leads to

$$\rho' = (1/N)\hat{P}_m \rho \hat{P}_m, \tag{5.32}$$

which was established by Goldberger and Watson, Wigner, and possibly by earlier authors. Goldberger and Watson argued that even if only a single measurement is made and a_n is obtained, Eq. (5.32) should still be used. This choice was called by Wigner the *moral* or *ethical* choice. (Some might use the term *democratic*. However, the *states* have no vote.)

Many operators (q and p, for example) are unbounded and yet represent observables. We take the view that any self-adjoint operator, bounded or unbounded—with discrete or otherwise spectrum, may represent an observable. This may be debatable. Wigner used to go around asking, "How would you measure the self-adjoint "observable" $\alpha q + \beta p$, α, β real numbers?" The direct measurement of observables is generally beyond present ingenuity, e.g., "How does one measure directly L^2, the total orbital angular momentum?" If A is a self-adjoint operator (unbounded or not), then for any (Borel) set S in the real line R, there exists a projector $\hat{P}_A(S)$ having the properties[26]:

$$\hat{P}_A(\phi) = 0, \quad \hat{P}_A(R) = 1 \tag{5.33}$$

$$\hat{P}_A(S_1)\hat{P}_A(S_2) = \hat{P}_A(S_2)\hat{P}_A(S_1), \tag{5.34}$$

$$\hat{P}_A(S_1 \cup S_2) = \hat{P}_A(S_1) + \hat{P}_A(S_2) - \hat{P}_A(S_1 \cap S_2), \tag{5.35}$$

$$\hat{P}_A(S_1 \cap S_2) = \hat{P}_A(S_1) \wedge \hat{P}_A(S_2) = \hat{P}_A(S_1)\hat{P}_A(S_2). \tag{5.36}$$

If the system is represented by a state ρ, if a measurement of A finds $a \in S$, and if the measurement is of minimal disturbance type, then the same arguments which led to Eq. (5.30)

lead to the following expression for ρ', the state immediately after the measurement:

$$\rho' = (1/N)\hat{P}_A(S)\,\rho\hat{P}_A(S), \tag{5.37}$$

$$N = \tau(\hat{P}_A(S)\,\rho). \tag{5.38}$$

In general measurements are not precise and not of minimal disturbance type. Further, it is sometimes desirable to make measurements on several noncommuting observables. For example, if the observables x, p, position and momentum, are simultaneously measured $\Delta x\Delta p \geqslant \hbar/2$, and usually $\Delta x\Delta p$ is much larger than $\hbar/2$. If a precise measurement yielded x_0, p_0 with $\Delta x\Delta p = \hbar/2$, then the system would be a pure state[27] $|\alpha\rangle$, where $|\alpha\rangle$ is a coherent state centered at x_0, p_0 in phase space and with Δx, Δp the spreads in x and p. Suppose, however, that a measurement determines only that x and p are in some region ω in phase space. We argue that ω should be covered in a minimal way with elementary areas α_K each having $\Delta x\Delta p = \hbar/2$, $\omega \in U\alpha_K$. Then to each elementary area α_K we may assign a coherent state $|\alpha_K\rangle$. We then construct the projector $\hat{P}_\omega$ corresponding to the span of $\{|\alpha_K\rangle\}$. If ρ is the state before the measurement and ρ' the state immediately after, then we expect that

$$\rho' = (1/N)\hat{P}_\omega\,\rho\hat{P}_\omega, \tag{5.39}$$

$$N = \tau(\hat{P}_\omega\,\rho). \tag{5.40}$$

The construction of $\hat{P}_\omega$ is performed using the methods developed in Sec. IV. We first form (assume, for convenience, a finite covering) $\hat{g} \equiv \Sigma_K |\alpha_K\rangle\langle\alpha_K|$, then find $\hat{P}_\omega$ through

$$\hat{P}_\omega = \underset{\epsilon \downarrow 0}{\text{s-lim}}\, \hat{g}_\epsilon^{-1}\,\hat{g}. \tag{5.41}$$

There is a question of lack of uniqueness of such a ρ', which we do not address here. There are other approaches to the question of imprecise or *fuzzy* measurements[28] and to measurements[29] which correspond to areas (volumes for three-dimensional or multiparticle problems). They are not simple and there is always the question of uniqueness—indeed even whether or not one should expect uniqueness.

The question of measurement of position alone (or of momentum alone) is much more clear cut. We consider a single particle moving in three-dimensional Euclidean space E^3, and we neglect other than position degrees of freedom. Such a particle in a pure state may be represented by its Schrödinger wave function $\psi(x,t)$. The time evolution of $\psi(x,t)$ develops via a unitary transformation $\hat{v}(t)$

$$\psi(x,t) = \hat{v}(t)\psi(x,0) \tag{5.42}$$

where $\hat{v}(t)$ is the extension (by continuity) of the operator $e^{-iHt/\hbar}$ and where H is the self-adjoint Hamiltonian. (The operator $e^{-iHt/\hbar}$ may be defined by its Taylor series expansion on a suitable dense set of the Hilbert space H). If V is any (Borel) volume in E^3, the projector $\hat{P}_q(V)$ corresponding to those described by Eqs. (5.33),(5.34),...,(5.36) and representing the self-adjoint position operator q is very simple; $\hat{P}_q(V)$ acts on ψ as a multiplication operator (suppressing t)

$$(\hat{P}_q(V)\psi)(x) = \chi_V(x)\psi(x), \tag{5.43}$$

where $\chi_V(x)$ is the characteristic function for the volume V

$$\chi_V(x) = \begin{cases} 1, & x \in V, \\ 0, & x \notin V. \end{cases} \tag{5.44}$$

This means, of course, in keeping with the standard interpretation, that if the state of the particle is ψ, then the probability $p(V)$ of finding the position q of the particle in V is just

$$p(V) = \langle\psi|\hat{P}_q(V)\psi\rangle = \int_{x\in V} \bar{\psi}(x)\psi(x)\,d^3x. \tag{5.45}$$

If the particle is in a mixed state represented at time t_1 by a density operator $\rho(t_1)$ and if a measurement of minimal disturbance type reveals only that the particle is in the volume V_1, then the state $\rho(t_1^+)$ just after the measurement will be given by

$$\rho(t_1^+) = (1/N_1)\,\hat{P}_q(V_1)\rho(t_1)\,\hat{P}_q(V_1). \tag{5.46}$$

The state will then evolve according to

$$\rho(t) = \hat{v}(t-t_1)\,\rho(t_1^+)\hat{v}^*(t-t_1), \quad t > t_1, \tag{5.47}$$

until a further measurement is made. If at that time $t_2 > t_1$ a measurement is made and reveals the particle to be in V_2, then the state $\rho(t_2^+)$ just after that measurement will be given by

$$\begin{aligned}\rho(t_2^+) &= (1/N_2)\,\hat{P}_q(V_2)\,\rho(t_2)\,\hat{P}_q(V_2) \\ &= (1/N_2N_1)\,\hat{P}_q(V_2)\hat{v}(t_2-t_1)\,\hat{P}_q(V_1)\,\rho(t_1)\,\hat{P}_q(V_1) \\ &\quad\times\hat{v}^*(t_2-t_1)\,\hat{P}_q(V_2).\end{aligned} \tag{5.48}$$

The above development clearly extends to any number of successive measurements.[23]

If a volume V is rotated and translated

$$V \to V' = RV + a \tag{5.49}$$

i.e. $x \in V$ undergoes a Euclidean transformation

$$x \to x' = Rx + a \in V' \tag{5.50}$$

where R is a rotation and a a translation, then there is a unitary $\hat{v}(a,R)$ such that

$$\hat{P}_q(V') = \hat{v}(a,R)\,\hat{P}_q(V)\hat{v}(a,R)^*. \tag{5.51}$$

This expresses the Euclidean invariance of the notion of position measurement. If $\psi' = \hat{v}(a,R)\psi = \hat{v}\psi$, then $\langle\psi'|\hat{P}_q(V')\psi'\rangle$, the probability that a measurement of position for the state ψ' will find the particle in V', is the same as the probability, for the state ψ, that the particle is in V:

$$\begin{aligned}\langle\psi'|\hat{P}_q(V')\psi'\rangle &= \langle\psi'|\hat{v}\,\hat{P}_q(V)\hat{v}^*\psi'\rangle \\ &= \langle\hat{v}^*\psi'|\,\hat{P}_q(V)v^*\psi'\rangle = \langle\psi|\,\hat{P}_q(V)\psi\rangle.\end{aligned} \tag{5.52}$$

The unitary maps $\{\hat{v}(a,R)\}$ form a representation of the Euclidean group in E^3, and the system of projectors $\hat{P}_q(V)$ obeying all of the above conditions is referred to by Wightman as a system of *imprimitivities* (due originally to Mackey). We refer the interested reader to the elegant paper[30] by Wightman, where many further aspects on the general question of localizability in quantum theory may be found.

Projectors have many other applications one of which is to the statistical mechanics of irreversibility. Here one may wish to *project* out of the full time developing statistical operator (or in classical theory, the phase space distribution function) a statistical operator corresponding to a reduced, or substatistical operator (such as a diagonal part). Zwanzig[31] has given a concise treatment of such problems in which he treats classical and quantum mechanical problems on the same footing. We follow his development.

Classical statistical mechanics makes use of a statistical distribution frunction $f(q,p,t)$ where q stands for coordinates $q^1,q^2,...,q^f$, p for canonical momenta $p_1, p_2,..., p_f$, and t is the time. The distribution function is a non-negative function in $L^1(\Omega)$ where Ω is the $2f$-dimensional phase space. It satisfies the Liouville equation

$$\frac{i\partial f}{\partial t} = -i\{f,H\}_{PB} \equiv Lf \tag{5.53}$$

where PB stands for Poisson bracket:

$$\{A,B\}_{PB} \equiv \sum_i \left(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}\right), \tag{5.54}$$

and L is the linear operator defined by Eq. (5.53). If f is a positive element in $L^1(\Omega)$, then $f^{1/2}$ is an element of $L^2(\Omega)$ and conversely if $\psi \in L^2(\Omega)$, $\psi^*\psi = f$ is an element of $L^1(\Omega)$. If $\psi \in L^2(\Omega)$ and satisfies

$$i\frac{\partial \psi}{\partial t} = L\psi, \tag{5.55}$$

then $\psi^*\psi = f$ is in $L^1(\Omega)$ and satisfies Eq. (5.53). Regarded as an operator in $L^2(\Omega)$, L is self-adjoint and e^{-iLt} is unitary. The similarity between ρ and f, $B_1(H)$ and $L^1(\Omega)$, and $B_2(H)$ and $L^2(\Omega)$ is striking. For an operator $\chi \in B_2(H)$ which satisfies

$$\frac{i\partial \chi}{\partial t} = L\chi \equiv \frac{[H,\chi]}{\hbar}, \tag{5.56}$$

where H is the Hamiltonian operator for the system, $[H,\chi]$ the commutator, $\chi^*\chi \equiv \rho$ is a positive operator in $B_1(H)$ and also satisfies

$$\frac{i\partial \rho}{\partial t} = L\rho = \frac{[H,\rho]}{\hbar}. \tag{5.57}$$

On operators $\chi \in B_2(H)$, L may be considered to be self-adjoint and e^{-iLt} unitary. Thus both the classical and quantum mechanical equations of evolution can be written in the form

$$\frac{id\chi}{dt} = L\chi, \tag{5.58}$$

where χ is an element of a Hilbert space [$B_2(H)$ or $L^2(\Omega)$] and L is a self-adjoint operator. The solution to Eq. (5.58) is given by

$$\chi(t) = e^{-itL}\chi(0), \tag{5.59}$$

and L self-adjoint implies that

$$\langle \chi(t)|\chi(t)\rangle = \langle \chi(0)|\chi(0)\rangle, \tag{5.60}$$

for the appropriate inner product.

Now if we are interested in $\hat{P}_1\chi \equiv \chi_1$, where $\hat{P}_1$ is a projector on the Hilber space ($\hat{P}_1$ not necessarily an orthogonal projector), then we may write

$$\chi = \chi_1 + \chi_2, \tag{5.61}$$

where

$$\chi_2 \equiv \hat{P}_2\chi \equiv \chi - \hat{P}_1\chi. \tag{5.62}$$

We take $\hat{P}_i$, $i = 1,2$, to be independent of time so that from Eq. (5.58) it follows that

$$\frac{i\partial \chi_1}{\partial t} = L_{11}\chi_1 + L_{12}\chi_2 \tag{5.63}$$

and

$$\frac{i\partial \chi_2}{\partial t} L_{21}\chi_1 + L_{22}\chi_2, \tag{5.64}$$

where $L_{ij} \equiv \hat{P}_i L \hat{P}_j$, $i,j = 1,2$. The idea now is to eliminate χ_2 and obtain an equation involving χ_1 alone [and, in general, an initial value of χ_2, $\chi_2(0)$]. Let

$$\chi_2 = e^{-iL_{22}t}U_2, \tag{5.65}$$

so that from Eq. (5.64),

$$e^{-iL_{22}t}\frac{i\partial U_2}{\partial t} = L_{21}\chi_1. \tag{5.66}$$

Then

$$\begin{aligned} U_2(t) &= U_2(0) - i\int_0^t e^{iL_{22}\tau}L_{21}\chi_1(\tau)\,d\tau \\ &= \chi_2(0) - i\int_0^t e^{iL_{22}\tau}L_{21}\chi_1(\tau)\,d\tau, \end{aligned} \tag{5.67}$$

and if we substitute χ_2 from Eq. (5.65) using the above result, we find the following integrodifferential equation for χ_1:

$$\begin{aligned} \frac{i\partial \chi_1}{\partial t} &= L_{11}\chi_1 + L_{12}e^{-iL_{22}t}\chi_2(0) \\ &\quad - iL_{12}\int_0^t e^{-iL_{22}\tau}L_{21}\chi_1(t-\tau)\,d\tau. \end{aligned} \tag{5.68}$$

Because $\hat{P}_i\chi_i = \chi_i$ and $\hat{P}_i^2 = \hat{P}_i$, $i = 1,2$, we may write Eq. (5.68) in Zwanzig's form

$$\begin{aligned} \frac{i\partial \chi_1(t)}{\partial t} &= \hat{P}_1L\chi_1(t) + \hat{P}_1Le^{-it\hat{P}_2L}\chi_2(0) \\ &\quad - i\int_0^t d\tau\, P_1Le^{-P_2L\tau}\hat{P}_2L\chi_1(t-\tau)\,d\tau. \end{aligned} \tag{5.69}$$

This equation has been used as a starting point for statistical mechanical theories of irreversibility, although as it stands no irreversibility has yet been introducted.

Measurement theory can also be considered from the standpoint of the quantum theory of scattering in which projectors and isometries play an important role. We content ourselves to refer the reader to some modern references.[32]

VI. CONFIGURATION SPACE AND SECOND QUANTIZATION

Fock in a classic paper[33] showed clearly and explicitly the relationship between configuration space and the space of second quantization for a system of identical elementary particles. We show this same relationship through the use of simple mapping and projector techniques. Although no new results are obtained, new insights and procedures are obtained which can be applied to more complicated situations. Further, Fermi–Dirac and Bose–Einstein particles are treated on the same basis, so a single formalism applies to both.

Configuration space in quantum theory refers to the positions (and spins) x of particles in a given system. Fock space F refers to a Hilbert space consisting of sequences $\psi \equiv \{\psi_n(x_1,...,x_n)\}_{n=0}^{\infty}$ of wave functions (ψ_0 = a complex number). The inner product in F is given by

$$\langle\psi_1|\psi_2\rangle = \sum_{n=0}^{\infty}\int\cdots\int dx_n\ \bar{\psi}_{1n}(x_1,\ldots,x_n)\psi_{2n}(x_1,\ldots,x_n)$$

$$= \sum_{n=0}^{\infty}\langle\psi_{1n}|\psi_{2n}\rangle. \tag{6.1}$$

We may generate F from a complete orthonormal set of one-particle wave functions $\phi_i(x)\equiv\phi(i)$, $i = 1,2,3,\cdots,$. Set

$$|ni\rangle = |i_1,\ldots,i_n\rangle\equiv\phi(i_1)\ldots\phi(i_n), \tag{6.2}$$

where the arguments in the function have been suppressed for printing convenience. So $\phi(i_1,\ldots,i_n)$ is a function of $x_1,x_2,\ldots,x_n$, and for $n = 0$, $\phi_0\equiv 1$. The set of functions $|ni\rangle$ are orthonormal and complete. We represent the completeness by

$$\sum_{ni}|ni\rangle\langle ni| = 1_F.$$

Matrix elements of operators are most simply computed using this fully unsymmetrized basis $\{|ni\rangle\}$, whereas physical symmetry can be handled simply and automatically through the use of second quantization.

Let S (space of second quantization) be the Hilbert space generated from a unique vacuum state $|0\rangle$ by the application of the usual particle creation operators $a(i)^*$. The operator $a(i)^*$ creates a particle in the state whose wave function is $\phi(i)$. The operators $a(i)^*$ and their adjoints $a(i)$ satisfy canonical commutation relations for bosons and anticommutation relations for fermions:

$$[a(i),a(j)^*]_{\mp} = \delta(i,j). \tag{6.3}$$

We introduce a set of vectors in S which is overcomplete but which has very simple properties. Let

$$|n,i;S\rangle = |i_1,\ldots,i_n;S\rangle\equiv(1/\sqrt{n!})a(i_1)^*a(i_2)^*\cdots a(i_n)^*|0\rangle. \tag{6.4}$$

Then the inner product of two such vectors is given by

$$\langle n,i;S|n',i';S\rangle = \delta_{nn'}\delta_n^S(i,i') \tag{6.5}$$

where δ_n^S is the completely symmetrized Kronecker δ for bosons:

$$\delta_n^S(i,i') = \frac{1}{n!}\sum_P\delta(i_1,Pi_1')\delta(i_2,Pi_2')\cdots\delta(i_n,Pi_n'), \tag{6.6}$$

and the completely antisymmetrized Kronecker δ for Fermions:

$$\delta_n^S(i,i') = \frac{1}{n!}\sum_P\epsilon(P)\delta(i_1,Pi_1')\delta(i_2,Pi_2')\cdots\delta(i_n,Pi_n'). \tag{6.7}$$

In the above two expressions the sum Σ_P goes over all permutation P of $(1,2,\ldots,n)$ and $\epsilon(P)$ is the signature of P. The vectors $|n,i;S\rangle$ have the nice property[34] that

$$\sum_i|n,i;S\rangle\langle n,i;S| = \hat{P}_n, \tag{6.8}$$

where $\hat{P}_n$ is the projector on the n-particle subspace of S. Clearly[34]

$$\sum_{n=0}^{\infty}\hat{P}_n = 1_S. \tag{6.9}$$

The projectors $\hat{P}_n$ have the following properties with respect to $a(i)$ and $a(i)^*$:

$$a(i)\hat{P}_n = \hat{P}_{n-1}a(i) \tag{6.10}$$

and

$$a(i)^*\hat{P}_n = P_{n+1}a(i)^*, \tag{6.11}$$

as is directly verified by using elements of the defining set $\{|ni;S\rangle\}$. Since every vector $|S\rangle$ in S is a linear combination of the $|ni;S\rangle$, we see that $|S\rangle$ satisfies

$$\hat{P}_S|S\rangle = |S\rangle, \tag{6.12}$$

where $\hat{P}_S$ is defined by its action on $|ni;S\rangle$:

$$\hat{P}_S|n,i;S\rangle\equiv\frac{1}{n!}\sum_P\lambda_P P|ni;S\rangle$$

$$= \frac{1}{n!}\sum_P\lambda_P|Pi_1,Pi_2,\ldots,Pi_n;S\rangle, \tag{6.13}$$

where $\lambda_P = 1$ for bosons and $\lambda_P = \epsilon(P)$ for fermions. Thus

$$\hat{P}_S|n,i;S\rangle = |n,i;S\rangle, \tag{6.14}$$

and Eq. (6.12) follows.

We now introduce a map Ω^*: $F\to S$, by means of the expression:

$$Q^* = \sum_{n,i}|n,i;S\rangle\langle n,i|. \tag{6.15}$$

The adjoint $(\Omega^*)^* = \Omega$ is given by

$$\Omega = \sum_{n,i}|n,i\rangle\langle n,i;S|. \tag{6.16}$$

We observe that Ω^* is onto S and Ω is into F. It is established directly from Eqs. (6.15) and (6.16) that

$$\Omega^*\Omega = 1_S. \tag{6.17}$$

That is Ω is an isometry from S to F. Therefore, Ω^* is a partial isometry from F to S

$$(\Omega^*)^*\Omega^* = \Omega\Omega^* = \hat{P}, \tag{6.18}$$

where $\hat{P}$ is a projector, the projector onto the completely symmetric (bosons) or completely antisymmetric (fermions) subspace of F. The projector $\hat{P}$ satisfies

$$\hat{P}\Omega = \Omega \tag{6.19}$$

and

$$\Omega^*\hat{P} = \Omega^*. \tag{6.20}$$

Thus $\hat{P}$ projects F onto the physical states in F. That is

$$\hat{P} = \sum_n\frac{1}{n!}\sum_{P\in S_n}\lambda_P P, \tag{6.21}$$

S_n being the symmetric permutation group on n elements.

If $|\psi F\rangle$ is a physical state in F, $\hat{P}|\psi F\rangle = |\psi F\rangle$, or

$$|\psi F\rangle = \Omega(\Omega^*|\psi F\rangle) = \Omega|\psi S\rangle \tag{6.22}$$

where

$$|\psi S\rangle = \Omega^*|\psi F\rangle. \tag{6.23}$$

Conversely, if $|\psi S\rangle$ is any vector in S,

$$|\psi F\rangle = \Omega|\psi S\rangle \tag{6.24}$$

is a vector in F; in fact, $|\psi F\rangle$ is in $\hat{P}F$, the physical subspace of F. That is the case follows from

$$\hat{P}|\psi F\rangle = \hat{P}\Omega|\psi S\rangle = \Omega|\psi S\rangle = |\psi F\rangle, \tag{6.25}$$

where we have used Eq. (6.19).

We observe that $\hat{PF}$ and S are unitarily equivalent. That is, if Ω^* is restricted to $\hat{PF}$, it becomes a unitary map from $\hat{PF}$ to S. This equivalence between $\hat{PF}$ and S expresses concisely the relationship between configuration space and the space of second quantization.

If $|\psi Ft\rangle$ is a physical state in F which satisfies Schrödinger's equation

$$i\hbar\frac{\partial}{\partial t}|\psi Ft\rangle = H_F|\psi Ft\rangle, \tag{6.26}$$

then, since $\hat{P}|\psi Ft\rangle = \Omega\Omega^*|\psi Ft\rangle$,

$$i\hbar\frac{\partial}{\partial t}\{\Omega^*|\psi Ft\rangle\} = (\Omega^* H_F \Omega)\{\Omega^*|\psi Ft\rangle\}, \tag{6.27}$$

or

$$i\hbar\frac{\partial}{\partial t}|\psi St\rangle = H_S|\psi St\rangle, \tag{6.28}$$

where $|\psi St\rangle \equiv \Omega^*|\psi Ft\rangle$ is the second quantized expression for the time evolving state. The Hamiltonian in second quantized form is just

$$H_S = \Omega^* H_F \Omega \tag{6.29}$$

and H_F may have any form on its domain in F so long as it is self-adjoint. That is, H_F may create or destroy particles. The second quantized form for other operators may be obtained in a similar manner. Note that Ω^* is not restricted to $\hat{PF}$ in Eq. (6.29) which makes the computation of H_S simple, since H_S may be expressed in terms of matrix elements of H_F between states $|in\rangle$:

$$H_S = \sum_{ni}\sum_{n'i'}\frac{1}{\sqrt{n!n'!}}a(i_1)^*\cdots a(i_n)^*\hat{P}_0 \times\langle i_1,\ldots,i_n|H_F|i'_1,\ldots,i'_{n'}\rangle a(i'_{n'})\cdots a(i'_1), \tag{6.30}$$

where $\hat{P}_0 = |0\rangle\langle 0|$ is the projector on the vacuum state $|0\rangle$ of S. The matrix elements $\langle i_1,\ldots,i_n|H_F|i'_1,\ldots,i'_{n'}\rangle$ are simply expressed, using Eq. (6.2):

$$\langle i_1,\ldots,i_n|H_F|i'_1,\ldots,i_{n'}\rangle = \int\cdots\int dx_1\cdots dx_n\, dx'_1\cdots dx_{n'} \times\bar{\phi}_{i_1}(x_1)\cdots\bar{\phi}_{i_n}(x_n)H_F\phi_{i'_1}(x'_1)\cdots\phi_{i'_{n'}}(x_{n'}). \tag{6.31}$$

The second quantized form of H given by Eq. (6.30) is extremely general. Let us examine a fairly general, but typical H_F, which is the sum of $H_F(l)$, $l = 1,\ldots,m$, each $H_F^{(l)}$ preserving the number of particles,

$$H_F^{(l)}|i(1),i(2),\ldots,i(n)\rangle = \sum_{k_1<k_2<\cdots<k_l}\sum_{i(k_1)',i(k_2)',\ldots,i(k_l)'} |i(1),i(2),\ldots,i(k_1)',\ldots,i(k_2)',\ldots,i(k_l)',\ldots,i(n)\rangle \times\langle i(k_1)',i(k_2)',\ldots,i(k_l)'|V^{(l)}|i(k_1),\ldots,i(k_l)\rangle, \tag{6.32}$$

where $V^{(l)}$ acts only on the $k_1,\ldots,k_l$ variables. The above equation looks somewhat complicated, but it has a simple structure. We write $i(k)$ rather than i_k in order to avoid double subscripts in printing.

Before proceeding to the general case, we treat the case $l = 1$ in complete detail. The general case is algebraically more complicated, but is not complicated in principle, and can be carried out in a similar way. For $l = 1$, $H_F = \Sigma_r V(r)$, where $V(r)$ acts only on the coordinates and spin x_r. Thus the matrix element in Eq. (6.32) becomes $\langle i(k'_1)|V^1|i(k_1)\rangle$ and Eq. (6.32), for $l = 1$, becomes

$$H_F^{(1)}|i(1),i(2),\ldots,i(n)\rangle = \sum_{k_1=1}^{n}\sum_{i(k_1)'}|i(1),i(2),\ldots,i(k_1)',\ldots,i(n)\rangle\langle i(k_1)'|V^1|i(k_1)\rangle,$$

where $i(k_1)'$ is in the k_1th place. From the above result it follows that $[i(k)\equiv i_k]$

$$\langle i_1,\ldots,i_n|H_F^{(1)}|j_1,\ldots,j_n\rangle = \sum_{k=1}^{n}\sum_{j_k'}\delta(i_1,j_1)\delta(i_2,j_2)\cdots \times\delta(i_k,j'_k)\cdots\delta(i_n,j_n)\langle j'_k|V^1|j_k\rangle.$$

Therefore, from Eq. (6.30) one obtains

$$H_S^{(1)} = \sum_{n,i}\frac{1}{n!}\sum_{k=1}^{n}\sum_{j_k}a(i_1)^*\cdots a(i_k)^*\cdots a(i_n)^*\hat{P}_0 \times\langle i_k|V^1|j_k\rangle a(i_n)\cdots a(j_k)\cdots a(i_1).$$

Since the a^* all commute (bosons) or all anticommute (fermions) and since $a(i_k)^*$ and $a(j_k)$ are similarly positioned relative to $a(i_1)^*$, resp. $a(i_1)$, each of the terms in the sum over k is the same. Hence

$$H_S^{(1)} = \sum_{i,j}\sum_{n>1}a(i)^*\langle i|V^1|j\rangle \times\sum_{i_2,\ldots,i_n}\frac{1}{(n-1)!}a(i_2)^*\cdots a(i_n)^*\hat{P}_0 a(i_n)\cdots a(i_2)a(j).$$

However,

$$\frac{1}{(n-1)!}\sum_{i_2,\ldots,i_n}a(i_2)^*\cdots a(i_n)^*\hat{P}_0 a(i_n)\cdots a(i_2)$$

is just $\hat{P}_{n-1}$, the projector on the $(n-1)$-particle subspace of S, and since $\Sigma_{n>1}\hat{P}_{n-1} = \hat{1}_S$, we have the well-known result

$$H_S^{(1)} = \sum_{i,j}a(i)^*\langle i|V^1|j\rangle a(j).$$

This method of deriving results in second quantized form is direct, unambiguous and purely algebraic. All combinatorics are incorporated in the foundations so all one need do is "turn the crank." We now proceed to the general $H^{(l)}$ case.

We now use Eq. (6.32) in Eq. (6.30), use the orthogonality of $|ni\rangle$, and collect $n(n-1)\cdots(n-l+1)/l! = \binom{n}{l}$ equivalent terms to obtain

$$H_S^{(l)} = \sum_{i(1),\ldots,i(n),n}\binom{n}{l}\frac{1}{n!}a[i(1)]^*a[i(2)]^*\cdots \times a[i(n)]^*\hat{P}_0 a[i(n)]\cdots a[i(l+1)] \times a[i(l)']a[i(l-1)']\cdots a[i(1)'] \times\langle i(1),\ldots,i(l)|V^{(l)}|i(1)',i(2)',\ldots,i(l)'\rangle. \tag{6.33}$$

Next, observe that

$$\hat{P}_{n-l} \equiv \frac{1}{(n-l)!} \sum_{i(l+1),\ldots,i(n)} a[i(l+1)]^* \cdots a[i(n)]^* \times \hat{P}_0 a[i(n)] \cdots a[i(l+1)] \tag{6.34}$$

is just the projector on the $(n-l)$-particle subspace of S and that, since $a^* \hat{P}_{n-l} = \hat{P}_{n-l+1} a^*$, etc., Eq. (6.34) becomes

$$H_S^{(l)} = \frac{1}{l!} \sum_{i_1,\ldots,i_l;i_1',i_2',\ldots,i_l'} a(i_1)^* \cdots a(i_l)^* \times \langle i_1,\ldots,i_l | V^{(l)} | i_1',\ldots,i_l' \rangle a(i_l') \cdots a(i_1'). \tag{6.35}$$

This well-known result, Eq. (6.35), holds for bosons and fermions alike.

In particular, for $l = 2$,

$$H_S^{(2)} = \frac{1}{2} \sum a_l^* a_m^* \langle lm | V^{(2)} | l'm' \rangle a_{m'} a_{l'}, \tag{6.36}$$

where

$$\langle lm | V^{(2)} | l'm' \rangle = \int\int dx\, dy\, \bar{\phi}_l(x) \bar{\phi}_m(y) V(x,y) \times \phi_{l'}(x) \phi_{m'}(y). \tag{6.37}$$

We have shown that the relation between configuration space and the space of second quantization is effected by a simple map Ω^* from the Fock space F to the space of second quantization S and that S is isometrically isomorphic to a subspace $P = \Omega\Omega^* F$ of F. The subspace P consists of completely symmetrized (bosons) or completely antisymmetrized (fermions) Fock vectors. The mapping then allows a simple transformation to be made from physical expressions in F to physical expressions in S.

VII. GENERALIZED WANNIER AND BLOCH FUNCTIONS

In this section we will discuss symmetrical orthogonalization as developed by Löwdin[35] and later by others.[36] Surveys on this procedure are given by Löwdin[37] and more recently by Jørgensen.

Consider a set of states $|\nu\rangle$, $\nu = 1,\ldots,M$. The vectors are normalized $\langle \nu|\nu \rangle = 1$, but are not necessarily linearly independent. The index ν may be considered as specifying localization sites and the bound states at those sites. The closed span of these vectors is denoted by M. In the manifold M, we define the bounded strictly positive operator

$$\hat{g} \equiv \sum_{\nu=1}^{M} |\nu\rangle\langle\nu|. \tag{7.1}$$

According to the results in Sec. IV,

$$\text{s-lim}_{\epsilon \downarrow 0} (\hat{g} + \epsilon)^{-1} \hat{g} = \hat{P}_M, \tag{7.2}$$

where $\hat{P}_M$ is the projector onto M. Further, if we define the vectors $|\tilde{\nu}|$ by means of

$$|\tilde{\nu}\rangle = \lim_{\epsilon \downarrow 0} (\hat{g} + \epsilon)^{-1/2} |\nu\rangle, \tag{7.3}$$

then

$$\sum_{\nu=1}^{M} |\tilde{\nu}\rangle\langle\tilde{\nu}| = \hat{P}_M. \tag{7.4}$$

If the vectors $|\nu\rangle$ are linearly independent, then we know that the set $\{|\tilde{\nu}\rangle\}_{\nu=1}^{M}$ is an orthonormal set. In this case we call the vectors $|\tilde{\nu}\rangle$ generalized Wannier functions. They are the usual Wannier functions of solid state physics when the sites correspond to a periodic lattice, whereas for a few sites the $|\tilde{\nu}\rangle$ are atomic orbitals.

As we previously pointed out, the Schmidt process for orthogonalization of a set of vectors is *undemocratic*, being highly dependent upon the order of vectors in the selection. In symmetric orthogonalization, the overlap matrix (Gram matrix) with elements $\langle \nu|\nu' \rangle$ plays a central role. If we set

$$|\phi\rangle \equiv \sum_{\nu=1}^{M} |\nu\rangle \phi(\nu), \tag{7.5}$$

then the inner product of two such vectors $|\phi\rangle$, $|\phi'\rangle$ is given by

$$\langle \phi | \phi' \rangle = \sum_{\nu,\nu'=1}^{M} \bar{\phi}(\nu) \langle \nu|\nu' \rangle \phi'(\nu'), \tag{7.6}$$

which we will use later.

Now let $f_\lambda(\nu)$ be the orthonormalized eigenvectors of the overlap matrix:

$$\sum_{\nu'=1}^{M} \langle \nu|\nu' \rangle f_\lambda(\nu') = m_\lambda f_\lambda(\nu), \tag{7.7}$$

$\lambda = 1,\ldots,M$, $\nu = 1,\ldots,M$. Since $(\langle \nu|\nu' \rangle)$ is a positive semidefinite Hermitian matrix, $m_\lambda \geqslant 0$. The f_λ satisfy

$$\sum_{\nu=1}^{M} \bar{f}_\lambda(\nu) f_{\lambda'}(\nu) = \delta(\lambda, \lambda') \tag{7.8}$$

(orthonormality) and

$$\sum_{\lambda=1}^{M} f_\lambda(\nu) \bar{f}_\lambda(\nu') = \delta(\nu, \nu') \tag{7.9}$$

(completeness).

In Eq. (7.5) we put $\phi(\nu) = f_\lambda(\nu) c_\lambda$ and define the corresponding $|\phi\rangle$ to be $|\lambda\rangle$ [i.e., $|\lambda\rangle = \Sigma_{\nu=1}^{M} |\nu\rangle f_\lambda(\nu) c_\lambda$]. Then Eq. (7.6) becomes

$$\langle \lambda | \lambda' \rangle = \sum_{\nu,\nu'=1}^{M} \bar{c}_\lambda \bar{f}_\lambda(\nu) \langle \nu|\nu' \rangle f_{\lambda'}(\nu') c_{\lambda'} = \sum_{\nu} \bar{c}_\lambda c_{\lambda'} \bar{f}_\lambda(\nu) m_\lambda f_{\lambda'}(\nu) = m_\lambda |c_\lambda|^2 \delta(\lambda, \lambda'). \tag{7.10}$$

Vectors $|\lambda\rangle$ for which $m_\lambda = 0$ are null vectors, since from Eq. (7.10) $\| |\lambda\rangle \|^2 = 0$, and therefore $|\lambda\rangle = 0$. Let the null space of the overlap matrix have dimension d, and label $m_\lambda = 0$ by $\lambda = 1,\ldots,d$. Then

$$\sum_{\nu=1}^{M} |\nu\rangle f_\lambda(\nu) = 0, \quad \lambda = 1, \ldots, d. \tag{7.11}$$

Each nonzero eigenvalue m_λ of the overlap matrix gives rise to a normalized state $|\lambda\rangle$ in M given by choosing $c_\lambda = m_\lambda^{-1/2}$;

$$|\lambda\rangle = \sum_{\nu=1}^{M} |\nu\rangle f_\lambda(\nu) m_\lambda^{-1/2}, \quad \lambda = d+1, \ldots, M. \tag{7.12}$$

The orthonormal set of vectors $|\lambda\rangle$ defined by Eq. (7.12) will be called Bloch vectors (or functions) in the generalized sense. For periodically arranged sites these vectors are the usual Bloch functions of solid state physics while for a few sites they are the molecular orbitals.

The states $|\lambda\rangle$, $M-d$ in number are complete in M, so any $|\psi\rangle \in M$ has the expansion

$$|\psi\rangle = \sum_{\lambda=d+1}^{M} |\lambda\rangle\langle\lambda|\psi\rangle. \tag{7.13}$$

From Eq. (7.12),

$$\langle\nu|\lambda\rangle = m_\lambda^{1/2} f_\lambda(\nu), \tag{7.14}$$

and from Eq. (7.13),

$$|\nu\rangle = \sum_{\lambda=d+1}^{M} |\lambda\rangle\langle\lambda|\nu\rangle = \sum_{\lambda=d+1}^{M} |\lambda\rangle \bar{f}_\lambda(\nu) m_\lambda^{1/2}. \tag{7.15}$$

It then follows directly that

$$\begin{aligned}|\tilde{\nu}\rangle &= \lim_{\epsilon \downarrow 0} (\hat{g}+\epsilon)^{-1/2}|\nu\rangle \\ &= \lim_{\epsilon \downarrow 0} \sum_{\lambda=d+1}^{M} \frac{1}{\sqrt{m_\lambda+\epsilon}} |\lambda\rangle\langle\lambda|\nu\rangle \\ &= \sum_{\lambda=d+1}^{M} |\lambda\rangle \bar{f}_\lambda(\nu), \quad \nu = 1, \ldots, M. \end{aligned} \tag{7.16}$$

The above equations may be inverted with the result

$$|\lambda\rangle = \sum_{\nu=1}^{M} |\tilde{\nu}\rangle f_\lambda(\nu), \quad \lambda = d+1, \ldots, M. \tag{7.17}$$

Therefore,

$$\sum_{\nu=1}^{M} |\tilde{\nu}\rangle\langle\tilde{\nu}| = \sum_{\lambda=d+1}^{M} |\lambda\rangle\langle\lambda| = \hat{P}_M \tag{7.18}$$

and

$$\langle\tilde{\nu}|\tilde{\nu}'\rangle = \sum_{\lambda=d+1}^{M} f_\lambda(\nu)\bar{f}_\lambda(\nu'). \tag{7.19}$$

We observe that

$$\sum_{\nu} \langle\tilde{\nu}|\tilde{\nu}'\rangle\langle\tilde{\nu}'|\tilde{\nu}''\rangle = \langle\tilde{\nu}|\hat{P}_M\tilde{\nu}''\rangle = \langle\tilde{\nu}|\tilde{\nu}''\rangle,$$

so that $(\langle\tilde{\nu}|\tilde{\nu}'\rangle)$ is the (matrix) projector on the support of the overlap matrix.

If the set $\{|\nu\rangle\}$ is linearly independent ($d=0$), then Eq. (7.19) becomes

$$\langle\tilde{\nu}|\tilde{\nu}'\rangle = \delta(\nu,\nu'), \tag{7.20}$$

and Eqs. (7.16) and (7.17) become the statements of the isometric transformation between the $\{|\tilde{\nu}'\rangle\}$ basis (Wannier) and the $\{|\lambda\rangle\}$ basis (Bloch) for M.

We now illustrate that we have given the appropriate names (Wannier and Bloch) to these vectors. Consider the example of bound states $|\alpha_i, \mathbf{R}_i\rangle$, localized at N sites $\mathbf{R}_i$, $i=1,\ldots,N$, in a periodic lattice of volume V. The indices α_i each ranges from 1 to b, so our previous index ν corresponds to the pair $(\alpha_i, \mathbf{R}_i)$ and $M = bN$ is the total number of states. We suppose that the bound state wave functions satisfy periodic boundary conditions over the volume V. That is,

$$\psi_{(\alpha_i,\mathbf{R}_i)}(\mathbf{r}) = \langle\mathbf{r}|\alpha_i\mathbf{R}_i\rangle = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \phi_{\alpha i}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_i)}, \tag{7.21}$$

where $\mathbf{k}$ is the usual wave vector for periodic boundary conditions (over the entire volume V). At a given site $\mathbf{R}_i$ we assume orthonormality,

$$\langle\alpha, \mathbf{R}_i|\alpha', \mathbf{R}_i\rangle = \sum_{\mathbf{k}} \bar{\phi}_\alpha(\mathbf{k})\phi_{\alpha'}(\mathbf{k}) = \delta(\alpha,\alpha'). \tag{7.22}$$

The overlap matrix elements are given by

$$\begin{aligned}\langle\alpha, \mathbf{R}_i|\alpha', \mathbf{R}_{i'}\rangle &= \sum_{\mathbf{k}} \bar{\phi}_\alpha(\mathbf{k})\phi_{\alpha'}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_{i'})} \\ &= \sum_{\mathbf{k}\in \mathrm{FBZ}} (M(k))_{\alpha\alpha'} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_{i'})} \end{aligned} \tag{7.23}$$

where FBZ means the first Brillouin zone and

$$(M(\mathbf{k}))_{\alpha\alpha'} \equiv \sum_{\mathbf{K}} \bar{\phi}_\alpha(\mathbf{k}+\mathbf{K})\phi_{\alpha'}(\mathbf{k}+\mathbf{K}), \tag{7.24}$$

where $\mathbf{K}$ runs over the reciprocal lattice vectors. The components $f_{\mathbf{k}\beta}(\alpha\mathbf{R}_i)$ of the eigenvectors of the overlap matrix are given by

$$f_{\mathbf{k}\beta}(\alpha\mathbf{R}_i) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{R}_i} g_{\mathbf{k}\beta}(\alpha), \quad \mathbf{k}\in\mathrm{FBZ}, \tag{7.25}$$

where $g_{\mathbf{k}\beta}$ satisfies

$$\sum_{\alpha'=1}^{b} (M(\mathbf{k}))_{\alpha\alpha'} g_{\mathbf{k}\beta}(\alpha') = m_{\mathbf{k}\beta} g_{\mathbf{k}\beta}(\alpha), \quad \beta = 1, \ldots, b. \tag{7.26}$$

The $f_{\mathbf{k}\beta}$ are orthonormal,

$$\begin{aligned}\sum_{\alpha i} \bar{f}_{\mathbf{k}\beta}(\alpha\mathbf{R}_i) f_{\mathbf{k}'\beta'}(\alpha, \mathbf{R}_i) &= \delta(\mathbf{k},\mathbf{k}') \sum_{\alpha=1}^{b} \bar{g}_{\mathbf{k}\beta}(\alpha) g_{\mathbf{k}\beta'}(\alpha) \\ &= \delta(\mathbf{k},\mathbf{k}')\delta(\beta,\beta'), \end{aligned} \tag{7.27}$$

and complete,

$$\begin{aligned}\sum_{\mathbf{k}\beta} f_{\mathbf{k}\beta}(\alpha\mathbf{R}_i)\bar{f}_{\mathbf{k}\beta}(\alpha', \mathbf{R}_{i'}) &= \delta(\alpha,\alpha') \sum_{\mathbf{k}} \frac{1}{N} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_{i'})} \\ &= \delta(\alpha,\alpha')\delta(\mathbf{R}_i, \mathbf{R}_{i'}). \end{aligned} \tag{7.28}$$

Because of the linear independence of the $|\alpha_i\mathbf{R}_i\rangle \equiv |\alpha_i, i\rangle$, none of the $m_{\mathbf{k}\beta}$ vanish. Hence from Eq. (7.12) we find the orthonormal and complete set of vectors

$$|\mathbf{k}\beta\rangle = \sum_{\alpha,i} |\alpha i\rangle \frac{1}{\sqrt{N m_{\mathbf{k}\beta}}} e^{i\mathbf{k}\cdot\mathbf{R}_i} g_{\mathbf{k}\beta}(\alpha), \quad \mathbf{k}\in\mathrm{FBZ}. \tag{7.29}$$

These vectors clearly have the Bloch form. On the other hand, by Eq. (7.16)

$$|\tilde{\alpha}\mathbf{R}_i\rangle = \sum_{\beta,\mathbf{k}\in\mathrm{FBZ}} |\mathbf{k}\beta\rangle \frac{1}{\sqrt{N}} e^{-i\mathbf{k}\cdot\mathbf{R}_i} \bar{g}_{\mathbf{k}\beta}(\alpha), \tag{7.30}$$

which can be recognized as the original Wannier[38] form, since, for a single band $b=1$, $\bar{g}_{\mathbf{k}\beta}(\alpha)$ is simply $\delta(\beta,\alpha)$.

Further insight into the construction of $|\tilde{\alpha}\mathbf{R}_i\rangle$ is obtained by starting from definition, Eq. (7.3), for the tilded states. We first find the result of applying $\hat{g}$ to the vectors $|\alpha\mathbf{R}_i\rangle$:

$$\begin{aligned}\hat{g}|\alpha\mathbf{R}_i\rangle &= \sum_{\alpha' i'} |\alpha'\mathbf{R}_{i'}\rangle\langle\alpha'\mathbf{R}_{i'}|\alpha\mathbf{R}_i\rangle \\ &= \sum_{\alpha' i'} |\alpha'\mathbf{R}_{i'}\rangle \frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}_{i'}-\mathbf{R}_i)} (M(\mathbf{k}))_{\alpha'\alpha}. \end{aligned} \tag{7.31}$$

Let $M(\mathbf{k})^{-1/2}$ be the inverse of the matrix $M(\mathbf{k}) = (M(\mathbf{k})_{\alpha\alpha'})^{1/2}$. Then

$$|\tilde{\alpha}\mathbf{R}_i\rangle = \hat{g}^{-1/2}|\alpha\mathbf{R}_i\rangle$$
$$= \sum_{\alpha' i'} |\alpha'\mathbf{R}_{i'}\rangle \frac{1}{N}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}_{i'}-\mathbf{R}_i)}(M(\mathbf{k})^{-1/2})_{\alpha'\alpha}. \tag{7.32}$$

Next expand $M^{-1/2}$ in terms of the eigenvectors of M and make use of Eq. (7.3) to obtain

$$|\tilde{\alpha}\mathbf{R}_i\rangle = \sum_{\alpha'} |\alpha'\mathbf{R}_{i'}\rangle \frac{1}{N}\sum e^{i\mathbf{k}\cdot(\mathbf{R}_{i'}-\mathbf{R}_i)} g_{\mathbf{k}\beta}(\alpha') m_{\mathbf{k}\beta}^{-1/2} \bar{g}_{\mathbf{k}\beta}(\alpha)$$
$$\equiv \sum_{\mathbf{k}\beta} |\mathbf{k}\beta\rangle \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{R}_i}\bar{g}_{\mathbf{k}\beta}(\alpha). \tag{7.33}$$

We emphasize that Eqs. (7.29) and (7.32) are in particularly simple form, this being due to the fact that the $(bN)\times(bN)$ matrix problem reduces to just a $(b)\times(b)$ matrix problem. This reduction was possible because the sites were periodically arrayed and because we imposed periodic boundary conditions over the entire volume V. The first condition precludes disordered systems, and the second condition precludes the consideration of surface states. If either of these restrictions is removed, we return directly to the full $(bN)\times(bN)$ matrix problem. In that case, the direct calculation of Wannier functions by means of Eq. (7.3) is more efficient, since the operator $(\hat{g}+\epsilon)^{-1/2}$ can be expanded systematically and simply in a cluster expansion.

As an example of the direct use of Eq. (7.3), we consider now the computation of the Wannier functions for an infinite linear lattice of Gaussians. This problem was studied in detail by Wannier.[39] Our result differs from that of Wannier in that the Wannier functions studied here are reflection-invariant about their respective localization sites. This is brought about because the phase is fixed by the symmetrical orthogonalization process.

Consider the normalized bound states $|i\rangle$ localized at $R_i = ai, i = 0, \pm 1, \pm 2, \ldots$, with wave functions ψ_i given by

$$\psi_i(x) = \langle x|i\rangle = \gamma^{1/4} e^{-\pi\gamma(x-R_i)^2/2}$$
$$\equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, \phi(k) e^{ik(x-R_i)}, \tag{7.34}$$

so that the k-space wave functions $\langle k|i\rangle$ are given by

$$\langle k|i\rangle = \phi(k)e^{ikR_i} = e^{-(1/2\pi\gamma)k^2} e^{-ikR_i}/\sqrt{\pi}\,\gamma^{1/4}. \tag{7.35}$$

Then

$$\langle k|\hat{g}|i\rangle = \left\langle k\left|\sum_{i'}\right|i'\right\rangle\langle i'|i\rangle\rangle = \sum_{i'}\langle k|i'\rangle\langle i'|i\rangle$$
$$= \phi(k)\sum_{i'=-\infty}^{\infty} e^{-ikR_{i'}}\int_{-\infty}^{\infty} dk'\,|\phi(k')|^2 e^{ik'(R_{i'}-R_i)}$$
$$= \phi(k)\frac{2\pi}{a}\sum_{n=-\infty}^{\infty}\left|\phi\left(k+\frac{2\pi n}{a}\right)\right|^2 e^{-ikR_i}, \tag{7.36}$$

where the last line results from performing the sum over the lattice in the convergence strip of ϕ. We have, therefore,

$$\langle k|\tilde{i}\rangle = \langle k|\hat{g}^{-1/2}|i\rangle$$
$$= \phi(k)e^{-ikR_i}\left(\frac{2\pi}{a}\sum_n\left|\phi\left(k+\frac{2\pi n}{a}\right)\right|^2\right)^{-1/2}. \tag{7.37}$$

This result may now be expressed in terms of the theta function Θ_3, since[40]

$$\sum_{n=-\infty}^{\infty}\left|\phi\left(k+\frac{2\pi n}{a}\right)\right|^2$$
$$= \frac{1}{\sqrt{\pi}\,\gamma^{1/4}}\sum_{n=-\infty}^{\infty} e^{-(1/2\pi\gamma)(k+2\pi n/a)^2}$$
$$= \frac{e^{-k^2/\pi\gamma}}{\sqrt{\pi}\gamma^{1/4}}\Theta_3\left(i\frac{2k}{\pi a\gamma}\middle|\frac{4i}{a^2\gamma}\right). \tag{7.38}$$

Hence

$$\tilde{\psi}_i(x) = \langle x|\tilde{i}\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\langle k|i\rangle e^{ikx}dk$$
$$= \frac{a^{1/2}}{2\pi}\int_{-\infty}^{\infty}\frac{e^{ik(x-R_i)}}{[\Theta_3(i2k/\pi a\gamma|4i/a^2\gamma)]^{1/2}}. \tag{7.39}$$

The integral in Eq. (7.39) is invariant under inversion in k of the integrand, so $\tilde{\psi}_i(x)$ is invariant under reflection of x about R_i. The orthonormality of the Wannier functions $\tilde{\psi}_i$ is assured since the infinite lattice of gaussians forms a linearly independent set.

Another simple, but important, example consists of the two site problem first treated in the case of the hydrogen molecule by Heitler and London, and later by Hund and Mulliken.[41] We consider now two hydrogen s-states centered at $\mathbf{R}_1$ and $\mathbf{R}_2$. The solutions to the overlap matrix problem [Eq. (7.6)] are

$$f_\pm(1) = \frac{1}{\sqrt{2}} = \pm f_\pm(2), \tag{7.40}$$

where $m_\pm = 1 \pm \omega$, ω being the positive real matrix $\langle 1|2\rangle$. The corresponding "Bloch" functions are

$$|\pm\rangle = (|1\rangle \pm |2\rangle)1/\sqrt{2(1+\omega)} \tag{7.41}$$

and the "Wannier" functions are

$$|\tilde{1}\rangle = |+\rangle\frac{1}{\sqrt{2}} + |-\rangle\frac{1}{\sqrt{2}}$$
$$= \frac{1}{2}\left(\frac{1}{\sqrt{1+\omega}}+\frac{1}{\sqrt{1-\omega}}\right)|1\rangle$$
$$+ \frac{1}{2}\left(\frac{1}{\sqrt{1+\omega}}-\frac{1}{\sqrt{1-\omega}}\right)|2\rangle$$

and

$$|\tilde{2}\rangle = |+\rangle\frac{1}{\sqrt{2}} - |-\rangle\frac{1}{\sqrt{2}}$$
$$= \frac{1}{2}\left(\frac{1}{\sqrt{1+\omega}}-\frac{1}{\sqrt{1-\omega}}\right)|1\rangle$$
$$+ \frac{1}{2}\left(\frac{1}{\sqrt{1+\omega}}+\frac{1}{\sqrt{1-\omega}}\right)|2\rangle. \tag{7.42}$$

The former (Bloch) are called molecular orbitals and the latter (Wannier), atomic orbitals.

The relations inverse to Eqs. (7.42) and (7.41) are

$$|\pm\rangle = (|\tilde{1}\rangle \pm |\tilde{2}\rangle)(1/\sqrt{2}),$$
$$|1\rangle = |+\rangle\sqrt{\tfrac{1}{2}(1+\omega)} + |-\rangle\sqrt{\tfrac{1}{2}(1-\omega)}$$
$$= |\tilde{1}\rangle\tfrac{1}{2}(\sqrt{1+\omega}+\sqrt{1-\omega}) + |\tilde{2}\rangle\tfrac{1}{2}(\sqrt{1+\omega}-\sqrt{1-\omega}),$$

and

$$|2\rangle = |+\rangle\sqrt{\tfrac{1}{2}(1+\omega)} - |-\rangle\sqrt{\tfrac{1}{2}(1-\omega)}$$
$$= |\tilde{1}\rangle\tfrac{1}{2}(\sqrt{1+\omega} - \sqrt{1-\omega}) + |\tilde{2}\rangle\tfrac{1}{2}(\sqrt{1+\omega} + \sqrt{1-\omega}). \quad (7.43)$$

Let $|\alpha\beta\rangle$, $\alpha\beta = 1,2$, be the vector for which the first electron is in state $|\alpha\rangle$ and the second in state $|\beta\rangle$. Then the spatial part of the two electron singlet states is spanned by the following states:

$$|++\rangle = [1/2(1+\omega)]](|11\rangle + |12\rangle + |21\rangle + |22\rangle)$$
$$= \tfrac{1}{2}(|\widetilde{11}\rangle + |\widetilde{12}\rangle + |\widetilde{21}\rangle + |\widetilde{22}\rangle),$$
$$(|+-\rangle + |-+\rangle)(1/\sqrt{2})$$
$$= [1/\sqrt{2(1-\omega^2)}](|11\rangle - (|22\rangle)$$
$$= (1/\sqrt{2})(|\widetilde{11}\rangle - |\widetilde{22}\rangle),$$
$$|--\rangle = [1/2(1-\omega)]](|11\rangle - |12\rangle - |21\rangle + |22\rangle)$$
$$= \tfrac{1}{2}(|\widetilde{11}\rangle - |\widetilde{12}\rangle - |\widetilde{21}\rangle + |\widetilde{22}\rangle). \quad (7.44)$$

The first and third states in Eq. (7.44) are ${}^1\Sigma_g$ states, and the second state is a ${}^1\Sigma_u$ state. The original Heitler–London state is

$$\frac{|12\rangle + |21\rangle}{\sqrt{2(1+\omega^2)}}$$
$$= \{|++\rangle(1+\omega) - |--\rangle(1-\omega)\}\frac{1}{\sqrt{2(1+\omega^2)}}, \quad (7.45)$$

whereas the original Hund–Mulliken state is $|++\rangle$ of Eq. (7.44). Clearly each is an arbitrary linear combination of the two ${}^1\Sigma_g$ functions and thus would yield a higher value of the energy than the lower of the two energies obtained by diagonalizing the two electron–two fixed proton Hamiltonian, the diagonalization being done with linear combinations of $|++\rangle$ and $|--\rangle$.

We conclude this section by analyzing a somewhat nonphysical problem, that of a spatially uniformly continuous distribution of bound states. Consider the case for which the site index i is replaced by a continuous variable $\mathbf{R}$, which is uniformly distributed over all space. We take a normalized bound state wave function centered at $\mathbf{R}$ to be

$$\psi(\mathbf{r}-\mathbf{R}) \equiv \langle \mathbf{r}|\mathbf{R}\rangle \equiv \frac{1}{(2\pi)^{3/2}}\int \phi(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})}\, d^3\mathbf{k}. \quad (7.46)$$

Then the $\mathbf{r}$, $\mathbf{r}'$ matrix elements of $\hat{g}$ are given by

$$\langle \mathbf{r}|\hat{g}|\mathbf{r}'\rangle \equiv \int \phi(\mathbf{r}-\mathbf{R})\, d^3\mathbf{R}\, \bar{\phi}(\mathbf{r}'-\mathbf{R})$$
$$= \int |\phi(\mathbf{k})|^2 e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\, d^3\mathbf{k}, \quad (7.47)$$

and $\hat{g}$ to any real power α has matrix elements

$$\langle \mathbf{r}|\hat{g}^\alpha|\mathbf{r}'\rangle = \frac{1}{(2\pi)^3}\int [(2\pi)^3|\phi(\mathbf{k})|^2]^\alpha e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\, d^3\mathbf{k}, \quad (7.48)$$

provided the integral exists. We see that

$$\langle \mathbf{r}|(\hat{g}+\epsilon)^{-1/2}|\mathbf{r}'\rangle = \frac{1}{(2\pi)^{9/2}}\int \frac{e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\, d^3\mathbf{k}}{(|\phi(\mathbf{k})|^2+\epsilon)^{1/2}}. \quad (7.49)$$

Now,

$$\langle \mathbf{r}'|\mathbf{R}\rangle = \frac{1}{(2\pi)^{3/2}}\int e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{R})}\phi(\mathbf{k}')\, d^3\mathbf{k}',$$

and hence,

$$\langle \mathbf{r}|(\hat{g}+\epsilon)^{-1/2}|\mathbf{R}\rangle = \langle \mathbf{r}|\tilde{\mathbf{R}}\rangle = \tilde{\psi}_{\mathbf{R}}(\mathbf{r})$$
$$= \frac{1}{(2\pi)^6}\int\int\int e^{i\mathbf{k}\cdot\mathbf{r}} e^{-i\mathbf{k}\cdot\mathbf{r}'}\, d^3\mathbf{k} \times \frac{e^{i\mathbf{k}'\cdot(\mathbf{r}'-\mathbf{R})}}{(|\phi(\mathbf{k})|^2+\epsilon)^{1/2}}\, d^3\mathbf{k}'\, \phi(\mathbf{k}')\, d^3\mathbf{r}'$$
$$= \frac{1}{(2\pi)^3}\int \frac{\phi(\mathbf{k})}{(|\phi(\mathbf{k})|^2+\epsilon)^{1/2}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})}\, d^3\mathbf{k}.$$

Therefore,

$$\tilde{\psi}_{\mathbf{R}}(\mathbf{r}) = \lim_{\epsilon\downarrow 0}\langle \mathbf{r}|(\hat{g}+\epsilon)^{-1/2}|\mathbf{R}\rangle$$
$$= \frac{1}{(2\pi)^3}\int \frac{\phi(\mathbf{k})}{|\phi(\mathbf{k})|} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})}\, d^3\mathbf{k}. \quad (7.50)$$

We observe that $\tilde{\psi}_{\mathbf{R}}(\mathbf{r})$ is not in $L^2(R^3, d^3\mathbf{x})$ as could have been anticipated since in the expression

$$\int |\mathbf{R}\rangle d^3\mathbf{R}\langle \mathbf{R}| = \hat{g}, \quad (7.51)$$

the total measure $\int d^3\mathbf{R} = \infty$.

Even though we are dealing with improper functions in the L^2 sense, nevertheless,

$$\int \tilde{\psi}_{\mathbf{R}}(\mathbf{r})\tilde{\psi}_{\mathbf{R}}(\mathbf{r}')d^3\mathbf{R}$$
$$= \frac{1}{(2\pi)^6}\int\int \frac{\phi(\mathbf{k})}{|\phi(\mathbf{k})|} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} \times e^{-i\mathbf{k}'\cdot(\mathbf{r}'-\mathbf{R})}\frac{\bar{\phi}(\mathbf{k})}{|\phi(\mathbf{k})|}\, d^3\mathbf{R}\, d^3\mathbf{k}\, d^3\mathbf{k}'$$
$$= \frac{1}{(2\pi)^3}\int\int \frac{\phi(\mathbf{k})}{|\phi(\mathbf{k})|} e^{i\mathbf{k}\cdot\mathbf{r}}\delta(\mathbf{k}'-\mathbf{k})\frac{\bar{\phi}(\mathbf{k}')}{|\phi(\mathbf{k}')|}\, d^3\mathbf{k}'\, d^3\mathbf{k}$$
$$= \delta^{(3)}(\mathbf{r}-\mathbf{r}'), \quad (7.52)$$

which represents the unit projector on H. This result could have been anticipated, since the functions given by Eq. (7.46) are total for any reasonable choice of $\phi(\mathbf{k})$, i.e., a Gaussian.[42]

VIII. CONCLUSIONS

We have presented in an elementary manner some of the basic principles having to do with projection operators. In addition, we have shown how to apply these basic techniques to a variety of problems occuring in theoretical physics. We make a case for the unity of the subject and try to show the great utility of this unified approach.

[1] F. Jørgensen, Int. J. Quant. Chem. **12**, 397–426 (1977).

[2] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966), p. 26.

[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1980), p. 187.

[4] C. Misner, K. Thorn, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 66.

[5] P. A. M. Dirac, *Quantum Mechanics* (Oxford U. P., New York, 1947), 3rd ed.

[6]T. Kato, Ref. 2, pp. 32–4.
[7]N. I. Akhiezer, and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1961), Vol. I, p. 69.
[8]J. Dixmier, *Les algebres d'operateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1969), 2nd ed., p. 323.
[9]W. E. Brittin, *Mathematical Methods in Theoretical Physics—Lectures in Theoretical Physics, Vol. XIVB* (Colorado Associated U. P., Boulder, 1973), pp. 55–96.
[10]J. von Neumann, *Functional Operators, Vol. II: The Geometry of Orthogonal Spaces* (Princeton U. P., Princeton, NJ, 1950), pp. 51–59.
[11]R. N. Hill, J. Math. Phys. **18**, 2316 (1977).
[12]W. E. Brittin and A. Y. Sakakura, J. Math. Phys. **21**, 2063–2069 (1980).
[13]W. E. Brittin, A. Y. Sakakura, and W. Wyss (unpublished).
[14]J. des Cloizeaux, Nucl. Phys. **20**, 321 (1960).
[15]T. Kato, Ref. 2, p. 282.
[16]E. Montroll, private communication.
[17]P. O. Löwdin, Adv. Phys. **5**, 1 (1956).
[18]H. Feshbach, Ann. Phys. (N. Y.) **19**, 287 (1962).
[19]A. K. Kerman, *Lectures in Theoretical Physics, Vol. VIIIC*, edited by W. E. Brittin (The University of Colorado Press, Boulder, 1966), p. 565.
[20]W. H. Furry, *Lectures in Theoretical Physics, Vol. VIIIA*, edited by W. E. Brittin (The University of Colorado Press, Boulder, 1966), p. 1.
[21]O. Bratteli, and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics I* (Springer-Verlag, New York, 1979), Chap. 2.
[22]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1980), p. 198.
[23]M. L. Goldberger and K. M. Watson, *Phys. Rev.* **134**, B919 (1964).
[24]F. Herbut, Ann. Phys. (N. Y.) **55**, 271 (1969).
[25]S. Gudder, J. Marchand, and W. Wyss, J. Math. Phys. **20**, 1963 (1979).
[26]J. Bub, *The Interpretation of Quantum Mechanics* (Reidel, Boston, 1974).
[27]C. Y. She and H. Heffner, Phys. Rev. **152**, 1103 (1966).
[28]S. T. Ali and G. G. Emch, J. Math. Phys. **15**, 176 (1974).
[29]S. T. Ali and E. Prugovecki, J. Math. Phys. **18**, 219 (1977).
[30]A. S. Wightman, Rev. Mod. Phys. **34**, 845 (1962).
[31]R. W. Zwanzig, *Lectures in Theoretical Physics, Vol. III*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience, New York, 1961), pp. 106–41.
[32]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977); M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1979), Vol. III; W. Thirring, *A Course in Mathematical Physics* (Springer-Verlag, New York, 1981), Vol. III.
[33]V. Fock, Z. Phys. **75**, 622–47 (1932) [C. R. Smith, NBS Translation T8-60].
[34]R. P. Feynmann, *Statistical Mechanics* (Benjamin, Reading, MA, 1972), p. 170.
[35]P. O. Löwdin, J. Chem. Phys. **18**, 365 (1958).
[36]B. Brandow, Rev. Mod. Phys. **39**, 771 (1967).
[37]P. O. Löwdin, Adv. Quantum Chem. **5**, 185 (1970).
[38]G. H. Wannier, Phys. Rev. **52**, 191 (1937).
[39]G. H. Wannier, J. Math. Phys. **19**, 131 (1977).
[40]A. Erdelyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, pp. 355, 358, 370.
[41]J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1963), Vol. 1, pp. 41–59, 62–71.
[42]V. Bargmann, P. Butera, and L. Girardello, Rep. Math. Phys. **2**, 221 (1971).

Fractals and nonstandard analysis

L. Nottale and J. Schneider
Observatoire de Meudon, 92195 Meudon Principal Cedex, France

(Received 2 November 1982; accepted for publication 22 July 1983)

We describe and analyze a parametrization of fractal "curves" (i.e., fractal of topological dimension 1). The nondifferentiability of fractals and their infinite length forbid a complete description based on usual real numbers. We show that using nonstandard analysis it is possible to solve this problem: A class of nonstandard curves (whose standard part is the usual fractal) is defined so that a curvilinear coordinate along the fractal can be built, this being the first step towards the possible definition and study of a fractal space. We mention fields of physics to which such a formalism could be applied in the future.

PACS numbers: 02.40. + m

I. INTRODUCTION

The concept of fractals, introduced by B. Mandelbrot,[1–3] applies to any curve, object, or set "whose form is extremely irregular and/or fragmented at all scales." More precisely, let D be the fractal dimension (e.g., the Hausdorff–Besicovitch dimension); a fractal is defined[2,3] as a set for which D is greater than the topological dimension D_T.

Nowadays this concept is increasingly considered in physics for several reasons:

(i) Fractal curves are functions which are continuous but nowhere differentiable; this property has already been observed for some natural phenomena, such as particle trajectories in quantum mechanics.[4,5]

(ii) The length of a fractal curve is dependent on the resolution with which it is measured and diverges when the resolution tends to be infinite.

(iii) A fractal dimension can be any real number, so this concept may apply to fields of physics such as critical behavior phenomena, where noninteger dimension has become a necessity.

Mandelbrot[1–3] pointed out many examples of the contribution fractals can bring to the description of natural phenomena such as the length of a coastline, the distribution of matter in the universe, turbulence, moon craters... . Furthermore, the concept of Haussdorff dimension has been applied to QCD jets,[6] gauge theories,[7] critical behavior,[8] fluctuations of the early universe,[9] or quantum-mechanical paths.[5]

However, in most cases, the authors limit themselves to the calculation of a fractal dimension or use fractals in a purely descriptive way (but see Le Mehauté *et al.*[10,11]). Because of the wide domain where phenomena seem to exhibit a fractal behavior, one is entitled to wish that a more thorough use of this concept would be possible, e.g., by building a formalism based on fractals and suitable for physics. In fact, no explicit calculation is presently possible on the fractal *itself* (i.e., the limit object instead of one of its approximations). It is the aim of this paper to show that nonstandard analysis, as built up by Robinson[12] is well adapted to such calculations.

In this paper, we first parametrize fractal "curves" (i.e., fractals of topological dimension 1) in the Cesaro[13,2] way (Sec. II). Then some paradoxical properties of fractals are evidenced and are clarified by the use of nonstandard analysis (Sec. III) as a way to build intrinsic curvilinear coordinates along a fractal curve (Sec. IV). This is hopefully a first step towards the definition of a fractal space by its own, while so far fractals have been considered as subsets of an integer-dimensional space.

II. PARAMETRIC EQUATION OF A FRACTAL CURVE

Consider a generalized von Koch curve in the $\mathbb{R}^2$ (or $\mathbb{C}$) plane. It can be built from an initial curve F_1 made up of p segments of equal length $1/q$ which connect the origin to the point [0,1] (see Fig. 1). Let ω_{j+1} be the polar angle of the jth segment and $Z_j = X_j + iY_j = (1/q)\sum_{k=0}^{j-1} e^{i\omega_k}$ the complex coordinate of a breaking point P_j. Two conditions hold between these data:

$$\sum_{k=0}^{p-1} e^{i\omega_k} = q, \quad Z_{j+1} - Z_j = \frac{1}{q} e^{i\omega_j}. \tag{2.1}$$

A curve F_2 is obtained by substituting each segment of F_1 by F_2 itself, scaled at its length $1/q$, as illustrated in Fig. 2. The resulting curve of an infinite sequence of these steps (substitution of each segment of F_n by $q^{-n} F_1$ giving F_{n+1}) is the fractal F.

As indicated by Mandelbrot[3] in the case of the Peano curve, F can be parametrized by a real number $x \in [0,1]$ developed in the counting base p in the form (see Fig. 2):

$$x = 0.x_1x_2\cdots = \sum_{1}^{\infty} x_k\, p^{-k}. \tag{2.2}$$

The fractal will apparently be completely defined when the complex coordinate $Z(x)$ of the point on F parametrized by x is known. $Z(x)$ can be easily obtained, thanks to the above building process of F, under the form

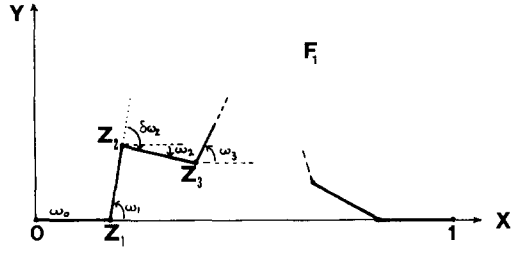


FIG. 1. Building of the basic structure of a fractal curve.

 0022-2488/84/051296-05$02.50

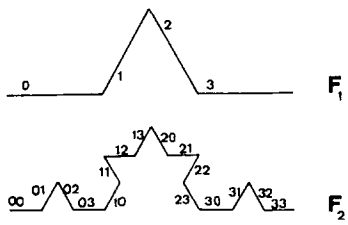


FIG. 2. Parametrization of a fractal curve.

$$Z(x) = Z_{x_1} + q^{-1}e^{i\omega_{x_1}}\left[Z_{x_2} + q^{-1}e^{i\omega_{x_2}}[Z_{x_3} + \cdots]\right], \tag{2.3}$$

so that we finally obtain

$$Z(x) = q\sum_{k=1}^{\infty} Z_{x_k} e^{i\sum_{l=1}^{k-1}\omega_{x_l}} q^{-k}. \tag{2.4}$$

If one defines new variables ρ_j, θ_j, and φ_{x_j} by writing

$$Z_j = q\rho_j e^{i\theta_j}, \quad \varphi_{x_j} = \theta_{x_j} + \sum_{l=1}^{j-1}\omega_{x_l}, \tag{2.5}$$

another interesting form for the parametric equations of a fractal can be given

$$X(x) = \sum_{k=1}^{\infty}\rho_{x_k}\cos(\varphi_{x_k})q^{-k},$$
$$Y(x) = \sum_{k=1}^{\infty}\rho_{x_k}\sin(\varphi_{x_k})q^{-k}. \tag{2.6}$$

It is well known that, with its dimension lying between 1 and 2, a fractal curve within a plane is intermediate between a line and a surface. Indeed, while it may be built by adding segments it may also be obtained by deleting surfaces.[2,3] This construction allows dealing with the problem of multiple points. Given an initial curve F_1, consider a polygon P_0 of surface $\mathcal{S}_0$ with one of its diagonals being the segment [0,1] and in which F_1 is included. Then build around each segment of F_1 a q-reduced scale version of P_0: We obtain a figure P_1 (for example of such a construction, see Fig. 3). An obviously sufficient condition for the absence of multiple points is that all polygons of P_1 are disjointed, since this will remain true in the figures $P_2, P_3, \ldots, P_n \ldots$. Let $\delta\omega_j = \omega_j - \omega_{j-1}$ be the angle between two segments in F_1 (see Fig. 1). Then for two adjacent segments, the above condition leads to (angles α, β, γ are defined in Fig. 3)

$$\alpha^+ + \beta^+ < \pi - \delta\omega_j, \quad \alpha^- + \beta^- < \pi + \delta\omega_j, \tag{2.7}$$

so that we get for all values of j the sufficient conditions

$$-\gamma^- < \delta\omega_j < \gamma^+. \tag{2.8}$$

Let us now illustrate the nonstandard character of fractals, answering (by the negative) the question: Do Eqs. (2.2) and (2.6) characterize all properties of F in the $\mathbb{R}^2$ plane?

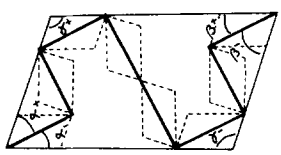


FIG. 3. Building of a fractal by deleting surfaces.

We first recall that, with its fractal dimension[2,3] $D = \log p/\log q$, the length of the fractal is infinite and its surface vanishes since the F_n length and surface are

$$\mathcal{L}_n/\mathcal{L}_0 = (p/q)^n = q^{n(D-1)},$$
$$\mathcal{L}_n/\mathcal{L}_0 = (p/q^2)^n = q^{-n(2-D)}. \tag{2.9}$$

Moreover, however small the difference of parameters $x(2) - x(1)$ for two points M_2 and M_1 on the fractal, though the distance in the plane $|Z(x_2) - Z(x_1)|$ vanishes, the distance along the fractal remains infinite.

However, let us build the following sequence:

$$a_n = a\mathcal{L}_n^{-1} = (aq^n)p^{-h} = 0.a_{n1}a_{n2}\cdots a_{np}, \tag{2.10}$$

where a is a nonzero number. Assume now that the a_n's are the parameters of a given sequence of points M_n on the curves F_n. The curvilinear coordinates of these points along F_n are equal to

$$l_n = \mathcal{L}_n a_n = a = \text{const.} \tag{2.11}$$

For example, in Fig. 4 the sequence of parameters 0.1, 0.03, 0.021, 0.0123, ... defines points at a constant distance 0.1 from the origin on the respective curves $F_1, F_2, F_3, F_4 \ldots$. Let now $n \to \infty$ and consider the limit point M of the sequence M_n on F. From Eq. (2.10) its parameter is $\lim(a_n) = 0$, i.e., *M coincides with the origin* 0; however, from Eq. (2.11), the curvilinear distance between O and M on the fractal is $a \neq 0$.

From this "paradox" we conclude that the real coordinate x is insufficient to describe thoroughly the fractal curve F: the distance along F between two points parametrized by two different x's is infinite, while points separated by a finite distance along F correspond to the same values of x. Thus another formalism is needed, and more precisely a set "larger" than $\mathbb{R}$: Nonstandard analysis[12] which allows dealing both with infinitesimals and infinite numbers, provides such a frame, well adapted to the study of fractals, as we show in the following sections.

III. NONSTANDARD ANALYSIS: A REMINDER

Nonstandard analysis (NSA) may be considered as the solution A. Robinson worked out for the old problem of infinitesimals. Leibnitz, founder of differential calculus, thought of "infinitely small" and "infinitely large" numbers as ideal numbers to which operations on usual numbers

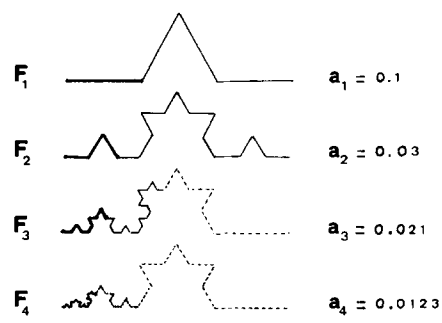


FIG. 4. A sequence of points on F_n with constant curvilinear distance to the origin.

would apply, though he was unable to build a coherent system: In fact the up to now accepted signification of these expressions involves the theory of limits and the so-called epsilon-delta method according to the works of Cauchy and Weierstrass. However, Robinson[12,14] has demonstrated that real numbers $\mathbb{R}$ can be extended to $^*\mathbb{R}$ which contains infinitely small and infinitely large numbers.

We will not try here to present a detailed description of NSA, but only recall some basic results useful in what follows independent of the precise way the theory is evolved, using, e.g., free ultrafilters and equivalence classes of sequences; see Robinson,[14] Stroyan and Luxemburg,[15] or using an axiomatic extension of the Zermelo set theory, see Nelson.[16]

The set $\mathbb{R}$ of hyper-real numbers is a totally ordered and non-Archimedean field. The set $\mathbb{R}$ of standard numbers is a subset of $^*\mathbb{R}$. $^*\mathbb{R}$ contains infinite elements, i.e, elements A such that $\forall n \in \mathbb{N}$, $|A| > n$. It also contains infinitesimal elements, i.e., B such that $\forall\, n \in \mathbb{N}\ (n \neq 0)$, $|B| < 1/n$. A finite element C is defined: $\exists\, m \in \mathbb{N}$, $|C| < m$. The set of infinitesimals is denoted by $\mathfrak{o}$, the set of finite numbers by $\mathfrak{O}$ and the set of infinite numbers by $^*\mathbb{R}_\infty$. Any finite number $a \in \mathfrak{O}$ can be split up in a single way as $a = r + \epsilon$, where $r \in \mathbb{R}$ and $\epsilon \in \mathfrak{o}$. In other words, the finite hyper–reals contain the ordinary reals with new numbers a clustered infinitesimally closely around each ordinary real r. Their set $\{a\}$ is called the monad of r. The real r is said to be the "standard part" of the hyper-real a, a function denoted by $r = \mathrm{st}(a)$. The "st" function is very useful for nonstandard demonstrations of standard theorems. For instance a sum $\Sigma_0^\infty f_n$ is said to converge if for different λ 's belonging to the set of infinite hypernatural numbers $^*\mathbb{N}_\infty$, st $(\Sigma_0^\lambda f_n)$ are all equal to the same finite number. Apart from the strict equality " = ", one introduces the equivalence relations "$\simeq$" meaning "infinitely close to," i.e., $a \simeq b$ $\langle\Leftrightarrow\rangle \mathrm{st}(a-b) = 0$, and the relation "$\sim$" meaning "of the order of," i.e., $a \sim b$ $\langle\Leftrightarrow\rangle$ $\exists\, k \in \mathbb{R}$, $k \neq 0$ such that $a = kb$.

Formal descriptions of NSA may be found in Refs. 14–17. There have also been some attempts of applications of NSA to physics, e.g., Kelemen and Robinson,[18,19] Moore,[20] Anderson.[21]

IV. NONSTANDARD COORDINATES ALONG FRACTALS

In Sec. II, a fractal in $\mathbb{R}^2$ was parametrized by a real number belonging to the interval [0,1]:

$$x = \frac{x_1}{p} + \frac{x_2}{p^2} + \cdots + \frac{x_n}{p^n} + \cdots . \tag{4.1}$$

Let us generalize the usual fractal by introducing a curve F_ω in $^*\mathbb{R}^2$, parametrized by an hyper-real number $x^* \in {}^*[0,1]$, as defined by the *-finite power series expansion

$$x^* = \frac{x_1}{p} + \cdots + \frac{x_n}{p^n} + \cdots + \frac{x_{\omega-m}}{p^{\omega-m}} + \cdots + \frac{x_\omega}{p^\omega}, \tag{4.2}$$

where $\omega \in {}^*\mathbb{N}_\infty$. In other terms, F_ω may be obtained by applying the building process in Sec. II (i.e., build F_{n+1} by substituting to each segment of F_n, F_1 scaled by $q^{-n})$ ω times. It should be noticed that F_ω is *not* a fractal in the nonstandard sense (since the fragmentation is * limited up to ω) but its standard part is identical to the usual fractal, i.e.,

$$F \equiv \mathrm{st}(F_\omega). \tag{4.3}$$

Then the study of F_ω allows us to study the properties of F, thanks to the standardization axiom.

A first advantage is that, while the length of F was undefined, the length of F_ω is defined:

$$\mathscr{L}_\omega = (p/q)^\omega = q^{\omega(D-1)}. \tag{4.4}$$

While the surface of F was zero, the surface of F_ω is an infinitesimal:

$$\mathscr{L}_\omega = (p/q^2)^\omega = q^{-\omega(2-D)}. \tag{4.5}$$

The curvilinear coordinate ξ of the point parametrized by x^* is now also defined on F_ω:

$$\xi = x^* \mathscr{L}_\omega = q^{-\omega}\{x_\omega + x_{\omega-1}\, p + \cdots + x_1 p^{\omega-1}\}. \tag{4.6}$$

This verifies that F_ω is built up with elementary segments of length $q^{-\omega}$. By using an infinitely great magnifying power, F_ω can be drawn exactly (while this was not the case for F) as in Fig. 5. The fractal is no more a limit concept.

Let us now utilize the new concept of F_ω to study or clarify some problems specific of fractals, which may be relevant for physical applications.

A. Finite distance along the fractal

In Sec. II we obtained a point separated from the origin by a finite nonzero distance along the fractal, while its parameter was $x = 0$. This situation may be clarified by defining M such that

$$\xi_M = \mathscr{L}_\omega\, p^{-M} = 1. \tag{4.7}$$

Then M is a solution of the equation

$$p^{\omega-M} = q^\omega, \tag{4.8}$$

so that

$$M = \omega(1 - \log p/\log q) = \omega(1 - 1/D). \tag{4.9}$$

Generally M is not an integer; therefore we define $\lambda \in {}^*\mathbb{N}_\infty$ as

$$\lambda = \mathrm{Int}(M) = \mathrm{Int}[\omega(1 - 1/D)], \tag{4.10}$$

where $\mathrm{Int}(X)$ is the integer part of X (it is straightforward to

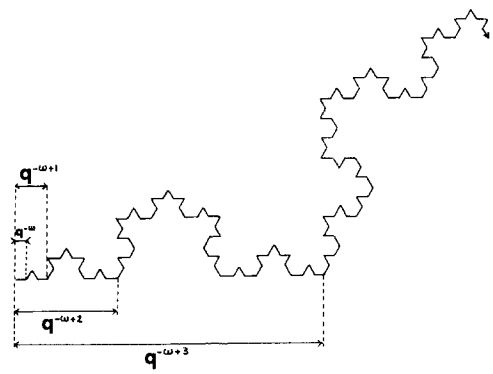


FIG. 5. Infinite magnification of curve F_ω, the standard part of which is the fractal F.

verify that this function is still defined in *$\mathbb{R}$). For any λ' such that $\lambda' \sim \lambda$ (the relation $\sim$ "of the order of" has been defined in Sec. III), the curvilinear distance along the fractal $\xi_{\lambda'} = \mathscr{L}_\omega p^{-\lambda'}$ belongs to $\mathbb{R}$. However, and this allows one to understand why $x = 0$ while $\xi \neq 0$, the corresponding distance in the $\mathbb{R}^2$ plane is an infinitesimal, $q^{-\lambda'}$, the standard part of which is thus zero.

Consider two points on F_ω separated by a curvilinear distance ξ. Depending on the power n of p in the expression (4.6) of ξ, three levels may be distinguished on F_ω:

—$n \in \mathbb{N}$: Finite distance in the $\mathbb{R}^2$ plane, infinite along the fractal.

—$n \sim \lambda$: Infinitesimal distance in the $\mathbb{R}^2$ plane, finite distance along the fractal.

—$n \sim \omega$: Infinitesimal distance in $\mathbb{R}^2$ and on F_ω. At this level the two distances are of the same order.

B. Intrinsic building of F

Parametric equations for F have been given in Sec. II (Eq. 2.6): It relates the parameter x to the coordinates X, Y in the $\mathbb{R}^2$ plane. An intrinsic building of F_ω, independent of the plane in which it is embedded, is possible: We only need to know the change in direction from each elementary segment of length $q^{-\omega}$ to the following one. In Sec. II, we had set $\delta\omega_j = \omega_j - \omega_{j-1}$, the angle between two segments in F_1 (see Fig. 1). Consider the infinitesimal segment of curvilinear coordinate ξ on F_ω, and define h such that $x_{\omega-h}$ is the first nonzero figure of the hypernatural number $q^\omega \xi$ in the base p:

$$\xi = q^{-\omega} \times \{0 + 0\times p + \cdots + 0\times p^{h-1} + x_{\omega-h} p^h + \cdots + x_1 p^{\omega-1}\}. \tag{4.11}$$

The relation for the angle we were looking for and which allows an intrinsic building of F_ω is simply given by

$$\delta\omega_j(\xi) = \delta\omega_{x_{\omega-h}}. \tag{4.12}$$

A somewhat paradoxical property of fractals is that this relative angle exists and may be computed for any value of ξ, while on the contrary the absolute angle $\omega_j(\xi)$ does not exist as soon as ξq^ω becomes infinite, since

$$\omega(\xi) = \sum_{q^\omega \xi' = 1}^{q^\omega \xi} \delta\omega_j(\xi') \tag{4.13}$$

so that the value of st $[\omega(\xi)]$ depends on that of ω (this is the nonstandard transfer of the nondifferentiability of F).

C. Family of curves F_ω and ϵ differentiability

The standard part of any curve differing from F_ω only by infinitesimals will also be the fractal F. This is true for any $F_{\omega'}$ with $\omega' \in {}^*\mathbb{N}_\infty$, $\omega' \neq \omega$, but might be generalized. In particular the "broken" aspect of F_ω can be given up and F_ω replaced by a smoothed nonstandard curve F'_ω, with $F = \text{st}(F'_\omega)$. A kind of differentiability can be defined for F'_ω.

Indeed, nonstandard analysis may be used to define differentiability of standard functions[15]:

$$\forall\ x \approx y \approx a, \frac{f(x)-f(a)}{x-a} \approx \frac{f(y)-f(a)}{y-a} \text{ finite},$$

then $f'(a) = \text{st}\left(\frac{f(x)-f(a)}{x-a}\right)$.

This definition could be generalized to an "ϵ differentiability" of nonstandard curves in the following way: If $\exists\ \epsilon \in \mathfrak{d}$, such that $\forall\ x, y, a \in {}^*\mathbb{R}$ verifying $|x-a| < \epsilon$ and $|y-a| < \epsilon$, and

$$\frac{f(x)-f(a)}{x-a} \approx \frac{f(y)-f(a)}{y-a}$$

finite, then

$$f'_\epsilon(a) = \text{st}\left(\frac{f(x)-f(a)}{x-a}\right).$$

In that sense, one can define a curve F'_ω which is ϵ differentiable everywhere, since the above definition is verified by taking $\epsilon = \epsilon_0$ with $\epsilon_0 q^\omega \approx 0$. It is clear that differentiability implies ϵ differentiability but that the reverse proposition is false, as shown by fractals.

V. GENERALIZATIONS AND CONCLUSION

To simplify the description of the concepts introduced in this paper, we have worked only with regular fractals of topological dimension 1 in $\mathbb{R}^2$. Some straightforward generalizations may be considered.

For instance a fractal curve in $\mathbb{R}^3$ is parametrized in the same way as in $\mathbb{R}^2$, the angles ω_j of the elementary segments of F_1 being replaced by rotation matrices R_j that define the new reference system tied to each segment $\mathbf{v}_j$; the elementary points are defined by vectors u_j, with $\mathbf{v}_j = \mathbf{u}_{j+1} - \mathbf{u}_j$. Thus, if p and q have the same definitions as above, a point of parameter $x = 0.x_1x_2\cdots x_n \cdots$ in the counting base p is referenced by a vector:

$$\mathbf{u}(x) = \sum_{n=0}^{\infty} R_{x_n} R_{x_{n-1}} \cdots R_{x_1} \mathbf{u}_{x_{n+1}} q^{-n}. \tag{5.1}$$

Such a curve has a topological dimension 1 and a fractal dimension $D = \log p/\log q$ lying between 1 and 3.

Another possible generalization is to give up self-similarity and to build a more general fractal curve in $\mathbb{R}^2$. F_1 is built up with p_0 segments, the coordinates of the "breaking points" being Z_{x_1}, with $x_1 = 0, 1, \ldots, p-1$. In each segment of F_1 referenced by x_1, we introduce p_{x_1} new segments defined by $Z_{x_1x_2}$, $x_2 = 0, 1, \ldots, p_{x_1} - 1$, and angle $\omega_{x_1x_2}$; we then build F_2. This construction is extended to the infinite, by giving ourselves the coordinates $Z_{x_1\cdots x_n}$, where $x_n \in \{0,1,\ldots, p_{x_1\cdots x_{n-1}} - 1\}$ and angles $\omega_{x_1\cdots x_n}$. There are $p_{x_1\cdots x_{n-1}}$ new segments of length $1/q_{x_1\cdots x_n}$. Since the base p varies from segment to segment, the sequence $x_1x_2\cdots x_n \cdots$ is now only a code and cannot be summarized by a number. The complex coordinate in the $\mathbb{R}^2$ plane of a point on F is then

$$Z(x_1x_2\cdots x_n \cdots) = \sum_{n=0}^{\infty} \frac{z_{x_1\cdots x_{n+1}} e^{i(\omega_{x_1} + \cdots + \omega_{x_1\cdots x_n})}}{q_{x_1} q_{x_1x_2} \cdots q_{x_1\cdots x_n}}. \tag{5.2}$$

This construction is obviously not unique. The dimension of F could be defined as $\langle p \rangle \langle q^{-D} \rangle = 1$ (see Mandelbrot[2]).

Up to now, only fractals with a globally given dimen-

sion D have been discussed. However the fractal dimension can clearly be a local property,[2] varying with the curvilinear coordinate $D = D(\xi)$; this is more easily understandable by remembering the underlying infinitesimal structure.

In the future, we will try to extend the parametrization presented here to fractals of topological dimensions greater than one. Instead of studying fractal *objects* embedded in an Euclidean space, the aim of such a work would be to define a *fractal space* intrinsically. Its dimension could then be a generic parameter and function of the coordinates in the same way as curvature occurs for curved spaces.

In a forthcoming paper,[22] one of the authors will consider physical applications of this formalism to quantum mechanics (uncertainty relations and theory of measure). We think the notions developed here could help in the field of quantization of gravity: the studies in this field usually assume without question an underlying, eventually foamy,[23] 4-manifold. Below some characteristic length, space itself could become a fractal. To be more specific one could for instance perform latticelike gauge field calculations[24] on a fractal. Let us notice that the path-integral ingredient of such calculations has also recently been formulated in terms of nonstandard analysis.[20] When applied to cosmology these ideas lead to the natural speculation that the very early universe experienced in its whole a noninteger-dimensional phase. To conclude we hope that these new trends will help to answer the question: Why does the macroscopic space appear to be three dimensional?

ACKNOWLEDGMENTS

We are very grateful to G. Lacroix for having brushed up the English of this paper. We acknowledge B. Mandelbrot for helpful discussion and comments.

[1] B. B. Mandelbrot, *Les objects fractals* (Flammarion, Paris, 1975).
[2] B. B. Mandelbrot, *Fractals* (Freeman, San Francisco, 1977).
[3] B. B. Mandelbrot, *The fractal geometry of nature* (Freeman, San Francisco, 1982).
[4] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[5] L. F. Abbott and M. B. Wise, Am. J. Phys. **49**(1), 37 (1981).
[6] K. Konishi, CERN preprint 1980.
[7] G. Parisi, Phys. Lett. B. **818**, 357 (1979).
[8] K. Keller, preprint, 1981.
[9] A. G. Doroshkevich, Ya. B. Zel'dovich, R. A. Syunyaev, and M. Yu. Khlopov, Pis'ma Astron. Zh. **6**, 457(1980) [Sov. Astron. Lett. **6**, 252 (1980)].
[10] A. Le Mehauté and G. Crepy, C. R. Acad. Sci. Paris **294**, 685 (1982).
[11] A. Le Mehauté, A. de Guibert, M. Delaye, and Ch. Filippi, C. R. Acad. Sci. Paris **294**, 835 (1982).
[12] A. Robinson, Proc. R. Acad. Sci. Amsterdam, Ser. A **64**, 432 (1961).
[13] E. Cesaro, 1905, Atti R. Accad. Sci. Fis. Mat. Napoli **XII**, 1 (1905).
[14] A. Robinson, *Non-standard Analysis* (North-Holland, Amsterdam, 1965).
[15] K. D. Stroyan and W. A. J. Luxemburg, *Introduction to the Theory of Infinitesimals* (Academic, New York, 1976).
[16] E. Nelson, Bull. Amer. Math. Soc. **83**, 1165 (1977).
[17] J. M. Henle and E. M. Kleinberg, *Infinitesimal Calculus* (MIT Press, Cambridge, MA, 1979).
[18] P. J. Kelemen and A. Robinson, J. Math. Phys. **13**, 1870 (1972).
[19] P. J. Kelemen and A. Robinson, J. Math. Phys. **13**, 1875 (1972).
[20] S. Moore, "N. S. Analysis applied to path integral," Universidad de Los Andes, preprint, 1981.
[21] R. M. Anderson, Israel J. Math. **25**, 15 (1976).
[22] L. Nottale (in preparation).
[23] S. Hawking, Nucl. Phys. B **144**, 349 (1978).
[24] J. H. Drouffe and C. Itzykson, Phys. Rep. **38**, 133 (1978).

Mean curvature and radiation field in SO(2) electrodynamics

D. Fauth and M. Sorg
Institut für Theoretische Physik der Universität Stuttgart, Pfaffenwaldring 57, D-7000 Stuttgart 80, West Germany

(Received 19 April 1983; accepted for publication 5 August 1983)

The trivial bundle of orthonormal frames over flat space-time is decomposed into two subbundles with structure groups SO(1,1) and SO(2), respectively. The curvature in the SO(2) bundle is identified with the electromagnetic field. It is shown that on certain conditions imposed upon the bundle decomposition the exterior derivative of the mean curvature 1-form in the SO(2) bundle is equal to the curvature 2-form in the SO(1,1) bundle. These conditions are (i) the Frobenius integrability of certain distributions generated by the splitting of the associated tangent bundle, and (ii) the vanishing of the mean curvature in the SO(1,1) bundle. For a single point charge the curvature in the SO(1,1) subbundle is identical to the radiation part of the Liénard–Wiechert field.

PACS numbers: 02.40. + m, 11.15.Kc, 03.50.De

I. INTRODUCTION

In this paper we further elaborate the idea of geometrizing the Liénard–Wiechert field of a single point charge along the lines discussed in a previous publication.[1] There the bundle of orthonormal frames over space-time was split into an SO(1,1) and an SO(2) subbundle. By explicit calculation using retarded coordinates the curvature 2-form in the SO(2) bundle could be identified with the dual of the Liénard–Wiechert field, whereas the curvature 2-form in the SO(1,1) bundle turned out to be the radiation part of the Liénard–Wiechert field. On the other hand, the radiative part could be obtained as well by exterior differentiation of the mean curvature 1-form in the SO(2) subbundle. The main result of this paper is to reveal the origin of these seemingly unrelated features of the radiation field. We give sufficient conditions to be imposed on the splitting of the tangent space of space-time, in order to get the above-mentioned relation between curvature 2-form in the SO(1,1) bundle and mean curvature 1-form in the SO(2) bundle.

We proceed as follows:

In Sec. II we briefly review the geometry of the bundle of orthonormal frames over a modified space-time M_4^-. The modification consists of removing the timelike world lines of charged particles from ordinary Minkowski space. M_4^- serves as base space for the principal bundle Λ_4^- of orthonormal frames which has as its structure group the Lorentz group SO(1,3). The associated tangent bundle τ_4^- is decomposed into the Whitney sum of two two-dimensional vector bundles with SO(1,1) and SO(2) as associated structure groups, respectively. That decomposition of τ_4^- is equivalent to reducing the Lorentz group SO(1,3) to the product group $SO(1,1)\times SO(2)$. We calculate the connection coefficients and curvature 2-forms in the subbundles and subsequently discuss their behavior under local $SO(1,1)\times SO(2)$ gauge transformations. We introduce linear combinations of the connection coefficients which have the property of constituting a representation of the factor group SO(2). These new potentials allow us to infer the existence of an integral 2-surface, independently of how the tangent space was split previously. Moreover, if expressed in terms of the new potentials, the curvature 2-form in the SO(2) bundle decomposes into two pieces, which in case of the Liénard–Wiechert field represent, respectively, the bound and the radiation parts thereof.

In Sec. II we take a closer look at distributions $\hat{\Delta}$ and $\tilde{\Delta}$ which are representation spaces for the reduced gauge groups SO(1,1) and SO(2), respectively. If we require both $\hat{\Delta}$ and $\tilde{\Delta}$ to be integrable, the Frobenius theorem yields conditions on the new potentials. Demanding integrability of a certain three-dimensional distribution $\bar{\Delta}$, in which the two-dimensional distribution $\tilde{\Delta}$ is embedded umbilically we obtain some further constraints on our potentials. These integrability conditions constitute one part of the assumptions needed for establishing the main result of the paper.

In Sec. IV we shortly review the notions of torsion 1-form and mean curvature 1-form and give the results for these quantities in both the subbundles under consideration.

In Sec. V we combine the integrability conditions, the property of umbilicality of $\tilde{\Delta}$ in $\bar{\Delta}$ and the requirement of vanishing mean curvature and thus prove the following assertion: on the conditions stated the mean curvature 1-form in the SO(2) bundle is a gauge transform of the connection 1-form in the SO(1,1) bundle. In other words, the curvature in the SO(1,1) bundle is the derivative of the mean curvature 1-form in the SO(2) bundle.

We finish the paper by discussing the physical implications of the geometric assumptions made.

II. THE SPLITTING OF THE SPACE-TIME TANGENT BUNDLE

Consider the trivial tangent bundle τ_4^- of the modified Minkowski space M_4^-. We denote the principal bundle of orthonormal frames associated to τ_4^- by Λ_4^-. Λ_4^- being trivial allows us to choose the usual canonical connection ω in Λ_4^-. The assignment of an orthonormal frame $\mathbf{e}(x) = \{\mathbf{e}_\mu(x)\}$ to each point $x \epsilon M_4^-$ defines a cross section in Λ_4^-. The local canonical connection $\omega(x)$ is obtained by taking the covariant derivative ∇ of the tetrad vectors $\mathbf{e}_\mu(x)$,

$$\nabla \mathbf{e}(x) = \mathbf{e}(x) \otimes \omega(x). \tag{2.1}$$

The connection 1-form $\omega(x)$ takes its values in the Lie algebra $\mathfrak{so}(1,3)$ of the Lorentz group SO(1,3) and can be decomposed

 0022-2488/84/051301-07$02.50

with respect to a basis of generators (T^i, t^i) of $\mathfrak{iso}(1,3)$:

$$\omega(x) = \mathbb{A}_{a,b}(x)T^a + \mathbb{B}_{a,b}(x)t^a. \tag{2.2}$$

The Lorentz group generators satisfy the commutation relations $(a,b,c = 1,2,3)$

$$\begin{aligned}
&[T^a, T^b] = \epsilon^{ab}{}_c T^c,\\
&[T^a, t^b] = \epsilon^{ab}{}_c t^c,\\
&[t^a, t^b] = -\epsilon^{ab}{}_c T^c.
\end{aligned} \tag{2.3}$$

If we make the identifications

$$\mathbf{e}_0 \equiv \mathbf{u}, \quad \mathbf{e}_1 \equiv \mathbf{v}, \quad \mathbf{e}_2 \equiv \mathbf{k}, \quad \mathbf{e}_3 = \mathbf{h}$$

with

$$(\mathbf{u}\cdot\mathbf{u}) = -(\mathbf{v}\cdot\mathbf{v}) = -(\mathbf{k}\cdot\mathbf{k}) = -(\mathbf{h}\cdot\mathbf{h}) = +1, \tag{2.4}$$

the Cartan structure equations (2.1) read in full detail

$$\begin{aligned}
\nabla\mathbf{u} &= \mathbf{v}\otimes\mathbb{B}_1 + \mathbf{k}\otimes\mathbb{B}_2 + \mathbf{h}\otimes\mathbb{B}_3,\\
\nabla\mathbf{v} &= \mathbf{u}\otimes\mathbb{B}_1 + \mathbf{k}\otimes\mathbb{A}_3 - \mathbf{h}\otimes\mathbb{A}_2,\\
\nabla\mathbf{k} &= \mathbf{u}\otimes\mathbb{B}_2 - \mathbf{v}\otimes\mathbb{A}_3 + \mathbf{h}\otimes\mathbb{A}_1,\\
\nabla\mathbf{h} &= \mathbf{u}\otimes\mathbb{B}_3 + \mathbf{v}\otimes\mathbb{A}_2 - \mathbf{k}\otimes\mathbb{A}_1.
\end{aligned} \tag{2.5}$$

As we are dealing with a trivial bundle the $\mathfrak{iso}(1,3)$-valued curvature Ω vanishes

$$\Omega = \mathrm{d}\omega + \omega\wedge\omega = 0. \tag{2.6}$$

We now split the trivial tangent bundle τ_4^- into the Whitney sum of two nontrivial bundles

$$\tau_4^- = \hat{\tau}_4 \oplus \tilde{\tau}_4. \tag{2.7}$$

In the principal bundle Λ_4^- that splitting of the tangent space of space-time M_4^- is equivalent to the reduction of the 4-frame $\mathbf{e}(x)$ to two 2-frames

$$\mathbf{e}(x) \to \hat{\mathbf{e}}(x)\times\tilde{\mathbf{e}}(x), \tag{2.8}$$

where $\hat{\mathbf{e}}(x) \equiv \{\mathbf{u}(x), \mathbf{v}(x)\}$ spans the distribution $\hat{\Delta}$ and, similarly, the subframe $\tilde{\mathbf{e}}(x) \equiv \{\mathbf{k}(x), \mathbf{h}(x)\}$ spans the distribution $\tilde{\Delta}$. The reduced gauge group $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ is the subgroup of all local Lorentz rotations Λ

$$\mathbf{e}'(x) = \mathbf{e}(x)\cdot\Lambda(x), \quad \Lambda\in\mathrm{SO}(1,1)\times\mathrm{SO}(2), \tag{2.9}$$

which leave invariant the distributions $\hat{\Delta}$ and $\tilde{\Delta}$. The local SO(2) gauge transformation in $\tilde{\tau}_4$ is given by

$$\begin{aligned}
\mathbf{k}' &= \cos\alpha\,\mathbf{k} + \sin\alpha\,\mathbf{h},\\
\mathbf{h}' &= -\sin\alpha\,\mathbf{k} + \cos\alpha\,\mathbf{h},
\end{aligned} \tag{2.10a}$$

whereas the local SO(1,1) gauge transformation in $\hat{\tau}_4$ reads

$$\begin{aligned}
\mathbf{u}' &= \cosh\beta\,\mathbf{u} + \sin\beta\,\mathbf{v},\\
\mathbf{v}' &= \sinh\beta\,\mathbf{u} + \cosh\beta\,\mathbf{v},
\end{aligned} \tag{2.10b}$$

where $\alpha(x)$ and $\beta(x)$ are space-time dependent gauge functions. Under the composite local $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ gauge transformation the potentials introduced in Eq. (2.2) transform as

$$\begin{aligned}
\mathbb{A}' &= \mathbb{A}_1 + \mathrm{d}\alpha,\\
\mathbb{A}_2' &= \cos\alpha\cosh\beta\,\mathbb{A}_2 + \sin\alpha\cosh\beta\,\mathbb{A}_3\\
&\quad + \sin\alpha\sinh\beta\,\mathbb{B}_2 - \cos\alpha\sinh\beta\,\mathbb{B}_3,\\
\mathbb{A}_3' &= -\sin\alpha\cosh\beta\,\mathbb{A}_2 + \cos\alpha\cosh\beta\,\mathbb{A}_3\\
&\quad + \cos\alpha\sinh\beta\,\mathbb{B}_2 + \sin\alpha\sinh\beta\,\mathbb{B}_3,\\
\mathbb{B}_1' &= \mathbb{B}_1 + \mathrm{d}\beta,\\
\mathbb{B}_2' &= -\sin\alpha\sinh\beta\,\mathbb{A}_2 + \cos\alpha\sinh\beta\,\mathbb{A}_3\\
&\quad + \cos\alpha\cosh\beta\,\mathbb{B}_2 + \sin\alpha\cosh\beta\,\mathbb{B}_3,\\
\mathbb{B}_3' &= -\cos\alpha\sinh\beta\,\mathbb{A}_2 - \sin\alpha\sinh\beta\,\mathbb{A}_3\\
&\quad - \sin\alpha\cosh\beta\,\mathbb{B}_2 + \cos\alpha\cosh\beta\,\mathbb{B}_3.
\end{aligned} \tag{2.11}$$

In order to construct connection forms in the subbundles $\hat{\tau}_4$ and $\tilde{\tau}_4$, the canonical flat connection ω is projected onto the $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ subalgebra of the Lorentz algebra

$$\begin{aligned}
\hat{\omega} &= \hat{P}\cdot\omega\cdot\hat{P} = \mathbb{B}_1 t^1,\\
\tilde{\omega} &= \tilde{P}\cdot\omega\cdot\tilde{P} = \mathbb{A}_1 T^1.
\end{aligned} \tag{2.12}$$

The nonvanishing curvatures of the two reduced bundles are given by

$$\hat{\Omega} = \mathrm{d}\mathbb{B}_1 t^1 = (-\mathbb{A}_2\wedge\mathbb{B}_3 + \mathbb{A}_3\wedge\mathbb{B}_2)t^1, \tag{2.13a}$$

$$\tilde{\Omega} = \mathrm{d}\mathbb{A}_1 T^1 = (-\mathbb{A}_2\wedge\mathbb{A}_3 + \mathbb{B}_2\wedge\mathbb{B}_3)T^1. \tag{2.13b}$$

In writing the curvature coefficients as on the right-hand side of (2.13) we made use of the vanishing of the $\mathfrak{iso}(1,3)$ curvature (2.6).

It can easily be checked that under the joint $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ gauge transformation the curvature fields $\mathrm{d}\mathbb{A}_1$ and $\mathrm{d}\mathbb{B}_1$ are invariant, as expected. If the specific bundle as constructed in Ref. 1 is used, the curvature $\tilde{\Omega}$ in the SO(2) bundle can be identified with the dual $^*\mathbb{F}$ of the electromagnetic field of a single point charge

$$^*\mathbb{F} = \mathrm{d}\mathbb{A}_1. \tag{2.14}$$

There it was also shown that the curvature $\hat{\Omega}$ in the normal bundle $\hat{\tau}_4$ is equal to the radiation part ${}^r\mathbb{F}$ of the single particle field

$${}^r\mathbb{F} = \mathrm{d}\mathbb{B}_1. \tag{2.15}$$

Keeping in mind this physical meaning of the geometric quantities we resume the investigation of the general geometric structure without referring to a single particle. The latter will serve only as an example to demonstrate our more general statements. It will be useful to introduce the new potentials

$$\begin{aligned}
\mathbb{K} &= -\mathbb{A}_3 - \mathbb{B}_2,\\
\mathbb{H} &= \mathbb{A}_2 - \mathbb{B}_3,
\end{aligned} \tag{2.16}$$

which under local $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ gauge transformations turn into

$$\begin{aligned}
\mathbb{K}' &= e^{\beta}(\cos\alpha\,\mathbb{K} + \sin\alpha\,\mathbb{H}),\\
\mathbb{H}' &= e^{\beta}(-\sin\alpha\,\mathbb{K} + \cos\alpha\,\mathbb{H}).
\end{aligned} \tag{2.17}$$

Thus we get

$$\mathbb{K}'\wedge\mathbb{H}' = e^{2\beta}\mathbb{K}\wedge\mathbb{H}, \tag{2.18}$$

from which it follows that the (**K**, **H**)-plane remains invariant under gauge transformations.

Under pure SO(2) transformations the potentials transform like the base vectors **k** and **h** of the original tetrad. This observation hints at the possibility that the same 2-plane $\tilde{\Delta}$ may be spanned by the 2-form $\mathbb{k} \wedge \mathbb{h}$ as well as by $\mathbb{K} \wedge \mathbb{H}$.

Equation (2.17) further tells us that under pure SO(1,1) transformations the potentials $\mathbb{H}$ and $\mathbb{K}$ transform like the inverse of a lightlike distance ρ. The problem of assigning a distance $\rho(x,y)$ to two events **x**,**y** separated by a lightlike vector $\mathbf{N} = \mathbf{x} - \mathbf{y}$ $(\mathbf{N}\cdot\mathbf{N} = 0)$ can be solved by first choosing some timelike vector **u**. Then a distance ρ between these two events can be defined as[2]

$$\rho(x,y) = (\mathbf{u}\cdot\mathbf{N}). \tag{2.19}$$

Once a time axis **u** has been chosen there is a spacelike unit vector **v** in the (**u**,**N**)-plane such that

$$\begin{aligned} &(\mathbf{u}\cdot\mathbf{v}) = 0, \\ &\mathbf{N} = \rho(\mathbf{u} + \mathbf{v}) =: \rho\mathbf{n} \quad (\mathbf{n}\cdot\mathbf{n}) = 0. \end{aligned} \tag{2.20}$$

Obviously, the length ρ of the vector **N** strongly depends on the choice of the time axis **u**. At this point the question arises as to how ρ changes when **u** is rotated according to an SO(1,1) transformation (2.10b). A simple calculation yields the following transformation property for ρ:

$$\rho' = (\mathbf{N}\cdot\mathbf{u}') = e^{-\beta}\rho. \tag{2.21}$$

Since the dimension of the potentials $\mathbb{H}$ and $\mathbb{K}$ is that of an inverse length, it is tempting to consider the dilation factor e^{β} of the SO(1,1)-transformed potentials $\mathbb{H}'$, $\mathbb{K}'$ [(2.16)] as originating from the inverse length constituent of those potentials. This assumption is easily verified for a single particle[1] where one finds indeed

$$\mathbb{K} = \rho^{-1}\mathbb{k}, \quad \mathbb{H} = \rho^{-1}\mathbb{h}. \tag{2.22}$$

Furthermore, when we eliminate $\mathbb{A}_2$ and $\mathbb{A}_3$ in the expression for the total field (2.13b) in favor of $\mathbb{H}$ and $\mathbb{K}$ we are left with a new form for the total field in the general case

$$^*\mathbb{F} = \mathbb{H} \wedge \mathbb{K} + \mathbb{H} \wedge \mathbb{B}_2 - \mathbb{K} \wedge \mathbb{B}_3. \tag{2.23}$$

Similarly we get for the curvature in the normal bundle (2.13a)

$$^r\mathbb{F} = \mathbf{d}\mathbb{B}_1 = \mathbb{B}_3 \wedge \mathbb{H} + \mathbb{B}_2 \wedge \mathbb{K}. \tag{2.24}$$

In the single particle case the terms bilinear in $\mathbb{H}$, $\mathbb{K}$ constitute the bound field $^b\mathbb{F} \sim \rho^{-2}$, whereas the terms containing $\mathbb{H}$ and $\mathbb{K}$ linearly make up the radiation part $^r\mathbb{F} \sim \rho^{-1}$.

Let us now exhibit another aspect of the potentials $\mathbb{H}$ and $\mathbb{K}$. By again exploiting the vanishing of the curvature 2-form in the original SO(1,3) bundle we find

$$\mathbf{d}(\mathbb{K} \wedge \mathbb{H}) = 2\mathbb{B}_1 \wedge \mathbb{K} \wedge \mathbb{H}. \tag{2.25}$$

This is just a Frobenius integrability condition[3]: it states that the distribution $\Delta^{\perp}$, say, orthogonal to the (**K**,**H**)-plane, is always integrable, no matter how the splitting of the tangent space was performed. Thus it seems that it is the distribution spanned by the vectors **H** and **K**, rather than $\tilde{\Delta}$, that is relevant for geometrically describing an electromagnetic field configuration by means of a system of two-dimensional surfaces in space-time. In the Appendix we give an example where this conjecture is shown to be true.

In the single particle case the orthogonal distribution $\Delta^{\perp}$ is spanned by the vectors $\mathbf{u}(x)$, $\mathbf{v}(x)$ and hence coincides with the distribution $\hat{\Delta}$. In the general situation, however, one cannot expect the distribution $\hat{\Delta}$ to be tangent to the integral surfaces of $\Delta^{\perp}$. Nonetheless, to require integrability of the distribution $\hat{\Delta}$ is one of the assumptions we need in order to get the desired relation between SO(1,1) curvature $^r\mathbb{F}$ and SO(2) mean curvature $\tilde{\sigma}$. For this reason it becomes necessary to study Frobenius integrability conditions in more detail.

III. INTEGRABILITY CONDITIONS

Having found an always existing integral submanifold in our system under investigation, we now present conditions for the distributions $\tilde{\Delta}$ and $\hat{\Delta}$ to be integrable as well. We once again make use of the Frobenius theorem in the version for differential forms[3] and find

$$\begin{aligned} &\mathbf{d}(\mathbb{h} \wedge \mathbb{k}) \\ &\quad = \mathbf{u} \wedge \mathbf{v} \wedge \{[\mathbb{B}_3(\mathbf{n}) + \mathbb{H}(\mathbf{u})]\mathbb{k} - [\mathbb{B}_2(\mathbf{n}) + \mathbb{K}(\mathbf{u})]\mathbb{h}\} \\ &\qquad + \mathbb{k} \wedge \mathbb{h} \wedge \{-[\mathbb{B}_3(\mathbf{h}) + \mathbb{B}_2(\mathbf{k})]\mathbf{n} \\ &\qquad - [\mathbb{K}(\mathbf{k}) + \mathbb{H}(\mathbf{h})]\mathbf{v}\}, \end{aligned} \tag{3.1}$$

where $\mathbb{n} := \mathbf{u} + \mathbf{v}$ is a lightlike 1-form $[\mathbb{n}(\mathbf{n}) = 0]$. We conclude that $\hat{\Delta}$ is involutive if

$$\begin{aligned} &\mathbb{B}_3(\mathbf{n}) + \mathbb{H}(\mathbf{u}) = 0, \\ &\mathbb{B}_2(\mathbf{n}) + \mathbb{K}(\mathbf{u}) = 0. \end{aligned} \tag{3.2}$$

The analogous consideration for the distribution $\tilde{\Delta}$ yields

$$\begin{aligned} \mathbf{d}(\mathbf{u} \wedge \mathbf{v}) = &\{[\mathbb{K}(\mathbf{h}) - \mathbb{H}(\mathbf{k})]\mathbf{u} + [\mathbb{B}_2(\mathbf{h}) \\ &- \mathbb{B}_3(\mathbf{k})]\mathbf{n}\} \wedge \mathbf{k} \wedge \mathbf{h} - \{[\mathbb{K}(\mathbf{v}) + \mathbb{B}_2(\mathbf{n})]\mathbf{k} \\ &+ [\mathbb{B}_3(\mathbf{n}) + \mathbb{H}(\mathbf{v})]\mathbf{h}\} \wedge \mathbf{u} \wedge \mathbf{v}, \end{aligned} \tag{3.3}$$

from which the integrability conditions for $\tilde{\Delta}$ follow:

$$\begin{aligned} &\mathbb{B}_2(\mathbf{h}) - \mathbb{B}_3(\mathbf{k}) = 0, \\ &\mathbb{K}(\mathbf{h}) - \mathbb{H}(\mathbf{k}) = 0. \end{aligned} \tag{3.4}$$

Inspired by the situation in the one-particle case we can think of still another involutive hypersurface. There it is the light cones emanating from the particle's world line that are the integral surfaces of the distribution $\bar{\Delta}$, spanned by the vectors **h**, **k**, **n**. Demanding $\bar{\Delta}$ to be integrable in the general case imposes the following additional constraints on $\mathbb{H}$ and $\mathbb{K}$:

$$\mathbb{K}(\mathbf{n}) = \mathbb{H}(\mathbf{n}) = 0. \tag{3.5}$$

Equations (3.5) imply that if the potentials $\mathbb{H}$, $\mathbb{K}$ possess components extending beyond the codistribution $^*\tilde{\Delta}$, then these residual components in $^*\hat{\Delta}$ point into the null direction $\mathbb{n}$:

$$\begin{aligned} &\mathbb{K} = \tilde{\mathbb{K}} + \hat{K}\mathbb{n}, \\ &\mathbb{H} = \tilde{\mathbb{H}} + \hat{H}\mathbb{n}. \end{aligned} \tag{3.6}$$

Besides the integrability conditions (3.2), (3.4), and (3.5) we

further require that $\tilde{\Delta}$ be umbilical with respect to its light-like normal section $\mathbf{n}$ in $\bar{\Delta}$.[4] This amounts to demanding that the derivative of the normal section $\mathbf{n}$ projected onto $\tilde{\Delta}$ is of the form

$$\tilde{P}^{\mu}{}_{\sigma}\nabla_{\mu}n^{\nu}\tilde{P}^{\lambda}{}_{\nu} = \rho^{-1}\tilde{P}_{\lambda\sigma}, \tag{3.7}$$

where $\tilde{P}$ is the projector onto $\tilde{\Delta}$:

$$\tilde{P}^{\mu}{}_{\nu} = -k^{\mu}k_{\nu} - h^{\mu}h_{\nu}. \tag{3.8}$$

By comparison of (3.7) with the first two of Cartan's structure equations (2.5) it follows that $\tilde{\mathbb{K}}$ and $\tilde{\mathbb{H}}$ [cf. (3.6)] are of equal magnitude and proportional to $\mathbb{k}$ and $\mathbb{h}$, respectively,

$$\tilde{\mathbb{K}} = \rho^{-1}\mathbb{k}, \quad \tilde{\mathbb{H}} = \rho^{-1}\mathbb{h}. \tag{3.9}$$

Thus the general structure of the new potentials becomes

$$\begin{aligned}\tilde{\mathbb{K}} &= \rho^{-1}\mathbb{k} + \hat{K}\mathbb{n},\\ \mathbb{H} &= \rho^{-1}\mathbb{h} + \hat{H}\mathbb{n}.\end{aligned} \tag{3.10}$$

This means the integrability conditions worked out so far do not suffice to confine $\mathbf{H}$ and $\mathbf{K}$ to the distribution $\tilde{\Delta}$. Since this is what we need in order to achieve our goal we have to give a further constraint on how the tangent space splitting must be done. As we shall see, imposing a condition on the mean curvature 1-form $\tilde{\sigma}$ of the normal bundle will do the job.

IV. MEAN CURVATURE

Let us now pass to some further geometric concepts,[4] the torsion 1-form Σ and the mean curvature 1-form σ. The torsion 1–form Σ is defined by taking the covariant derivative of the base 1-forms in the appropriate cotangent bundles $(*\tilde{\tau}_4)$ and $(*\hat{\tau}_4)$, respectively,

$$\begin{aligned}\tilde{\mathbb{D}}\tilde{\mathbb{E}} &= \mathbf{d}\tilde{\mathbb{E}} + \tilde{\omega}\wedge\tilde{\mathbb{E}} =: \tilde{\Sigma}\wedge\tilde{\mathbb{E}},\\ \hat{\mathbb{D}}\hat{\mathbb{E}} &= \mathbf{d}\hat{\mathbb{E}} + \hat{\omega}\wedge\hat{\mathbb{E}} =: \hat{\Sigma}\wedge\hat{\mathbb{E}},\end{aligned} \tag{4.1}$$

where $(a,b,c| = 0,1;\ i,j,k\ | = 2,3)$,

$$\hat{\mathbb{E}} = \{\mathbb{E}^a\} = \begin{pmatrix}\mathbf{u}\\ -\mathbf{v}\end{pmatrix}, \quad \tilde{\mathbb{E}} = \{\tilde{\mathbb{E}}^i\} = \begin{pmatrix}-\mathbb{k}\\ -\mathbb{h}\end{pmatrix}. \tag{4.2}$$

Note that the integrability conditions (3.2) and (3.4) have already been incorporated in Eq. (4.1).

The geometric meaning of the torsion 1-form has been discussed in Ref. 1; thus we restrict ourselves to presenting the results for $\hat{\Sigma}$ and $\tilde{\Sigma}$ in terms of the potentials $\mathbb{H}$, $\mathbb{K}$, and $\mathbb{B}_a$. A convenient decomposition of the torsion 1-forms as given in Ref. 1 reads

$$\tilde{\Sigma} = \mathbb{S}_a\mathbb{E}^a, \quad \hat{\Sigma} = \mathbb{S}_i\hat{\mathbb{E}}^i. \tag{4.3}$$

This representation of the torsion 1-forms shows clearly that $\tilde{\Sigma}$ annihilates the distribution $\tilde{\Delta}$ and likewise $\hat{\Sigma}$ annihilates $\hat{\Delta}$. Taking the convariant derivatives as in Eq. (4.1) yields for the S-matrices

$$\begin{aligned}\mathbb{S}_0 &= \begin{pmatrix}\mathbb{B}_2(\mathbf{k}) & \mathbb{B}_2(\mathbf{h})\\ \mathbb{B}_3(\mathbf{k}) & \mathbb{B}_3(\mathbf{h})\end{pmatrix}, \quad \mathbb{S}_1 = \begin{pmatrix}\mathbb{A}_3(\mathbf{k}) & \mathbb{A}_3(\mathbf{h})\\ -\mathbb{A}_2(\mathbf{k}) & -\mathbb{A}_2(\mathbf{h})\end{pmatrix},\\ \mathbb{S}_2 &= \begin{pmatrix}\mathbb{B}_2(\mathbf{u}) & \mathbb{B}_2(\mathbf{v})\\ -\mathbb{A}_3(\mathbf{u}) & -\mathbb{A}_3(\mathbf{v})\end{pmatrix}, \quad \mathbb{S}_3 = \begin{pmatrix}\mathbb{B}_3(\mathbf{u}) & \mathbb{B}_3(\mathbf{v})\\ \mathbb{A}_2(\mathbf{u}) & \mathbb{A}_2(\mathbf{v})\end{pmatrix}.\end{aligned} \tag{4.4}$$

It follows that, in general, the torsion 1-forms are not Lie algebra valued, i.e., $\hat{\Sigma}(\tilde{\Sigma})$ does not take its values in $\mathfrak{so}(1,1)$ $[\mathfrak{so}(2)]$. However, if we investigate the transformation behavior of the S-matrices we find for the SO(2)-transformed matrices,

$$\mathbb{S}'_a = \mathrm{ad}_{\tilde{\Lambda}^{-1}}\mathbb{S}_a, \quad \mathbb{S}'_i = \mathbb{S}_j\tilde{\Lambda}^{j}{}_{i}, \quad \tilde{\Lambda}\in\mathrm{SO}(2), \tag{4.5}$$

whereas an SO(1,1) transformation $\hat{\Lambda}$ [(2.10b)] changes the S-matrices according to

$$\mathbb{S}'_a = \hat{\Lambda}^{b}{}_{a}\mathbb{S}_b, \quad \mathbb{S}'_i = \mathrm{ad}_{\hat{\Lambda}^{-1}}\mathbb{S}_i, \quad \hat{\imath}\in\mathrm{SO}(1,1). \tag{4.6}$$

Thus we get for the SO(2) [SO(1,1)] transformed torsion 1-form $\tilde{\Sigma}(\hat{\Sigma})$,

$$\tilde{\Sigma}' = \mathrm{ad}_{\tilde{\Lambda}^{-1}}\tilde{\Sigma}, \quad \hat{\Sigma}' = \mathrm{ad}_{\hat{\Lambda}^{-1}}\hat{\Sigma}. \tag{4.7}$$

Equations (4.7) show that although $\tilde{\Sigma}$ and $\hat{\Sigma}$ are not Lie algebra valued 1-forms they nevertheless transform as if they were. Hence it makes sense to take their exterior covariant differentials

$$\tilde{\mathbb{D}}\tilde{\Sigma} := \mathrm{d}\tilde{\Sigma} + \tilde{\omega}\wedge\tilde{\Sigma} + \tilde{\Sigma}\wedge\tilde{\omega} \tag{4.8}$$

(and similarly for $\hat{\Sigma}$), for these covariant derivatives transform as if they were tensorial Lie algebra valued 2-forms. We can derive the first Bianchi identity by observing that the repeated application of the covariant derivative yields, on the one hand, the curvature $\tilde{\Omega}$,

$$\tilde{\mathbb{D}}\tilde{\mathbb{D}}\tilde{\mathbb{E}} = \tilde{\Omega}\wedge\tilde{\mathbb{E}}. \tag{4.9}$$

On the other hand it follows from the first structure equation (4.1) that

$$\tilde{\mathbb{D}}\tilde{\mathbb{D}}\tilde{\mathbb{E}} = (\tilde{\mathbb{D}}\tilde{\Sigma} + \tilde{\Sigma}\wedge\tilde{\Sigma})\wedge\tilde{\mathbb{E}}. \tag{4.10}$$

Combining the last two equations we end up with the first Bianchi identity

$$(\tilde{\mathbb{D}}\tilde{\Sigma} + \tilde{\Sigma}\wedge\tilde{\Sigma})\wedge\tilde{\mathbb{E}} = \tilde{\Omega}\wedge\tilde{\mathbb{E}}. \tag{4.11}$$

As a further consequence of the specific transformation behavior [(4.5), (4.6)] of the S-matrices with respect to the product group $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$, we find that the traces of the torsion 1-forms are invariant under $\mathrm{SO}(1,1)\times\mathrm{SO}(2)$ gauge transformations. Therefore they are characteristics of how the tangent space was broken up. They share this property with the integrability conditions of that tangent space splitting. The traces of the torsion 1-forms are called the mean curvature 1-forms $\hat{\sigma}$, $\tilde{\sigma}$,[4]

$$\hat{\sigma} = \tfrac{1}{2}\,\mathrm{tr}\,\hat{\Sigma}, \quad \tilde{\sigma} = \tfrac{1}{2}\,\mathrm{tr}\,\tilde{\Sigma}. \tag{4.12}$$

Inserting the expressions (4.4) for the S-matrices one readily finds

$$2\tilde{\sigma} = [\mathbb{B}_2(\mathbf{k}) + \mathbb{B}_3(\mathbf{h})]\mathbb{n} + [\mathbb{H}(\mathbf{h}) + \mathbb{K}(\mathbf{k})]\mathbb{v}, \tag{4.13a}$$

$$2\hat{\sigma} = -[\mathbb{K}(\mathbf{v}) + \mathbb{B}_2(\mathbf{n})]\mathbb{k} - [\mathbb{H}(\mathbf{v}) + \mathbb{B}_3(\mathbf{n})]\mathbb{h}. \tag{4.13b}$$

Our third basic requirement besides the conditions of Sec. II to be imposed on the tangent space splitting is to demand the mean curvature $\hat{\sigma}$ of the normal bundle to be zero: $\hat{\sigma}\equiv 0$. As a consequence we find the relations

$$\begin{aligned}\mathbb{K}(\mathbf{v}) &= -\mathbb{B}_2(\mathbf{n}),\\ \mathbb{H}(\mathbf{v}) &= -\mathbb{B}_3(\mathbf{n}).\end{aligned} \tag{4.14}$$

If this result is combined with the integrability conditions

(3.2) for the distribution $\hat{\Delta}$ we obtain

$$\mathbb{K}(\mathbf{m}) = \mathbb{H}(\mathbf{m}) = 0, \tag{4.15}$$

where $\mathbf{m} = -\mathbf{u} + \mathbf{v}$ is a second null vector of $\hat{\Delta}$ independent of $\mathbf{n}$. It follows that the projections of $\mathbb{H}$ and $\mathbb{K}$ onto $^*\hat{\Delta}$ vanish, i.e., the new potentials are entirely contained in the distribution $\tilde{\Delta}$,

$$\begin{aligned} \mathbb{K} &= \rho^{-1}\mathbb{K}, \\ \mathbb{H} &= \rho^{-1}\mathbb{h}. \end{aligned} \tag{4.16}$$

Thus the mean curvature 1-form [(4.13a)] becomes

$$\tilde{\sigma} = \tfrac{1}{2}[\mathbb{B}_2(\mathbf{k}) + \mathbb{B}_3(\mathbf{h})]\,\mathbf{n} - \rho^{-1}\mathbf{v}, \tag{4.17}$$

and hence is included in $^*\hat{\Delta}$.

In the following we need the result that the 1-form $\mathbb{C}$,

$$\mathbb{C} = \rho^{-1}\,\mathrm{d}\rho + \rho^{-1}\mathbf{v} + \mathbb{B}_1, \tag{4.18}$$

is an element of $^*\hat{\Delta}$ as well. For a proof of this statement take twice the covariant derivative of the lightlike section $\mathbf{n}$,

$$\nabla\nabla n^{\mu} = \Omega^{\mu}{}_{\nu} n^{\nu}, \tag{4.19}$$

observe (4.16), and use the fact that the canonical connection ω is flat ($\Omega = 0$). One finds

$$\mathbb{C} \wedge (\mathbf{k} \otimes \mathbf{k} + \mathbf{h} \otimes \mathbf{h}) = \mathbb{B}_2 \wedge \mathbf{n} \otimes \mathbf{k} + \mathbb{B}_3 \wedge \mathbf{n} \otimes \mathbf{h}. \tag{4.20}$$

From this equation it follows that $\mathbb{C}$, and hence $\mathbb{C} - \rho^{-1}\mathbf{v}$, do not have any components in $^*\tilde{\Delta}$,

$$\rho^{-1}\,\mathrm{d}\rho + \mathbb{B}_1 \in {}^*\hat{\Delta}. \tag{4.21}$$

V. RESULT AND DISCUSSION

We close the argument by simply taking the exterior derivative of the 2-form $\mathbb{K} \wedge \mathbb{H}$,

$$\mathbf{d}(\mathbb{K} \wedge \mathbb{H}) = 2\left(\tilde{\sigma} - \frac{\mathbf{d}\rho}{\rho}\right) \wedge \mathbb{K} \wedge \mathbb{H}. \tag{5.1}$$

In doing so we utilized the equality

$$\mathbf{d}(\mathbf{k} \wedge \mathbf{h}) = 2\tilde{\sigma} \wedge \mathbf{k} \wedge \mathbf{h} \tag{5.2}$$

that follows easily from the definition (4.12) of the torsion 1-form.

By comparision of (5.1) with Eq. (2.25) and observing (4.21) we now find that on the assumptions stated above the mean curvature 1-form can be identified with a SO(2) gauge transform of the potential $\mathbb{B}_1$:

$$\mathbb{B}_1 = \tilde{\sigma} - \frac{\mathbf{d}\rho}{\rho}. \tag{5.3}$$

The exterior derivative of $\mathbb{B}_1$ is just the gauge independent curvature $^r\mathbb{F}$ in the SO(1,1) bundle [cf. Eq. (2.15)]. Thus we arrive at the result we set out to prove

$$^r\mathbb{F} = \mathrm{d}\mathbb{B}_1 = \mathrm{d}\tilde{\sigma}, \tag{5.4}$$

namely, the normal bundle curvature $^r\mathbb{F}$ is equal to the exterior derivative of the mean curvature $\tilde{\sigma}$ in the SO(2) bundle.

In retrospect, we realize that essentially three kinds of assumptions were necessary to derive this result: (i) integrability of the distributions $\hat{\Delta}$, $\tilde{\Delta}$, $\bar{\Delta}$; (ii) umbilicality of $\tilde{\Delta}$ in $\bar{\Delta}$; and (iii) the vanishing of the mean curvature $\hat{\sigma}$ in the SO(1,1) bundle. Observe that all these conditions are gauge invariant.

Now the question arises if there is any physical meaning that can be attributed to these purely mathematical conditions. The answer can be given most easily for the case of a pointlike particle.[1] There the requirement on $\bar{\Delta}$ and $\tilde{\Delta}$ to be integrable—the latter being a subdistribution of the former—has to be interpreted as follows: the integral surfaces of $\tilde{\Delta}$ are propagation surfaces of the electromagnetic field, which in the course of time sweep over the light cones emanating from the particle's world line. The light cones themselves are just the integral surfaces of the three-dimensional distribution $\bar{\Delta}$, and hence they are the collection of all propagation 2-surfaces originating in a single event on the particle world line.

We need to look at the integrability of the distribution $\hat{\Delta}$ in some more detail. Let us first remark that, at each point $x \in M_4$, the tangent space splitting induces two null directions $\mathbf{n} = \mathbf{u} + \mathbf{v}$ and $\mathbf{m} = -\mathbf{u} + \mathbf{v}$. Provided the tangent space is decomposed in a smooth way, then both the vector fields $\mathbf{n}(x)$ and $\mathbf{m}(x)$ are integrable. The corresponding integral curves are null trajectories that we may interpret as photon paths ending at or emerging from particle world lines and reaching out to infinity.

On the other hand, the system of photon trajectories that fills the whole of the manifold M_4 prescribes a splitting of the tangent space at each point $x \in M_4$ by defining the distribution $\hat{\Delta}$ via the lightlike vectors $\mathbf{n}$ and $\mathbf{m}$. Integrability of that distribution $\hat{\Delta}$ means that the entirety of photon lines can be foliated into 2-sheets, the integral surfaces of $\hat{\Delta}$.

In this context the following observation is worth noticing. Taking the covariant derivative $\hat{\nabla}$ with respect to the reduced connection $\hat{\omega}$ in the SO(1,1) bundle of the photon line tangent vectors $\hat{n}$ and $\hat{m}$ yields

$$\begin{aligned} \hat{\nabla}\mathbf{n} &= \mathbf{n} \otimes \mathbb{B}_1, \\ \hat{\nabla}\mathbf{m} &= -\mathbf{m} \otimes \mathbb{B}_1. \end{aligned} \tag{5.5}$$

Interpreted geometrically Eqs. (5.5) state that the photon lines parallel transport their own tangent vectors. Thus the photon trajectories are the null geodesics in the integral 2-surfaces of $\hat{\Delta}$. In other words, photons travel in a minimal way on surfaces that themselves are minimal in the well-known geometrical sense that their mean curvature $\hat{\sigma}$ vanishes. This condition of $\hat{\sigma}$ being identically zero is just the one we had to impose on our system. It remains to be investigated whether any real electromagnetic configuration leads to a similar geometric "minimal" structure.

APPENDIX: A STATIC TWO-PARTICLE CONFIGURATION

In order to once more demonstrate the geometric significance of the potentials $\mathbb{H}$ and $\mathbb{K}$ we study as a further example the geometry of two static point particles located on the x-axis at equal distances from the origin. Because of the time independence of the system we now consider the principal bundle of orthonormal 3-frames where the rotation group SO(3) serves as the structural group. As a base space

we take the modified Euclidean space E_3^- which differs from the ordinary E_3 by having two points excluded where the particles reside. Since we restrict ourselves to the static case we look for a distribution $\tilde{\Delta}$ in E_3^- which is uniquely characterized by its spacelike unit normal $\mathbf{v}$ contained in E_3^-. The curvature in the SO(2) subbundle $\tilde{\Delta}$ is then to be identified with the dual of the electromagnetic field of the two-point particles. From the unit normal $\mathbf{v}(x)$,

$$\mathbf{v} = \cos\Theta\,\mathbf{E}_1 + \sin\Theta\cos\Phi\,\mathbf{E}_2 + \sin\Theta\sin\Phi\,\mathbf{E}_3$$
$$(\nabla\mathbf{E}_i = 0, \quad i = 1,2,3), \tag{A1}$$

the standard gauge $\tilde{\mathbf{e}}(x) = \{\mathbf{k}(x), \mathbf{h}(x)\}$ in the SO(2) bundle is readily obtained as

$$\mathbf{k}(x) = \frac{\partial\mathbf{v}}{\partial\Theta}, \quad \mathbf{h}(x) = \frac{1}{\sin\Theta}\frac{\partial\mathbf{v}}{\partial\Phi}. \tag{A2}$$

A special distribution $\tilde{\Delta}$ is fixed by putting

$$\Phi = 2\varphi,$$
$$\cos\Theta = p\cos\theta_1 + q\cos\theta_2, \tag{A3}$$
$$P + q = 1,$$
$$p, q| = \text{const}.$$

Here θ_1 (θ_2) denotes the angle between the x-axis and the radius vector $\mathbf{r}_1$ ($\mathbf{r}_2$) that reaches from the locus of particle 1 (2) to an arbitrary point $x \in E_3^-$. φ is just the azimuthal angle of ordinary spherical polar coordinates (see Fig. 1).

With the specific cross section $\tilde{\mathbf{e}}(x)$ of Eq. (A2) fixed we can calculate the corresponding connection coefficients in the SO(3) bundle:

$$\mathbb{A}_1 = \cos\Theta\,\mathrm{d}\varphi,$$
$$\mathbb{A}_2 = -\sin\Theta\,\mathrm{d}\varphi, \tag{A4}$$
$$\mathbb{A}_3 = \mathrm{d}\Theta.$$

The curvature 2-form in the reduced SO(2) bundle follows from (A4) analogously to the procedure in Sec. II:

$$^*\mathbb{F} = \mathrm{d}\mathbb{A}_1 = -\mathbb{A}_2\wedge\mathbb{A}_3 = -\sin\Theta\,\mathrm{d}\Theta\wedge\mathrm{d}\Phi. \tag{A5}$$

Inserting (A3) yields

$$^*\mathbb{F} = -2p\sin\theta_1\,\mathrm{d}\theta_1\wedge\mathrm{d}\varphi - 2q\sin\theta_2\,\mathrm{d}\theta_2\wedge\mathrm{d}\varphi, \tag{A6}$$

which we recognize as the sum of the individual electric fields of two particles of charges $2p$ and $2q$, respectively. Therefore Maxwell's equations are satisfied automatically:

$$\mathrm{d}^*\mathbb{F} = \delta\,^*\mathbb{F} = 0. \tag{A7}$$

Computing the charge contained in any 2-cycle C^2 via

$$Z = \frac{1}{4\pi}\oint_{C^2} {}^*\mathbb{F}, \tag{A8}$$

we find $Z = -2$ if both the points excluded from E_3 are circumscribed by C^2. In this case the expression (A8) is just the winding number of the mapping (A1) along with (A3), $\mathbf{v}$: $C^2 \to S^2$, and thus it is automatically an integer. If we try to calculate the charge of only one particle there emerges a line singularity on the x-axis where the mapping $\mathbf{v}$ is not defined. In this way $\mathbf{v}$ maps noncompact manifolds onto each other and therefore no winding number exists any longer. Consequently the individual charges $2p$ and $2q$ are not quantized.

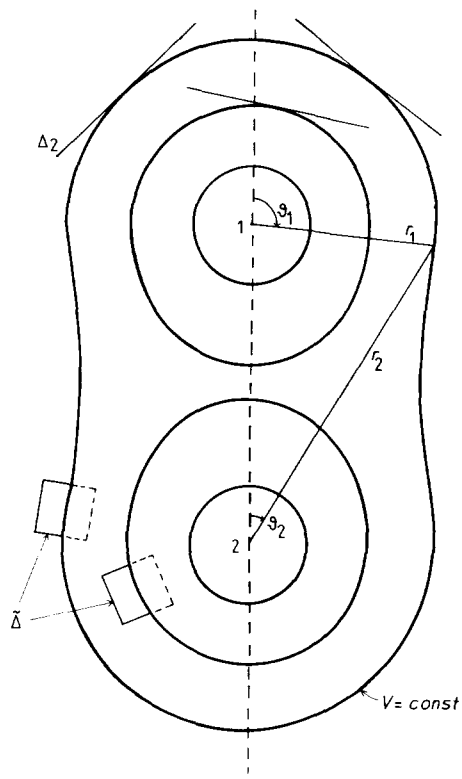


FIG. 1. Cross section of the surfaces of constant potential $V = 2p/r_1 + 2q/r_2$. The distribution Δ_2 spanned by the vectors $\mathbf{K},\mathbf{H}$ is tangent to the equipotential surfaces whereas, in general, the planes of the distribution $\tilde{\Delta}$ intersect the integral manifold V of Δ_2.

However, as will be shown in a forthcoming paper, the singularity can be removed, thereby providing a condition on the numbers p and q such that the individual changes become integers as well. In the given configuration of two particles it is easy to see from Eq. (A6) that if the values $p = q = \frac{1}{2}$ are adopted both the particles will assume a charge of -1. If we want to describe particles of opposite charge we simply have to insert a relative minus sign between the cosines of Eq. (A3) which gives us a total charge of zero.

If we take a closer look at the distribution spanned by the vectors $\mathbf{k}$ and $\mathbf{h}$ we find, in contrast to the one-particle case,[1] that $\tilde{\Delta}$ is no longer integrable, i.e., $[\mathbf{h},\mathbf{k}] \notin \tilde{\Delta}$. On the other hand, from the physical point of view, there must exist integral manifolds in the system, namely, the surfaces of constant electrostatic potential. If we examine the potentials $\mathbb{H}$ and $\mathbb{K}$ that in our time-independent configuration reduce to (observe $\mathbb{B}_a \equiv 0$)

$$\mathbb{H} = \mathbb{A}_2, \quad \mathbb{K} = -\mathbb{A}_3, \tag{A9}$$

we find

$$^*(\mathbb{K}\wedge\mathbb{H}) = \mathrm{d}(1/r_1 + 1/r_2). \tag{A10}$$

In geometric terms Eq. (A10) signifies that the planes spanned by $\mathbf{K}$ and $\mathbf{H}$ are tangent to the equipotential surfaces. We recall from Eq. (2.25) that in the general four-dimensional case there always exist additional 2-surfaces perpendicular to the $(\mathbf{K},\mathbf{H})$-distribution. The lesson we learn from this simple example is the following: if we want to consider classical point-particle electrodynamics as a geometric

structure in space-time, the relevant surfaces within this structure are determined by the (**K**,**H**)-planes rather than by the distribution $\tilde{\Delta}$ which originally was used to generate the geometric structure. It seems rather incidental that in the single-particle case the (**K**,**H**)-planes coincide with the (**k**,**h**)-planes [cf. Eq. (4.8)].

[1]M. Sorg, Z. Naturforsch. A **37**, 201 (1982).
[2]J. L. Synge, Ann. Math. **84**, 33 (1970).
[3]C. von Westenholz, *Differential Forms in Mathematical Physics* (North-Holland, Amsterdam, 1978).
[4]B. Chen, *Geometry of Submanifolds* (Marcel Dekker, New York, 1973).

Equivalent random force and time-series model in systems far from equilibrium

Kuniharu Kishida
Department of Applied Mathematics, Faculty of Engineering, Gifu University, 1-1 Yanagido, Gifu 501-11, Japan

(Received 20 July 1983; accepted for publication 9 December 1983)

Under the condition that observed time-series data is given, a stochastic Markovian equation for a physical system can be transformed into an observable non-Markovian equation used in the time-series analysis. The physical random force satisfying the fluctuation dissipation theorem is also transformed into a stochastically equivalent random force in the derivation of the time-series model of observable variables. Statistical quantities, i.e., correlation and power spectral density functions for observable variables, can be expressed not only by the physical random force, but also by the equivalent random force. A relation between the variance of physical random force and that of equivalent random force is also found.

PACS numbers: 02.50. + s, 05.70.Ln

I. INTRODUCTION

This is the second paper of a series in which we hope to bridge the gap between a stochastic model in physics and a time-series model in statistics or control, and clarify the physics involved in the time-series model in systems far from equilibrium. We will consider in this paper macroscopic steady-state systems such as chemical plants and nuclear power reactors. They are open systems in a far from equilibrium state and usually have complicated reaction networks. Though a time-series model is often used for the identification of system or for the understanding of a situation of reaction processes, the physical properties of model is not clear because of its methodical nature. It treats a complicated practical system as a black box; a time-series model is determined from observed data in the manner as the method of least-squares analysis. Therefore, a physical foundation of time-series model is needed for actual system identification and making a reliable diagnosis possible.

As in the previous paper[1] (hereafter denoted as I), we derived a time-series model which is called the AR-MA (autoregressive moving average) model from a "physical" [2] Langevin equation, by using the coarse graining in time, the elimination of irrelevant variables, and the projection of relevant variables on a space spanned by observable variables. That is, in the physical approach,

$$\begin{aligned} P\circ E\circ C\ [\text{Langevin equation}] &= P\circ E\ [\text{State equation}] \\ &= P\ [\text{Non-Markovian equation}] \\ &= \text{AR-MA model}, \end{aligned}$$

or, in the control or mathematical approach,

$$\begin{aligned} E\circ P\circ C\ [\text{Langevin equation}] &= E\circ P\ [\text{State equation}] \\ &= E\ [\text{Kalman filter}] \\ &= E\ [\text{Markovian representation}] \\ &= \text{AR-MA model}, \end{aligned}$$

where P denotes the projection on the observable variable space, E the elimination of irrelevant variables, and C the time coarse graining operator.

One key point of the derivation of AR-MA model from the physical model is the introduction of an innovation statistically equivalent to the physical random force. Since the first and the second moments are equivalent to the probability distribution in a Gaussian process, the innovation can be defined so as to coincide with the first and the second moments of physical random force. Hence, we call hereafter the innovation the equivalent random force. In this paper, we will show the equivalence in terms of the correlation function and the power spectral density, since the correlation function and the power spectral density are the most fundamental statistical quantities in the statistical analysis of physical and engineering fields. It is important to know an essential role or a mathematical meaning of equivalent random force in the statistical physics. It is well known that the random force of physical model satisfies the generalized Einstein relation (I.21), which is one of the expressions called the fluctuation dissipation theorem in the statistical physics. On the other hand, the equivalent random force satisfies the Riccati type equation (I.25), as mentioned in I. Hence, we will clarify the relationship between both equations for the physical understanding of AR-MA model.

There are mainly two different approaches in the practical analysis of chemical and/or nuclear plants. One is the Kalman filter method[3] for the coefficient or parameter estimation of physical model, and the other is the time series method[4] for the identification of physical processes. Then, a systematic viewpoint is necessary for the analysis or diagnosis of macroscopic systems. As mentioned above, these approaches are related with each other. In Sec. II, we will show the foundation of the autoregressive moving average model from the Kalman filter explicitly. In Sec. III, we will first clarify a relation between variances of physical random force and equivalent random force, and then show the mathematical properties of the equivalent random force in terms of correlation functions. We will examine roles of equivalent

 0022-2488/84/051308-06$02.50

random force in the power spectral density in Sec. IV. The final section is for concluding remarks.

II. CONTRACTION OF INFORMATION AND PHYSICAL TIME-SERIES MODEL

By using the system size expansion method[5] and taking a normal scaling,[6] macroscopic systems can have these mathematical properties[7]: (1) Markovian process, (2) linearity, (3) Gaussian process, and (4) whiteness of random force. That is, a physical model in stochastic systems far from equilibrium is a linear Langevin equation with a white Gaussian random force. A sufficient number of state variables satisfying the Langevin equation are needed for the Markovian description of a macroscopic system. In the system far from equilibrium such as chemical and/or nuclear plants operated in a steady state, less than needed variables are usually measured. Owing to this observation, we have a problem of the contraction of information in macroscopic systems as mentioned in I. Moreover, signals in practical plants such as a cement rotary kiln and a boiler of a power station are sampled with the interval Δt and processed by a digital computer for quick handling of large number of time-series data. Therefore, we have the time coarse grained basic equations [cf. Eqs. (I.5) and (I.6)]:

$$x(n) = \Phi x(n-1) + f(n), \tag{1}$$

$$y(n) = Hx(n). \tag{2}$$

Equation (1) is the discrete-time state equation, $x(n)$ a d-dimensional state variable at time n, Φ the time coarse grained regression $d \times d$ matrix (rank $\Phi = d$), H the observation $q \times d$ matrix ($q < d$, rank $H = q$), $y(n)$ a q-dimensional observable state variable, and a d-dimensional Gaussian random force $f(n)$ has the following statistical properties: $E\{f(n)f(m)^T\} = V\delta_{nm}$, $E\{f(n)\} = 0$, where $E\{\cdots\}$ is the ensemble average, V the time coarse grained diffusion $d \times d$ matrix (rank $V = d$), T the tranposition, and δ_{nm} the Kronecker's symbol.

In I, a physical time-series model was derived from the state equation (1) by using the projection upon an observable state space and the elimination of irrelevant variables. However, the derivation of AR-MA model from the state equation via a Kalman filter was skipped in I. Hence, we will show it explicitly under the stationarity assumption, i.e., the system is stationary with initial time in the infinitely remote past.

Projecting the state equation (1) on the observable state space, we obtain the Kalman filter (I.24)

$$x(n|n) = \Phi x(n-1|n-1) + P(\infty)H^T\Gamma(\infty)^{-1}\gamma(n), \tag{3}$$

where $x(n|n)$ is the least-square estimator under the condition that $Y(n)^T = [y(n)^T, y(n-1)^T, y(n-2)^T, \cdots]$ is given:

$$x(n|n) = E\{x(n)|Y(n)\},$$

and where the innovation is defined by

$$\gamma(n) := y(n) - y(n|n-1),$$

$$E\{\gamma(n)\gamma(n)^T\} := \Gamma(n) = HP(n)H^T, \tag{4}$$

$$P(n) := E\{(x(n) - x(n|n-1))(x(n) - x(n|n-1))^T\},$$

where $P(n)$ satisfies the Riccati equation (I.25)

$$P(n) = \Phi\{P(n-1) - P(n-1)H^T\Gamma(n-1)^{-1}HP(n-1)\}\Phi^T + V, \tag{5}$$

and $P(\infty) = \lim_{n\to\infty} P(n)$. It should be noted that a steady-state solution of evolution equation (5) with an almost everywhere initial value corresponds to a stable fixed point in an aged system, though we have fixed points of the nonlinear equation (5). We have used the stable solution as $P(\infty)$ in Eq. (3). In the derivation of Eq. (3), we have used the following formula (7). Suppose Y_a and Y_b are independent random variables and let X be another random variable. If we set

$$X_{ab} = \int XP\{X|Y_a, Y_b\}dX,$$

$$X_a = \int XP\{X|Y_a\}dX,$$

$$X_b = \int xP\{X|Y_b\}dX,$$

we can have

$$\int X_{ab}\, P\{Y_a, Y_b\}\, dY_a dY_b = \int X_a\, P\{Y_a\}\, dY_a + \int X_b\, P\{Y_b\}\, dY_b, \tag{6}$$

from the relation of probability density functions:

$$P\{X|Y_a, Y_b\} := \frac{P\{X, Y_a, Y_b\}}{P\{Y_a, Y_b\}} = \frac{P\{X, Y_a\} + P\{X, Y_b\}}{P\{Y_a\}P\{Y_b\}} = \frac{P\{X|Y_a\}}{P\{Y_b\}} + \frac{P\{X|Y_b\}}{P\{Y_a\}}.$$

From Eq. (6), a linear estimator of X with given Y_a and Y_b can be evaluated by

$$E\{X|Y_a, Y_b\} = E\{X|Y_a\} + E\{X|Y_b\}. \tag{7}$$

The observation equation (2) is also projected as

$$y(n) = Hx(n|n). \tag{8}$$

The next step for the derivation of AR-MA model is the elimination of irrelevant variables. The contraction of information in a macroscopic system is expressed by the elimination of $x(n|n)$ from Eqs. (3) and (8). This elimination of state variables brings the basic equation the non-Markovian nature as in the damping theory of quantum mechanics.[8] The transformation of Luenberger method[9] is useful in this procedure because of the minimum realization of model. We define a transformation matrix

$$T = [h_1, (\Phi^T)h_1, \ldots, (\Phi^T)^{\sigma_1 - 1}h_1, \ldots, h_q, \ldots, (\Phi^T)^{\sigma_q - 1}h_q]^T, \tag{9}$$

where $\Sigma_{i=1}^{q}\sigma_i = d$ and h_i^T is the ith row vector of H, and where σ_i is the Kronecker index so that rank $T = d$. Defining a new estimator $x'(n|n)$ through the inverse matrix $T^{-1}(= S)$ as $x(n|n) = Sx'(n|n)$, we can transform the basic equations (3) and (8) into the Markovian representation (I.27):

$$x'(n|n) = \Phi' x'(n-1|n-1) + F'\gamma(n), \tag{10}$$

and the observation equation (I.28):

$$y(n) = H' x'(n|n), \tag{11}$$

where the coefficient matrices are given by

$$\Phi' = T\Phi S, \quad F' = TP(\infty)H^T\Gamma(\infty)^{-1}, \quad \text{and} \quad H' = HS.$$

For the component-wise representation of Eqs. (10) and (11), we use the notations

$$x'(n|n) = \begin{pmatrix} x'_1(n|n) \\ \vdots \\ x'_q(n|n) \end{pmatrix}, \quad \text{where } x'_i(n|n) = \begin{pmatrix} x'_{i,1}(n|n) \\ \vdots \\ x'_{i,\sigma_i}(n|n) \end{pmatrix} \quad (i = 1,2,...,q),$$

and the time shift operator z

$$zx'(n|n) = x'(n+1|n+1).$$

Then, Eqs. (10) and (11) are rewritten as

$$zx'_{i,k}(n|n) = x'_{i,k+1}(n|n) + z\sum_{j=1}^{q} b_{ij,k}\,\gamma_j(n) \quad (k = 1,2,...,\sigma_i - 1),$$

$$zx'_{i,\sigma_i}(n|n) = \sum_{j=1}^{q}\sum_{l=1}^{\sigma_j} a_{ij,l}\,x'_{j,\sigma_j+1-l}(n|n) + z\sum_{j=1}^{q} b_{ij,\sigma_i}\,\gamma_j(n),$$

$$y_i(n) = x'_{i,1}(n|n) \quad (i = 1,2,...,q),$$

where

$$a_{ij,l} = h_i^T \Phi^{\sigma_i} s_{k_j - l + 1} \quad \left(k_j = \sum_{i=1}^{j}\sigma_i\right),$$

s_i is the ith column vector of S, and $b_{ij,l}$ is the $(k_{i-1}+l, j)$th element of the $d \times q$ matrix F'. In the similar manner mentioned in Appendix of I, we can obtain the physical AR-MA model with the equivalent random force by eliminating $x'(n|n)$ from the component-wise equations:

$$A(z)y(n) = B(z)\gamma(n), \tag{12}$$

where $A(z) = [A_{ij}(z)]$, $B(z) = [B_{ij}(z)]$, and

$$A_{ij}(z) = z^{\sigma_i}\delta_{ij} - \sum_{l=1}^{\sigma_j} a_{ij,l} z^{\sigma_j - l},$$

$$B_{ij}(z) = \sum_{l=0}^{\sigma_i - 1} b_{ij,\sigma_i - l} z^{l+1} - \sum_{m=1}^{q}\sum_{l=1}^{\sigma_m}\sum_{k=1}^{\sigma_m - l} a_{im,l} b_{mj,\sigma_m + 1 - l - k}\, z^k \quad (i,j = 1,2,...,q).$$

Putting $M = \max(\sigma_i)$ and setting $A(z) = \sum_{i=0}^{M} A_i z^{M-i}$, and $B(z) = \sum_{i=0}^{M-1} B_i z^{M-i}$, we can symbolically rewrite Eq. (12) as the AR-MA model of order $(M, M-1)$:

$$\sum_{i=0}^{M} A_i\, y(n-i) = \sum_{i=0}^{M-1} B_i\,\gamma(n-i). \tag{13}$$

It is concluded that the steady-state system described by the linear, Markov-Gaussian state equation (1) with the whiteness of random force and the observable condition (9) has the AR-MA representation (12) or (13), when we have measured the system by the observation matrix (2).

III. CORRELATION FUNCTION AND EQUIVALENT RANDOM FORCE

In Sec. II, we have shown the relationship among the state equation, the Kalman filter, the Markovian representation, and the AR-MA model. In the derivation, we have often used mathematical properties of the Gaussian process in a steady macroscopic system. In the Gaussian process the first and the second moments or mean values and correlation functions are equivalent to the probability distribution function. Namely, time-varying statistical quantities are, in a steady Gaussian process, limited to correlation functions. In this section, we will examine correlation functions obtained from physical system from the viewpoint of the contraction of information, and show that the innovation in Eqs. (3), (10), and (13) is equivalent to the physical random force in Eq. (1).

The system is assumed to be in a stable and steady state. The eigenvalues of Φ lie inside the unit circle. Since $x(n)$ is the fluctuation, we can have the zero mean values: $E\{x(n)\} = 0$. It is well known that, from Eq. (1), the correlation function matrix of $x(n)$ is

$$C_{xx}(k,l) := E\{x(k)x(l)^T\} = C_{xx}(k-l) = E\{x(k-l)x(0)^T\} = \begin{cases} \Phi^{k-l}C_{xx}(0) & k \geqslant l, \\ C_{xx}(0)(\Phi^T)^{l-k} & k \leqslant l, \end{cases} \tag{14}$$

where $C_{xx}(0)$ satisfies the generalized Einstein relation (I.21):

$$C_{xx}(0) = \Phi C_{xx}(0)\Phi^T + V. \tag{15}$$

Equation (15) corresponds to the fluctuation dissipation theorem in the discrete time sampling, and is also called the Lyapunov equation in the control theory. The solution of Eq. (15) is formally given by

$$C_{xx}(0) = \sum_{i=0}^{\infty} \Phi^i V(\Phi^T)^i. \tag{16}$$

From Eq. (2), we have the correlation function matrix of observable variables

$$C_{yy}(k-l) := E\{y(k)y(l)^T\} = HE\{x(k)x(l)^T\}H^T = \begin{cases} H\Phi^{k-l}C_{xx}(0)H^T & k \geqslant l, \\ HC_{xx}(0)(\Phi^T)^{l-k}H^T & k \leqslant l. \end{cases} \tag{17}$$

It is obvious from Sec. II that the partial correlation function matrix C_{yy} obtained from the state equation must be identified with that of the Kalman filter, of the Markovian representation, or of the AR-MA model. We will prove the identification explicitly, and show that the innovation in Eqs. (3), (10), and (13) is a statistically equivalent random force which produces the same correlation function as the physical random force f in Eq. (1).

From Eqs. (3) and (8), the observable state variable can be expressed by the innovation series:

$$y(n) = \sum_{i=0}^{\infty} H\Phi^i K\gamma(n-i), \tag{18}$$

where $K = P(\infty)H^T\Gamma(\infty)^{-1}$. Equation (18) is also obtained from Eqs. (10) and (11); that is,

$$y(n) = \sum_{i=0}^{\infty} H'(\Phi')^i F'\gamma(n-i) = \sum_{i=0}^{\infty} H\Phi^i K\gamma(n-i).$$

Therefore, the correlation function matrix for the Kalman filter, the Markovian representation, or the AR-MA model is obtained as

$$\begin{aligned} &E\{y(k)y(l)^T\} \\ &\quad = E\left\{\sum_{i=0}^{\infty} H\Phi^i K\gamma(k-i)\left(\sum_{j=0}^{\infty} H\Phi^j K\gamma(l-j)\right)^T\right\} \\ &\quad = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} H\Phi^i K\Gamma(\infty)\delta_{k-i-l+j}K^T(\Phi^T)^j H^T. \end{aligned} \tag{19}$$

From the stationary condition, we have, for $k \leqslant l$,

$$\begin{aligned} C_{yy}(k-l) &:= E\{y(0)y(l-k)^T\} \\ &= \sum_{i=0}^{\infty} H\Phi^i K\Gamma(\infty)K^T(\Phi^T)^{i-k+l}H^T \\ &= H\left\{\sum_{i=0}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i\right\}(\Phi^T)^{l-k}H^T. \end{aligned} \tag{20}$$

To verify the identification of Eq. (17) with Eq. (20), we will find a following relation between the steady-state Riccati equation (5) and the generalized Einstein relation (15). By formal calculation of perturbation of steady-state equation (5), we have

$$\begin{aligned} P(\infty) &= V - \Phi K\Gamma(\infty)K^T\Phi^T + \Phi P(\infty)\Phi^T \\ &= \sum_{i=0}^{\infty}\Phi^i V(\Phi^T)^i - \sum_{i=1}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i. \end{aligned} \tag{21}$$

Substituting Eq. (16) into Eq. (21), we obtain the relation between $C_{xx}(0)$ and $P(\infty)$:

$$P(\infty) + \sum_{i=1}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i = C_{xx}(0). \tag{22}$$

From the definitions of $\Gamma(\infty)$ and K, we have the relations

$$\text{(a)}\quad HK = HP(\infty)H^T\Gamma(\infty)^{-1} = \Gamma(\infty)\Gamma(\infty)^{-1} = I, \tag{23}$$

$$\begin{aligned} \text{(b)}\quad HP(\infty) &= \Gamma(\infty)\Gamma(\infty)^{-1}HP(\infty) = \Gamma(\infty)K^T \\ &= HK\Gamma(\infty)K^T. \end{aligned} \tag{24}$$

Using Eqs. (22)–(24), we can prove the identification of correlation functions (17) and (20):

$$\begin{aligned} C_{yy}(k-l) &= H\left\{\sum_{i=1}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i\right\}(\Phi^T)^{l-k}H^T \\ &\quad + H\{K\Gamma(\infty)K^T\}(\Phi^T)^{l-k}H^T \\ &= H\left\{\sum_{i=1}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i\right\}(\Phi^T)^{l-k}H^T \\ &\quad + HP(\infty)(\Phi^T)^{l-k}H^T \\ &= HC_{xx}(0)(\Phi^T)^{l-k}H^T. \end{aligned} \tag{25}$$

In much the same manner, we can obtain the identification of Eq. (20) with Eq. (17) for $k \geqslant l$:

$$\begin{aligned} C_{yy}(k-l) &= E\{y(k-l)y(0)^T\} \\ &= H\Phi^{k-l}\sum_{i=0}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i H^T \\ &= H\Phi^{k-l}\sum_{i=1}^{\infty}\Phi^i K\Gamma(\infty)K^T(\Phi^T)^i H^T \\ &\quad + H\Phi^{k-l}P(\infty)H^T \\ &= H\Phi^{k-l}C_{xx}(0)H^T. \end{aligned} \tag{26}$$

It is concluded that both the random force f and the innovation γ give the same correlation functions of observable variables, though components of f are more than those of γ. And the number of components of γ is equal to that of measured variables. Therefore, the innovation γ is the statistically equivalent random force for the observable correlation functions.

IV. POWER SPECTRAL DENSITY AND EQUIVALENT RANDOM FORCE

In the physical and engineering fields, fluctuations in steady phenomena are usually expressed in the frequency domain rather than the time domain. In this section, we will examine the power spectral density in the frequency domain instead of the correlation function. It is well known that the power spectral density matrix is expressed by the Fourier transformation of the correlation function matrix according to the Weiner–Khintchin's theorem,[8] and is defined in the discrete time sampling by

$$P_{..}[\tilde{z}] = \sum_{n=-\infty}^{\infty} C_{..}(n)\tilde{z}^{-n}, \quad \tilde{z} = e^{i\omega\Delta t}. \tag{27}$$

Substituting Eq. (14) into Eq. (27), we have

$$\begin{aligned} &P_{xx}[\tilde{z}] \\ &\quad = C_{xx}(0)(I - \tilde{z}\Phi^T)^{-1} + (I - \tilde{z}^{-1}\Phi)^{-1}C_{xx}(0) - C_{xx}(0) \\ &\quad = (I - \tilde{z}^{-1}\Phi)^{-1}\{C_{xx}(0) - \Phi C_{xx}(0)\Phi^T\}(I - \tilde{z}\Phi^T)^{-1} \\ &\quad = (I - \tilde{z}^{-1}\Phi)^{-1}V(I - \tilde{z}\Phi^T)^{-1}. \end{aligned} \tag{28}$$

Here we have used the relation (15) and the property that Φ is the contraction mapping in a stable state. If we put

$$G_{xx}[\tilde{z}] = (I - \tilde{z}^{-1}\Phi)^{-1}, \tag{29}$$

then Eq. (28) is

$$P_{xx}[\tilde{z}] = G_{xx}[\tilde{z}]V(G_{xx}[\tilde{z}^{-1}])^T, \tag{30}$$

and G_{xx} is a kind of propagator of Eq. (1) in the frequency domain. On the other hand, by substitution of Eq. (17) into Eq. (27), we have

$$P_{yy}[\tilde{z}] = G_{yy}[\tilde{z}]V(G_{yy}[\tilde{z}^{-1}])^T, \tag{31}$$

where

$$G_{yy}[\tilde{z}] = H(I - \tilde{z}^{-1}\Phi)^{-1} = HG_{xx}[\tilde{z}], \tag{32}$$

which is called the transfer function in the control theory.

In this section we will express the equivalent representation of Eq. (31) in terms of the innovation. From the steady-state Riccati equation (5) and the identity

$$P(\infty) - \Phi P(\infty)\Phi^T$$
$$= (I - \tilde{z}^{-1}\Phi)P(\infty)(I - \tilde{z}\Phi^T)$$
$$+ \tilde{z}^{-1}\Phi P(\infty)(I - \tilde{z}\Phi^T) + (I - \tilde{z}^{-1}\Phi)P(\infty)\Phi^T\tilde{z},$$

we have the equivalent representation of the diffusion matrix V:

$$V = \Phi P(\infty)H^T\Gamma(\infty)^{-1}HP(\infty)\Phi^T$$
$$+ (I - \tilde{z}^{-1}\Phi)P(\infty)(I - \tilde{z}\Phi^T)$$
$$+ \tilde{z}^{-1}\Phi P(\infty)(I - \tilde{z}\Phi^T) + (I - \tilde{z}^{-1}\Phi)P(\infty)\Phi^T\tilde{z}. \quad (33)$$

Multiplying Eq. (33) by $H(I - \tilde{z}^{-1}\Phi)^{-1}$ and $(I - \tilde{z}\Phi^T)^{-1}H^T$ on both sides, we obtain another expression of Eq. (31):

$$P_{yy}[\tilde{z}] = H(I - \tilde{z}^{-1}\Phi)^{-1}V(I - \tilde{z}\Phi^T)^{-1}H^T$$
$$= H(I - \tilde{z}^{-1}\Phi)^{-1}\Phi P(\infty)H^T\Gamma(\infty)^{-1}HP(\infty)$$
$$\times\Phi^T(I - \tilde{z}\Phi^T)^{-1}H^T + HP(\infty)H^T$$
$$+ H(I - \tilde{z}^{-1}\Phi)^{-1}\tilde{z}^{-1}\Phi P(\infty)H^T$$
$$+ HP(\infty)\Phi^T\tilde{z}(I - \tilde{z}\Phi^T)^{-1}H^T. \quad (34)$$

If we use the relations of $\Gamma(\infty) = HP(\infty)H^T$ and $K = P(\infty)H^T\Gamma(\infty)^{-1}$, then Eq. (34) becomes

$$P_{yy}[\tilde{z}] = H(I - \tilde{z}^{-1}\Phi)^{-1}\Phi K\Gamma(\infty)$$
$$\times K^T\Phi^T(I - \tilde{z}\Phi^T)^{-1}H^T + \Gamma(\infty)$$
$$+ H(I - \tilde{z}^{-1}\Phi)^{-1}\Phi K\tilde{z}^{-1}\Gamma(\infty)$$
$$+ \tilde{z}\Gamma(\infty)K^T\Phi^T(I - \tilde{z}\Phi^T)^{-1}H^T$$
$$= G^K_{yy}[\tilde{z}]\Gamma(\infty)(G^K_{yy}[\tilde{z}^{-1}])^T, \quad (35)$$

where $G^K_{yy}[\tilde{z}] = I + H(I - \tilde{z}^{-1}\Phi)^{-1}\Phi K\tilde{z}^{-1}$. From Eq. (23), the factor $G^K_{yy}[\tilde{z}]$ of Eq. (35) is rewritten as

$$G^K_{yy}[\tilde{z}] = I + H\sum_{i=0}^{\infty}\tilde{z}^{-i}\Phi^i\Phi K\tilde{z}^{-1}$$
$$= I + H\sum_{i=1}^{\infty}\tilde{z}^{-i}\Phi^i K$$
$$= H\sum_{i=0}^{\infty}\tilde{z}^{-i}\Phi^i K$$
$$= H(I - \tilde{z}^{-1}\Phi)^{-1}K. \quad (36)$$

We can find that $G^K_{yy}[\tilde{z}]$ is a propagator of Eqs. (3) and (8) in the frequency domain and that the propagator (36) can be derived from Eq. (18). It can be concluded that Eq. (35) is the equivalent representation of Eq. (31) by using the equivalent random force.

V. CONCLUDING REMARKS

(1) Any combination of the Kronecker index $\{\sigma_i\}$ of transformation matrix T, with the condition $\Sigma_{i=1}^{q}\sigma_i = d$, can determine an AR-MA model as in Sec. II. Therefore, we must examine the uniqueness of AR-MA model related to the physical state equation. Let T and T'' be the transformation matrices corresponding to the combinations of $\{\sigma_i\}$ and $\{\sigma_i''\}$, respectively, then we have the transformations of the conditional state variable

$$x'(n|n) = Tx(n|n) \quad \text{or} \quad x(n|n) = Sx'(n|n),$$

and

$$x''(n|n) = T''x(n|n) \quad \text{or} \quad x(n|n) = S''x''(n|n),$$

where $ST = I$ and $S''T'' = I$. Setting the transformation matrix $Q(= TS'')$; $x'(n|n) = Qx''(n|n)$, we can obtain another Markovian representation of $x''(n|n)$ from the Markovian representation (10) of $x'(n|n)$.

The transformation matrix Q in the Markovian representations may correspond to a unimodular matrix in the AR-MA models. As in the same way in Sec. II, we have AR-MA models according to the combinations of $\{\sigma_i\}$ and $\{\sigma_i''\}$, respectively:

$$A(z)\ y(n) = B(z)\ \gamma(n),$$
$$A''(z)\ y(n) = B''(z)\ \gamma(n),$$

where z is the time shift operator in Sec. II. The coefficient matrices of AR-MA models are transformed into each other by a unimodular matrix. That is,

$$U(z)A(z) = A''(z),$$
$$U(z)B(z) = B''(z),$$

where the unimodular matrix $U(z)$ has the polynomial elements of z, and $\det U(z) = \text{constant}$. The uniqueness of AR-MA model derived from the state equation is not guaranteed except in special cases.

In contrast to AR-MA models, the correlation function matrix $C_{yy}(n)$ and the power spectral density $P_{yy}[\tilde{z}]$ are invariant with the choice of the Kronecker indices, since $C_{yy}(n)$ and $P_{yy}[\tilde{z}]$ in the Kalman filter (3), the Markovian representation (10), and the AR-MA model (13) are the same as in Secs. III and IV. Therefore, an additional physical condition is needed for the correspondence of an AR-MA model fitting to the time-series data to a physical AR-MA model uniquely.

(2) As mentioned in Sec. III, the number of components of the equivalent random force is equal to that of observable variables. Inversely, another AR-MA model may be determined for various combinations of observable variables. Therefore, we can have various types of AR-MA models for the same physical process according to measurements. In macroscopic plants, a time-series model is often used for the diagnosis of plant operations or the determination of physical processes. In this situation, it is important to distinguish a physical random force with an equivalent random force.

(3) The power spectral density of Eq. (3) may be defined by

$$P^K_{xx}[\tilde{z}] = (I - \tilde{z}^{-1}\Phi)^{-1}K\Gamma(\infty)K^T(I - \tilde{z}\Phi^T)^{-1}.$$

The rank of $P^K_{xx}[\tilde{z}]$ is equal to that of $P_{yy}[\tilde{z}]$ and is less than that of $P_{xx}[\tilde{z}]$. This is because $P_{yy}[\tilde{z}] = HP_{xx}[\tilde{z}]H^T = HP^K_{xx}[\tilde{z}]H^T$. On the other hand, the propagator $G_{yy}[\tilde{z}]$, Eq. (32), is not equal to $G^K_{yy}[\tilde{z}]$, Eq. (36).

(4) The equivalent representation of power spectral density proved in Sec. IV is related to the spectral factorization in the control theory which is reviewed by Kailath[10] and Brockett.[11] Hence, the operations and calculations in Sec. IV are similar to the spectral factorization problem in the control theory. However, the observation equation (2) has no

random force for measurements, so that the power spectral density of observable variables is inevitably factorized as in Sec. IV.

(5) An AR (autoregressive) model, which is

$$\sum_{i=0}^{M} A_i\, y(n-i) = \gamma(n) \quad (A_0 = I)\,,$$

is often used for practical cases, since its simple and fast recursion algorithms can be applied.[12] In an algorithm, the AR model is fitted to correlation functions obtained from physical systems. Hence, the uniqueness problem of (1) does not remain in the AR model fitting procedure. An AR model is identified with an equivalence class of AR-MA model derived from the physical equation.

ACKNOWLEDGMENTS

The author would like to thank Dr. S. Yamada and Mr. K. Bekki, Osaka University, for the essential comments of Eqs. (33)–(36). This work was partially supported by the Grant-in-Aid for Scientific Research (No. 58780203) from the Ministry of Education, Science, and Culture.

[1] K. Kishida, Phys. Rev. A **25**, 496 (1982).
[2] The term of "physical" means that a stochastic process is described by a master equation with given transition probabilities of physical system as mentioned in Ref. 1. This implies that regression and diffusion coefficient matrices are given in a Langevin equation or state equation (1).
[3] N. Suda and S. Shirai, J. Nucl. Sci. Technol. **8**, 438 (1971); W. L. Miller and J. B. Lewis, IEEE Trans. Automatic Control **16**, 841 (1971); H. Kokame and Y. Nishikawa, Nucl. Sci. Eng. **67**, 8 (1978).
[4] K. Fukunishi, Nucl. Sci. Eng. **62**, 215 (1977); S. Haykin, *Nonlinear Methods of Spectral Analysis*, Topics in Applied Physics, Vol. 34 (Springer-Verlag, Berlin, 1979).
[5] N. G. van Kampen, Can. J. Phys. **39**, 551 (1961); R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973).
[6] K. Tomita and H. Tomita, Prog. Theor. Phys. **51**, 1731 (1974); K. Kishida, S. Kanemoto, and T. Sekiya, J. Nucl. Sci. Technol. **13**, 19 (1976).
[7] R. Kubo, *Many Body Theory* (Benjamin, New York; Shokabo, Tokyo, 1966), pp. 1–16.
[8] R. Kubo, *Statistical Physics* (in Japanese), edited by M. Toda and R. Kubo (Iwanami, Tokyo, 1972), Secs. 5 and 6, to be published as *Statistical Physics II:Non-Equilibrium Statistical Mechanics* (Springer-Verlag, Berlin, 1984).
[9] D. G. Luenberger, IEEE Trans. Autom. Control **12**, 290 (1967).
[10] T. Kailath, IEEE Trans. Inf. Theory **20**, 146 (1974).
[11] R. W. Brockett, *Finite Dimensional Linear Systems* (Wiley, New York, 1970).
[12] N. Levinson, J. Math. Phys. **25**, 261 (1947); J. Durbin, Rev. Inst. Int. Statist. **28**, 233 (1960).

Maupertuis' principle of least action in stochastic calculus of variations

J.-C. Zambrini[a),b)]
Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland

(Received 1 February 1983; accepted for publication 8 April 1983)

Within the framework of stochastic calculus of variations for time-symmetric semimartingales $X(t,\omega)$, we consider two different stochastic versions of Maupertuis' least action principle, in Lagrangian and Hamiltonian terms. The general results are applied to classical statistical mechanics, where they coincide with those of classical calculus of variations, and to Nelson's stochastic mechanics, an approach to quantum mechanics where a time-symmetric semimartingale represents the position of a particle and the dynamics is expressed by a stochastic version of Hamilton's principle of least action. Some historic examples of old quantum theory are discussed.

PACS numbers: 02.50.Ey, 03.65. – w

INTRODUCTION

Motivated by the growing number of discoveries of interesting natural phenomena which seem to escape from the usual deterministic descriptions, we examine systematically in the present paper probabilistic extensions of variational principles used for the determination of possible dynamics in classical physics. A general framework was created recently by Yasue in this goal,[1,2] and we choose the same point of view.

The central object of our investigations will be Maupertuis' least action principle. It has the advantage of involving not only equations of motion but also the energy conservation law, one of the most important conservation laws in physics.

The first section will be devoted to a brief presentation of the class of stochastic processes taken into account in these variational principles. Essentially, it was discovered by Nelson in 1966,[3] but can be presented today in a general frame of semimartingale stochastic integrals,[4–7] a large extension of classical (Itô's) theory of stochastic integrals.

In Sec. 2 we will recall central results of stochastic calculus of variations, especially a stochastic version of Hamilton's principle of least action. This will be obtained by means of a variation which does not modify the time parameter in the action. For Maupertuis' principle, we need another variation so that the parameter has a variational status equivalent to the state variables. We introduce in this aim a noncontemporaneous variation and show that Hamilton's principle is equivalent to Maupertuis' one for a stochastic generalization of conservation systems.

Section 4 deals with the connection between variational principles and conservation of energy (defined in terms of the initially given Lagrangian).

The first application concerns classical statistical mechanics (Sec. 5), where the dynamics is given by an ordinary differential equation with random initial conditions. In this limiting case we obtain again principles of classical calculus of variations.

The best application will be to Nelson's stochastic mechanics, a realization of quantum mechanics where the notion of paths for particles is preserved (Sec. 6). We verify, for example, that Planck's recipe of quantization, in old quantum theory, still makes certain sense if we use the so-called stochastic quantization procedure.

1. A CLASS OF STOCHASTIC PROCESSES

Let $(\Omega, \mathcal{A}, P)$ be a base probability space and a stochastic process x in $\mathbb{R}^l$ some continuous application $t \to x_t \equiv x(t)$ from a time interval I into the Hilbert space $H = L^2((\Omega,P);\mathbb{R}^l)$. We consider two filtrations indexed by I, $\mathfrak{P}_t$ and $\mathfrak{F}_t$ with $\mathfrak{P}_s \subset \mathfrak{P}_t$ and $\mathfrak{F}_s \supset \mathfrak{F}_t$ for $s \leqslant t$ to which x is adapted.

By hypothesis, x is simultaneously a $\mathfrak{P}_t$ semimartingale and an $\mathfrak{F}_t$ semimartingale, in other words, it admits two Meyer's canonical decompositions[5]

$$x_t = x_0 + B_t + M_t, \tag{1.1}$$

where B_t is a process of bounded variation adapted to $\mathfrak{P}_t$, with $B_0 = 0$, and M_t a local $\mathfrak{P}_t$-martingale with $M_0 = 0$, and

$$x_t = x_0 + B_{*t} + M_{*t}, \tag{1.2}$$

where B_{*t} is a process of bounded variation adapted to $\mathfrak{F}_t$, with $B_{*0} = 0$, and M_{*t} a local $\mathfrak{F}_t$-martingale with $M_{*0} = 0$. Moreover, the process x_t will have the two mean velocities corresponding to different information available at time t,

$$Dx_t = \lim_{\Delta t \downarrow 0} E\left[\frac{x(t+\Delta t) - x(t)}{\Delta t}\middle| \mathfrak{P}_t\right] \tag{1.3}$$

and

$$D_* x_t = \lim_{\Delta t \downarrow 0} E\left[\frac{x(t) - x(t-\Delta t)}{\Delta t}\middle| \mathfrak{F}_t\right], \tag{1.4}$$

where $E[\cdots|\sigma]$ is a σ-conditional expectation.

These two limits exist in H and the mappings $t \to Dx_t$, $t \to D_* x_t$ are continuous from I into H by hypothesis. Since this class of stochastic processes was discovered by Nelson,[3,7,8] we will call them simply Nelson processes hereafter.

The following integration by parts formula will be used, for two Nelson processes X and Y,

$$E\left[X(t)Y(t)\Big|_{t_a}^{t_b}\right] = E\left[\int_{t_a}^{t_b} \{DX(t)\cdot Y(t) + X(t) D_* Y(t)\}\, dt\right], \tag{1.5}$$

[a)] Partially supported by the Swiss National Science Foundation.

[b)] Present address: Department of Mathematics, Princeton University, Princeton, NJ 08544.

 0022-2488/84/051314-09$02.50

where $E[\cdots]$ is the absolute expectation.[3]

More explicit descriptions of these processes and of their properties can be found in Refs. 6 and 9. For a really rigorous account on this class of time symmetric semimartingales, consult Refs. 7 and 10.

Let $L \in C^2(\mathbb{R}^{3l};\mathbb{R})$ be a given function and $X(t)$ a Nelson process on $I \supset [t_a,t_b]$. If the process $L(X(t),DX(t),D_*X(t))$ is integrable, we define the action for Lagrangian L by a real functional J,

$$J: x \mapsto E\left[\int_{t_a}^{t_b} L(X(t),DX(t),D_*X(t))\,dt\right]. \tag{1.6}$$

We suppose further that the components of the $\mathbb{R}^{3l}$-gradient $(\partial_1 L,\partial_2 L,\partial_3 L)(X(t),DX(t),D_*X(t))$ in the three variables of the Lagrangian are themselves Nelson process. All the Lagrangians used for different types of actions in this paper will satisfy these hypothesis.

By $D_{x_a}^{x_b}$ we will denote the totality of Nelson processes with fixed end points $X(t_a)=X_a \in H$, $X(t_b)=X_b \in H$, and by Δ that of Nelson processes $Z(t)$ such that $Z(t_a)=Z(t_b)=0$.

2. MAUPERTUIS' PRINCIPLE VIA HAMILTON'S PRINCIPLE

Our starting point will be the stochastic version of Hamilton's principle of least action.[1,2]

Since $L(x,y,z)$ is twice continuously differentiable, we can use its Taylor expansion and then employ it at the processes $X(t),DX(t),D_*X(t)$ so that the variation of the action functional J [given by Eq. (1.6)] in X on a Nelson process δX is

$$\delta J[X](\delta x) = E\left[\int_{t_a}^{t_b}\left(\frac{\partial L}{\partial DX}\delta DX + \frac{\partial L}{\partial D_*X}\delta D_*X + \frac{\partial L}{\partial X}\delta X\right)dt\right]. \tag{2.1}$$

A process $X=X(t)$ in $D_{x_a}^{x_b}$ is called a stationary point of J, or an extremal, if

$$\delta J[X](\delta X)=0 \tag{2.2}$$

for all the processes δX in Δ.

Theorem: Stochastic Hamilton's principle[1,2]: A necessary and sufficient condition for $X(t)\in D_{x_a}^{x_b}$ to be a stationary point of the action functional J is that, on $X(t)$,

$$D\frac{\partial L}{\partial D_*X(t)} + D_*\frac{\partial L}{\partial DX(t)} + \frac{\partial L}{\partial X(t)} = 0. \tag{2.3}$$

Equation (2.3) (the YEN equation) follows from this stochastic least action principle as the Euler–Lagrange equation follows from the classical Hamilton's principle.[9,11]

In order to interpret Eq. (2.3) as a stochastic generalization of the Euler–Lagrange equation, it is natural to introduce the following definitions.

The configuration of a stochastic dynamical system at time t is described by a point in the Euclidean-l space with coordinates $\{X_t^1,\ldots,X_t^l\}$.

This space is called the configuration space. A trajectory or a path is the continuous curve in the configuration space traced by this point as time changes. The state of the system at t is given by a point in the $3l$-space with coordinates

$$\{X_t^1,\ldots,X_t^l;\ DX_t^1,\ldots,DX_t^l;\ D_*X_t^1,\ldots,D_*X_t^l\}.$$

From the physical point of view, the admissible nonextremal processes used in Hamilton's principle are virtual: They give paths along which the system may be imagined to move without satisfying necessarily the law of motion (2.3).

Let us observe that Hamilton's principle remains unchanged if the Lagrangian depends explicitly on time t, namely for nonconservative systems. In the following, if not otherwise stated, we consider only conservative systems or, more precisely, conservative and holonomic systems.[9]

Then, by Eq. (2.1), Hamilton's principle can be put in the form

$$E\left[\int_{t_a}^{t_b}\delta L\,dt\right]=0, \tag{2.4}$$

where the variation of L is defined formally by the expression

$$\delta L = \frac{\partial L}{\partial X}\delta X + \frac{\partial L}{\partial DX}\delta DX + \frac{\partial L}{\partial D_*X}\delta D_*X. \tag{2.5}$$

Clearly, in such a variation, time t is not altered. However, even in classical mechanics, it is often necessary to consider time t as an auxiliary state variable. Let us introduce a new parameter u and u-dependent time $t:[u_a,u_b]\to\mathbb{R}$ which is a (deterministic) differentiable function such that

$$\frac{dt}{du}\equiv\phi>0. \tag{2.6}$$

Define a bijection $T_\phi: H\to H$ by

$$X(u)\mapsto X(t(u))\equiv\overline{X}(u). \tag{2.7}$$

In using the definitions of the velocities (1.3) and (1.4), we have simply, if the subscript u denotes mean derivatives with respect to u,

$$D_u\overline{X}=\phi\cdot DX, \tag{2.8}$$

$$D_{*u}\overline{X}=\phi\cdot D_*X. \tag{2.9}$$

In this way, we can define a new variation for a general Lagrangian $L\in C^2(\mathbb{R}^{3l}\times\mathbb{R};\mathbb{R})$ by

$$\delta_t L = \frac{\partial L}{\partial X}\delta_t X + \frac{\partial L}{\partial DX}\delta_t\left(\frac{D_u\overline{X}}{\phi}\right) + \frac{\partial L}{\partial D_*X}\delta_t\left(\frac{D_{*u}\overline{X}}{\phi}\right) + \frac{\partial L}{\partial t}\delta_t t, \tag{2.10}$$

with

$$\delta_t\left(\frac{D_u\overline{X}}{\phi}\right) = \frac{\delta_t D_u\overline{X}}{\phi} - \frac{D_u\overline{X}}{\phi}\frac{\delta_t(dt/du)}{\phi} = \frac{D_u\delta_t\overline{X}}{\phi} - DX\frac{d}{dt}(\delta_t t),$$

where we have utilized Eq. (2.8) and the commutation of derivatives and variations. By construction of the x-variation, we have $\delta_t\overline{X}=\delta X$ and so

$$\delta_t\left(\frac{D_u\overline{X}}{\phi}\right) = D\delta X - DX\frac{d}{dt}(\delta_t t). \tag{2.11}$$

In the same way we find

$$\delta_t\left(\frac{D_{*u}\overline{X}}{\phi}\right) = D_*\delta X - D_*X\frac{d}{dt}(\delta_t t). \tag{2.12}$$

After the substitution of Eqs. (2.11) and (2.12) into Eq. (2.10), and by the definition (2.5) of the δ-variation, we conclude that

$$\delta_t L = \delta L - \left(DX\frac{\partial L}{\partial DX} + D_*X\frac{\partial L}{\partial D_*X}\right) \times \frac{d}{dt}(\delta_t t) + \frac{\partial L}{\partial t}\delta_t t. \tag{2.13}$$

As is expected, these two variations coincide if t is not varied. We shall call this new variation δ_t in which time is varied as well as the state variables noncontemporaneous variation.

Now we introduce a particular class of Lagrangians, inspired by the usual situation in classical mechanics.[9] They are of the form

$$L(X,DX,D_*X) = T(X,DX,D_*X) - V(X), \tag{2.14}$$

where $T \in C^2(\mathbb{R}^{3l};\mathbb{R})$ is homogeneous of degree 2 in the velocities DX and D_*X are called the kinetic energy, and $V(X) \in C^2(\mathbb{R}^l;\mathbb{R})$ is the potential energy.

On the other hand, it is possible to verify, modulo some conditions on the form of the given Lagrangian L, that YEN equation (2.3) is equivalent to the following system of "stochastic Hamilton's equations"[9]:

$$\frac{1}{2}DX = \frac{\partial H}{\partial p}, \tag{2.15}$$

$$\frac{1}{2}D_*X = \frac{\partial H}{\partial p_*}, \tag{2.16}$$

$$\frac{1}{2}(Dp_* + D_*p) = -\frac{\partial H}{\partial X}, \tag{2.17}$$

where the generalized momenta p and p_* are defined by

$$\frac{1}{2}p = \frac{\partial L}{\partial DX}, \tag{2.18}$$

$$\frac{1}{2}p_* = \frac{\partial L}{\partial D_*X} \tag{2.19}$$

and the Hamiltonian H by

$$H = \tfrac{1}{2}pDX + \tfrac{1}{2}p_*D_*X - L. \tag{2.20}$$

In using Eq. (2.20) and Euler's theorem for homogeneous functions, the Hamiltonian H corresponding to the Lagrangian (2.14) and expressed in the variables (X,DX,D_*X) takes a compact form,

$$\mathscr{E}(X,DX,D_*X) = T(X,DX,D_*X) + V(X). \tag{2.21}$$

It will be called the energy function.

Now, we can prove that for this class of Lagrangians, Hamilton's principle implies the stochastic version of Maupertuis' principle. More precisely:

Theorem 1: For the Nelson processes γ: $t \to X(t)$ in $D_{x_b}^{x_a}$ such that

$$E[\mathscr{E}(X,DX,D_*X)] = h \quad \text{(a constant)} \tag{2.22}$$

where $\mathscr{E}$ is the energy function (2.21),

$$\delta E\left[\int_{t_a}^{t_b} L(X,DX,D_*X)\,dt\right] = 0 \Rightarrow \delta_t E\left[\int_\gamma 2T(X,DX,D_*X)\,dt\right] = 0.$$

In Maupertuis' principle, the transit times are varied for different processes since we use noncontemporaneous variations. Sometimes one speaks of "reduced action" for this new principle.

Proof: By Eq. (2.13) for the kinetic energy T,

$$\delta_t T = \delta T - \left(DX\frac{\partial T}{\partial DX} + D_*X\frac{\partial T}{\partial D_*X}\right)\frac{d}{dt}(\delta_t t),$$

that is, by homogeneity,

$$\delta_t T = \delta T - 2T\frac{d}{dt}(\delta_t t). \tag{2.23}$$

According to Eqs. (2.4) and (2.14), Hamilton's principle is

$$0 = E\left[\int_{t_a}^{t_b}(\delta T - \delta V)\,dt\right].$$

Since V depends only on the position, $\delta V = \delta_t V$, and, by Eq. (2.23),

$$0 = E\left[\int_{t_a}^{t_b}\left\{\delta_t T + 2T\frac{d}{dt}(\delta_t t) - \delta_t V\right\}dt\right].$$

It follows from the constraint (2.22) that $E[\delta_t T] = -E[\delta_t V]$ and then

$$0 = E\left[\int_{t_a}^{t_b} 2\{\delta_t T\,dt + T d(\delta_t t)\}\right]. \tag{2.24}$$

Now, by Eq. (2.6), $dt = \phi\,du$, and then

$$d(\delta_t t) = \frac{d}{du}(\delta_t t)du = \delta_t\phi \cdot du; \tag{2.25}$$

we get for Eq. (2.24)

$$0 = E\left[\int_0^1 2\{\delta_t T \cdot \phi + T\delta_t\phi\}\,du\right],$$

where we have fixed $u(t_b) = 1$, $u(t_a) = 0$. Finally,

$$0 = E\left[\int_0^1 2\delta_t(T\phi)\,du\right] = \delta_t E\left[\int_{t_a}^{t_b} 2T\,dt\right]. \quad \blacksquare$$

For remembering the energy constraint (2.22) involved in Maupertuis' principle, we will denote it by

$$\delta_{th} E\left[\int_\gamma 2T(X,DX,D_*X)\,dt\right] = 0. \tag{2.26}$$

Taking into account the above-mentioned Hamiltonian framework, Maupertuis' principle can also be expressed in the following way:

Theorem 2: With the same conditions as in Theorem 1, we have

$$\delta_{th}\frac{1}{2}E\left[{}^{\mathfrak{P}}\!\int p \cdot dX + {}^{\mathfrak{F}}\!\int p_* \cdot dX\right] = 0. \tag{2.27}$$

Here ${}^{\mathfrak{P}}$ and ${}^{\mathfrak{F}}$ denote, respectively, the $\mathfrak{P}_t$ and the $\mathfrak{F}_t$ integrals of Itô[4,9] for semimartingales, an extension of the classical notion of stochastic integrals (also due to Itô).

Proof: Since T is homogeneous,

$$2T = \frac{\partial T}{\partial DX}DX + \frac{\partial T}{\partial D_*X}D_*X.$$

Now, by Eq. (2.14) and Eqs. (2.18) and (2.19),

Maupertuis' principle (2.26) changes to

$$\delta_{th}\frac{1}{2}E\left[\int_\gamma p\,DX\,dt+\int_\gamma p_*\,D_*X\,dt\right]=0. \tag{2.28}$$

On the other hand, for X and Y two Nelson's processes we have, by continuity of Y,

$$Y(t)=\lim_{|\Delta|\to 0}\sum_{i=0}^{n-1}Y(t_i)\cdot 1_{[t_i,t_{i+1}]}(t),\quad t_a\leqslant t\leqslant t_b,$$

(here $1_{[\,,\,]}$ denotes a characteristic function) and[4]

$$\mathfrak{P}\int_{t_a}^{t_b}Y\cdot dX=\text{l.i.m.}\sum_{i=0}^{n-1}Y(t_i)\{X(t_{i+1})-X(t_i)\}, \tag{2.29}$$

where l.i.m. is the limit in the mean square, $|\Delta|=\max(t_{i+1}-t_i)$, and $t_0=t_a<t_1<\cdots<t_n=t_b$ a partition of $[t_a,t_b]$. But we know also that

$$E\left[\int_{t_a}^{t_b}Y(t)\cdot DX(t)\,dt\right]=\lim_{|\Delta|\to 0}E\left[\sum_{i=0}^{n-1}Y(t_i)DX(t_i)(t_{i+1}-t_i)\right]=\lim_{|\Delta|\to 0}E\left[\sum_{i=0}^{n-1}Y(t_i)\{X(t_{i+1})-X(t_i)\}\right].$$

Then

$$E\left[\mathfrak{P}\int_{t_a}^{t_b}Y\cdot dX\right]=E\left[\int_{t_a}^{t_b}Y(t)\,DX(t)\,dt\right]. \tag{2.30}$$

In the same way, one obtains the relation

$$E\left[\mathfrak{F}\int_{t_b}^{t_a}Y\cdot dX\right]=E\left[\int_{t_a}^{t_b}Y(t)\,D_*X(t)\,dt\right]. \tag{2.31}$$

For $Y\equiv p$ in Eq. (2.30), then $Y\equiv p_*$ in Eq. (2.31), the addition reduces Eq. (2.28) to the form (2.27). ■

We may also observe that, from the general relations between the $\mathfrak{P}_t$, $\mathfrak{F}_t$ and symmetric stochastic integrals,[4,6,9]

$$\int_{t_a}^{t_b}d(X\cdot Y)=\int_{t_a}^{t_b}Y\circ dX+\int_{t_a}^{t_b}X\circ dY \tag{2.32}$$

$$=\mathfrak{P}\int_{t_a}^{t_b}Y\cdot dX+\mathfrak{F}\int_{t_a}^{t_b}X\cdot dY, \tag{2.33}$$

where $\circ$ denotes the Fisk–Stratonovich symmetric integral, and Eqs. (2.30) and (2.31), we get

$$E\left[X(t)Y(t)\big|_{t_a}^{t_b}\right]=E\left[\int_{t_a}^{t_b}\{Y(t)\,DX(t)+X(t)\,D_*Y(t)\}\,dt\right], \tag{2.34}$$

namely, the integration by parts formula (1.5). Averaging Eq. (2.34) with the formula where X and Y are interchanged yields

$$E\left[X(t)Y(t)\big|_{t_a}^{t_b}\right]=E\left[\int_{t_a}^{t_b}\{Y(t)\cdot\tfrac{1}{2}(DX(t)+D_*X(t))+X(t)\cdot\tfrac{1}{2}(DY(t)+D_*Y(t))\}\,dt\right] \tag{2.35}$$

$$=E\left[\int_{t_a}^{t_b}Y\circ dX+X\circ dY\right]. \tag{2.36}$$

Now, we will give some emphasis to the reciprocal relationship between Maupertuis' principle and Hamilton's principle for conservative systems.

3. HAMILTON'S PRINCIPLE VIA MAUPERTUIS' PRINCIPLE

We will assume the validity of Maupertuis' principle, that is,

$$\delta_{th}E\left[\int_\gamma 2T(X,DX,D_*X)\,dt\right]=0 \tag{3.1}$$

and

$$\delta_t E\left[T(X,DX,D_*X)+V(X)\right]=0. \tag{3.2}$$

The classical method of Lagrange multipliers[9,12] suggests consideration of

$$\delta_t E\left[\int_{t_a}^{t_b}\mathscr{L}(X,DX,D_*X)\,dt\right]=0 \tag{3.3}$$

for the new Lagrangian

$$\mathscr{L}=2T+\lambda\,[T+V-h\,], \tag{3.4}$$

where h is the constant used in Eq. (2.22) and λ some unknown multiplier. By means of the time change $t=t(u)$ defined in the preceding section, Eq. (3.3) modifies to [in using Eqs. (2.8), (2.9), and the convention of Theorem 1 for the parameter u]

$$0=\delta_t E\left[\int_0^1\mathscr{L}\left(\overline{X},\frac{D_u\overline{X}}{\phi},\frac{D_{*u}\overline{X}}{\phi}\right)\phi\,du\right] \tag{3.3'}$$

$$=E\left[\int_0^1\delta_t\mathscr{L}\cdot\phi\,du\right]+E\left[\int_0^1\mathscr{L}\delta_t\phi\,du\right]. \tag{3.5}$$

It follows from

$$\delta_t\mathscr{L}(X,DX,D_*X)=\delta_t\mathscr{L}\left(\overline{X},\frac{D_u\overline{X}}{\phi},\frac{D_{*u}\overline{X}}{\phi}\right)$$

that

$$\delta_t\mathscr{L}=\frac{\partial\mathscr{L}}{\partial X}\delta_tX+\frac{\partial\mathscr{L}}{\partial DX}\frac{\delta_tD_u\overline{X}}{\phi}-\frac{\partial\mathscr{L}}{\partial DX}\frac{D_u\overline{X}}{\phi^2}\delta_t\phi+\frac{\partial\mathscr{L}}{\partial D_*X}\frac{\delta_tD_{*u}\overline{X}}{\phi}-\frac{\partial\mathscr{L}}{\partial D_*X}\frac{D_{*u}\overline{X}}{\phi^2}\delta_t\phi$$

and then the first term of the right-hand member of Eq. (3.5) becomes

$$E\left[\int_0^1\left\{\frac{\partial\mathscr{L}}{\partial X}\delta_tX\cdot\phi+\frac{\partial\mathscr{L}}{\partial DX}D_u(\delta_tX)-\frac{\partial\mathscr{L}}{\partial DX}DX\cdot\delta_t\phi+\frac{\partial\mathscr{L}}{\partial D_*X}D_{*u}(\delta_tX)-\frac{\partial\mathscr{L}}{\partial D_*X}D_*X\cdot\delta_t\phi\right\}\cdot du\right].$$

After integration by parts of the second and the fourth terms, we get

$$E\left[\int_{t_a}^{t_b}\left(\frac{\partial\mathscr{L}}{\partial X}-D_*\frac{\partial\mathscr{L}}{\partial DX}-D\frac{\partial\mathscr{L}}{\partial D_*X}\right)\delta_t X\,dt\right]+E\left[\left(\frac{\partial\mathscr{L}}{\partial DX}+\frac{\partial\mathscr{L}}{\partial D_*X}\right)\delta_t X\Big|_{t_a}^{t_b}\right]-E\left[\int_0^1\left(\frac{\partial\mathscr{L}}{\partial DX}DX+\frac{\partial\mathscr{L}}{\partial D_*X}D_*X\right)\delta_t\phi\,du\right].$$

In this way Eq. (3.5) reduces to

$$0=E\left[\int_{t_a}^{t_b}\left(\frac{\partial\mathscr{L}}{\partial X}-D_*\frac{\partial\mathscr{L}}{\partial DX}-D\frac{\partial\mathscr{L}}{\partial D_*X}\right)\delta_t X\,dt\right]+E\left[\left(\frac{\partial\mathscr{L}}{\partial DX}+\frac{\partial\mathscr{L}}{\partial D_*X}\right)\delta_t X\Big|_{t_a}^{t_b}\right]-E\left[\int_0^1\left(\frac{\partial\mathscr{L}}{\partial DX}DX+\frac{\partial\mathscr{L}}{\partial D_*X}D_*X-\mathscr{L}\right)\delta_t\phi\,du\right].$$

Now the $\delta_t X\equiv\delta X$ are arbitrary in Δ, and $\delta_t\phi$ is also arbitrary, which implies that

$$\frac{\partial\mathscr{L}}{\partial X}-D_*\frac{\partial\mathscr{L}}{\partial DX}-D\frac{\partial\mathscr{L}}{\partial D_*X}=0,\quad t_a\leqslant t\leqslant t_b,\tag{3.6}$$

$$E\left[\frac{\partial\mathscr{L}}{\partial DX}DX+\frac{\partial\mathscr{L}}{\partial D_*X}D_*X-\mathscr{L}\right]=0,\quad 0\leqslant u\leqslant 1.\tag{3.7}$$

Substituting for $\mathscr{L}$ from Eq. (3.4) in Eq. (3.7) and taking into account the homogeneity of T yields

$$E[\lambda(T+V-h)]-E[2T(1+\lambda)]=0.$$

This equation for λ has a simple solution. Indeed,

$$\lambda E[T+V-h]-(1+\lambda)E[2T]=0$$

is satisfied by $\lambda=-1$ since $E[T+V-h]$ vanishes by hypothesis.

One verifies immediately that for this value of λ Eq. (3.6) simplifies to

$$D\frac{\partial T}{\partial D_*X}+D_*\frac{\partial T}{\partial DX}-\frac{\partial T}{\partial X}+\frac{\partial V}{\partial X}=0,$$

that is, to the YEN equation (2.3) for the Lagrangian

$$L(X,DX,D_*X)=T(X,DX,D_*X)-V(X).$$

Then we proved

Theorem 3: For the Nelson processes $\gamma: t\to X(t)$ in $D_{x_a}^{x_b}$ such that Maupertuis' principle holds in the form (2.26), that is, with a Lagrangian L of the type (2.14), Hamilton's principle is also satisfied:

$$\delta E\left[\int_{t_a}^{t_b}L(X,DX,D_*X)\,dt\right]=0.$$

4. VARIATIONAL PRINCIPLES AND CONSERVATION OF ENERGY

In this section we consider the connection between the different variational principles and the conservation of the energy function (2.20), $\mathscr{E}=(\partial L/\partial DX)\,DX+(\partial L/\partial D_*X)\,D_*X-L$.

For this purpose, it will be interesting to compare the sense of the noncontemporaneous variation δ_t with the one of the δ variation.

Since we know already the sense of $\delta E\left[\int_{t_a}^{t_b}L(X,DX,D_*X)\,dt\right]$ thanks to Hamilton's principle, let us examine the noncontemporaneous variation of this action. As for Theorem 3, by Eq. (2.25) we have

$$\delta_t E\left[\int_{t_a}^{t_b}L\,dt\right]=E\left[\int_0^1\delta_t(L\phi)\,du\right]=E\left[\int_{t_a}^{t_b}\delta_t L\cdot dt\right]+E\left[\int_0^1 L\,d(\delta_t t)\right].\tag{4.1}$$

Taking into account the definition (2.13) of δ_t, the homogeneity of T in $L(X,DX,D_*X)=T(X,DX,D_*X)-V(X)$, and the relation $\mathscr{E}(X,DX,D_*X)=2T(X,DX,D_*X)-L(X,DX,D_*X)$, we find

$$E\left[\int_{t_a}^{t_b}\delta L\,dt-\int_0^1\mathscr{E}\,d(\delta_t t)\right].\tag{4.2}$$

Now, by Eqs. (2.35) and (2.36),

$$E\left[\int_0^1 d(\mathscr{E}\cdot\delta_t t)\right]=E\left[\int_0^1\mathscr{E}\circ d(\delta_t t)+\delta_t t\circ d\mathscr{E}\right]=E\left[\int_0^1\mathscr{E}\cdot d(\delta_t t)+\frac{1}{2}(D_u\mathscr{E}+D_{*u}\mathscr{E})\delta_t t\cdot du\right]\tag{4.3}$$

since $\delta_t t$ is of bounded variations.[13] Let us note that the integrand in the second term in the expectation is nothing else than the part of the variation of energy due to the time variation between the real and virtual processes, namely,

$$\tfrac{1}{2}(D_u\mathscr{E}+D_{*u}\mathscr{E})\cdot\delta_t t\equiv\Delta_t\mathscr{E}.\tag{4.4}$$

By means of Eq. (4.3), it may be seen that Eq. (4.2) becomes

$$E\left[\int_{t_a}^{t_b}\delta L\,dt-\int_0^1 d(\mathscr{E}\,\delta_t t)+\int_0^1\Delta_t\mathscr{E}\cdot du\right].$$

Hence, using Eq. (2.5) and integrations by part, we get finally:

Theorem 5: For conservative stochastic systems with Lagrangian (2.14), $L(X,DX,D_*X)=T(X,DX,D_*X)-V(X)$, the noncontemporaneous (first) variation of the action for L is given by

$$\delta_t E\left[\int_{t_a}^{t_b}L\,dt\right]=E\left[\left(\frac{\partial L}{\partial DX}+\frac{\partial L}{\partial D_*X}\right)\delta X\Big|_{t_a}^{t_b}\right]+E\left[\int_{t_a}^{t_b}\left(\frac{\partial L}{\partial X}-D_*\frac{\partial L}{\partial DX}-D\frac{\partial L}{\partial D_*X}\right)\delta X\,dt\right]-E[\mathscr{E}\,\delta_t t\,|_0^1]+E\left[\int_0^1\frac{1}{2}(D_u\mathscr{E}+D_{*u}\mathscr{E})\,\delta_t t\,du\right].\tag{4.5}$$

Naturally, if t is not varied, we obtain again Hamilton's principle. In this variational formulation, the symmetry between the couples of stochastic processes $(X;(p,p_*))$ [in using Eqs. (2.18) and (2.19)] and $(t;\mathscr{E})$ appears clearly. The main difference resides in the fact that $t=t(u)$ is a trivial (namely deterministic) process with differentiable paths,

which explains why there exists only one momentum $\mathscr{E}$ conjugate to the time process.

By definition, and for a given Largrangian $L = T - V$, a dynamical Nelson process in $D_{x_a,t_a}^{x_b,t_b}$ (that is, with fixed end points in position and time) will be characterized by the property

$$\delta_t E\left[\int_{t_a}^{t_b} L\,dt\right] = 0 \tag{4.6}$$

for all the suitable δx and $\delta_t t$.

It follows from Theorem 5 that a dynamical Nelson process $x(t)$ satisfies simultaneously

$$D\frac{\partial L}{\partial D_* X} + D_*\frac{\partial L}{\partial DX} - \frac{\partial L}{\partial X} = 0, \tag{4.7}$$

$$\frac{d}{dt}E\left[\epsilon(X,DX,D_* X)\right] = 0. \tag{4.8}$$

For obtaining (4.8), two differential versions of Eq. (1.5) were used, namely one for $Y = 1$,

$$\frac{d}{dt}E[X] = E[DX],$$

and another for $X = 1$,

$$\frac{d}{dt}E[Y] = E[D_* Y].$$

By averaging these two expressions for $X = Y = \epsilon$, we get indeed Eq. (4.8). When the system is not conservative,

$$\frac{d}{dt}E\left[\epsilon(X,DX,D_* X,t)\right] = E\left[\frac{\partial \epsilon}{\partial t}\right]. \tag{4.8'}$$

To emphasize the difference between the two types of variations, we can also observe that, according to the two last terms of (4.5) for a dynamical Nelson process,

$$E[\delta_t \mathscr{E}] = \delta_t E[\mathscr{E}(X,DX,D_* X)] = 0, \tag{4.9}$$

that is, the constraint (3.2) used for Maupertuis' principle, while for Hamilton's principle $\delta E[\epsilon] \neq 0$ generally, since it gives the variation of energy between any two admissible paths. In other words, it is really necessary to modify the time interval $[t_a, t_b]$ in order to maintain the energy constant.

For example, using similar procedures as for Theorem 5, it may be seen that, on an extremal,

$$\delta_t E\left[\int_{t_a}^{t_b} 2T\,dt\right] = E\left[\frac{1}{2}(p + p_*)\delta x\Big|_{t_a}^{t_b}\right] + E\left[\int_0^1 \delta_t \mathscr{E}\,du\right]. \tag{4.10}$$

But the usual development for the left-hand side and the definition of $\delta_t T$ modifies the latter to

$$\delta E\left[\int_{t_a}^{t_b} 2T\,dt\right] - E\left[\int_0^1 2T\,d(\delta_t t)\right].$$

Now by the definition $\delta_t \mathscr{E}$, the right-hand term of Eq. (4.10) is also

$$E\left[\frac{1}{2}(p + p_*)\,\delta x\Big|_{t_a}^{t_b}\right] + E\left[\int_{t_a}^{t_b} \delta\mathscr{E}\,dt\right] - E\left[\int_0^1 2T\,d(\delta_t t)\right].$$

Thus we find:

Theorem 6: For conservative stochastic systems with Lagrangian (2.14), the contemporaneous (first) variation of the reduced action on an extremal is

$$\delta E\left[\int_{t_a}^{t_b} 2T\,dt\right] = E\left[\frac{1}{2}(p + p_*)\,\delta x\Big|_{t_a}^{t_b}\right] + E\left[\int_{t_a}^{t_b} \delta\mathscr{E}\,dt\right]. \tag{4.11}$$

Even for fixed end points ($\delta X \in \Delta$) the second term on the right-hand side does not vanish generally.

5. APPLICATION TO CLASSICAL STATISTICAL MECHANICS

The first illustrations of these stochastic variational principles concerns some "cryptodeterministic" processes, in Whittaker's terminology,[14] used in classical statistical mechanics.

Indeed, a simple kinematic assumption on the nature of motion is that

$$dX(t) = v(X(t),t)\,dt, \tag{5.1}$$

where v is a smooth (unspecified) function of $\mathbb{R}^{l+1}$ and the only random element involved is the initial condition

$$x(t_0) = x_0 \in H. \tag{5.2}$$

In this case the two filtrations $\mathfrak{P}_t$ and $\mathfrak{F}_t$ are equal and do not vary with time. One speaks of "deterministic filtrations." The two canonical decompositions (1.1) and (1.2) coincide and are trivial to the extent that $M_t = M_{*t} = 0$, $\forall t \in I$. Since X_t is of bounded variation, the paths are (a.e.) differentiable. Therefore, the mean velocities (1.3) and (1.4) also coincide,

$$DX_t = D_* X_t = \frac{dX}{dt}. \tag{5.3}$$

It is then natural to choose for this case of Lagrangian $L(X,DX,D_* X,t)$ symmetrical in $DX, D_* X$,

$$L(X,DX,D_* X,t) = \tfrac{1}{4}M\,|DX|^2 + \tfrac{1}{4}M\,|D_* X|^2 - V(X,t) \tag{5.4}$$

since, by Eq. (5.3), it is nothing else than the Lagrangian for a particle with mass M in a given potential $V(X,t)$. Note that the "kinetic energy" $T(DX,D_* X) = \frac{1}{4}M\,|DX|^2 + \frac{1}{4}M\,|D_* X|^2$ is homogeneous of degree 2 in DX and $D_* X$ [compare with Eq. (2.14)]. It follows from the stochastic Hamilton's principle that the dynamics of this stochastic dynamical system will be given by the YEN equation (2.3), namely, here by

$$M\frac{d^2 X}{dt^2} = -\nabla V(X,t). \tag{5.5}$$

Thus, Newton's dynamical law follows naturally from the kinematic assumption (5.1).

For noncontemporaneous variations, we use a time change (2.7), $T_\phi : X(u) \to X(t(u)) \equiv \bar{X}(u)$ such that, by Eq. (2.6),

$$dt = \phi\,du, \tag{5.6}$$

where ϕ is some unspecified strictly positive smooth function on $\mathbf{R}$. Provided that (5.6) has a unique solution, Eq. (2.8) [or (2.9)] is satisfied since

$$d\bar{X}(u) = v(\bar{X}(u),t(u)) \cdot \phi \cdot du \tag{5.7}$$

is equivalent to Eq. (5.1).

On the other hand, it is well known that under mild assumptions about Eq. (5.1), the probabilistic evolution of such a system can also be obtained in solving an initial value problem, namely, the Liouville equation

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}(v\rho), \tag{5.8}$$

where $\rho = \rho(X,t)$ is a probability density for the process $X(t)$.

In this way the framework for the variational description of a probabilistic system (5.1) [or (5.7)] is well defined. One verifies easily that the different principles proposed in the four first sections restore the main variational results used in classical mechanics.[12]

Consider the strictly deterministic situation where the initial condition $X_0 \in H$ degenerates to a true ($\in \mathbb{R}^l$) constant x_a for a conservation system.

Then, since we can write the symmetrical Lagrangian (5.4),

$$L(X, DX = \dot{X}, D_* X = \dot{X}) \equiv L_c(X, \dot{X}), \tag{5.9}$$

where the subscript c denotes the classical definition, the two momenta coincide and

$$p = p_* = p_c. \tag{5.10}$$

The function v of Eq. (5.1) is a gradient. Indeed, if one considers the (classical) action for the Lagrangian L_c along extremal $X^x(s)$ trajectories between the fixed point (t_a, x_a) and the free final point (t,x), one obtains a real valued function of x and t,

$$\int_{t_a}^{t} L_c(X^x(s), \dot{X}^x(s))\, ds, \tag{5.11}$$

well defined for $|t - t_a|$ sufficiently small,[11] with the property

$$Mv(x,t) = \operatorname{grad} \int_{t_a}^{t} \{L_c(X^x(s), \dot{X}^x(s))\}\, ds. \tag{5.12}$$

As heuristic example, we will use Theorem 6 for the one-dimensional case and a variation between two dynamical paths of different constant energies in a periodic motion of period τ. If I is the reduced action, we have

$$\delta I = \delta \mathscr{E} \int_{t_a}^{t_a+\tau} dt = \delta \mathscr{E} \cdot \tau \tag{5.13}$$

by hypothesis of periodicity, which means that

$$\tau = \frac{\partial I}{\partial \mathscr{E}}. \tag{5.14}$$

But, by homogeneity of T, $2T_c = p_c \dot{X}$ and then

$$I = \int_{t_a}^{t_a+\tau} p_c \cdot \dot{X}\, dt = \oint p \cdot dX, \tag{5.15}$$

where $\oint$ denotes an integration over one period.

Historically, the action (5.15) was the main object of the old quantum theory.[15] Indeed, Planck's recipe for quantizing the oscillator (and other one-dimensional periodic systems[16]) was to put

$$I = \oint p \cdot dX = nh, \quad n \in \mathbb{N}, \tag{5.16}$$

where h is Planck's constant. In the limit of Bohr's correspondence principle (namely when n is a large number), the change of I and $\mathscr{E}$, using Eqs. (5.14) and (5.15), was supposed to be given by

$$\Delta I = \Delta n \cdot h, \quad \Delta \mathscr{E} = \Delta n \cdot h\nu, \tag{5.17}$$

where ν denotes a frequency τ^{-1}. Thus, since for an harmonic oscillator with potential $V(X) = \frac{1}{2}M\omega^2 X^2$,

$$I = \mathscr{E}\tau + \text{const}, \quad \text{where} \quad \tau = 2\pi/\omega, \tag{5.18}$$

the recipe (5.16) gives the quantized levels of energy

$$E = n\hbar\omega + \text{const}. \tag{5.19}$$

Today we know that the constant is $\frac{1}{2}\hbar\omega$.

It must be emphasized that no trace of Eq. (5.16) survives in conventional quantum mechanics, except in the study of the semiclassical limit (WKB method; cf. Refs. 11, 17, and 18).

6. APPLICATION TO STOCHASTIC MECHANICS

The second illustration is given by Nelson's stochastic mechanics, created in 1966.[3,7,8] It involves a more general kinematic assumption than Eq. (5.1), namely,

$$dX(t) = b(X(t),t)\, dt + \sqrt{\hbar/M}\, dW(t) \tag{6.1}$$

and $X(0) = X_0 \in H$, independent of W.

Here W is a Brownian motion on $\mathbb{R}^l$, $\hbar$ and M two positive constants, and b a smooth (unspecified) function on $\mathbb{R}^{l+1}$. Equation (6.1) is a differential form of the canonical decomposition (1.1) for a filtration $\mathfrak{P}_t$ containing $\sigma\{X(s), s \leqslant t\}$ and W the $\mathfrak{P}_t$-martingale part of $X(t)$. After integration, the first integral of the right-hand member of Eq. (6.1) is a Stieltjes integral for each sample and the second one a $\mathfrak{P}_t$-martingale integral of Itô. Furthermore, $X(t)$ given by Eq. (6.1) is a Markoff process.

A filtration $\mathfrak{F}_t = \sigma\{X(u); u \geqslant t\}$ contains the future of the process $X(t)$, and the associated Meyer's canonical decomposition (1.2) takes the differential form

$$dX(t) = b_*(X(t),t)\, dt + \sqrt{\hbar/M}\, dW_*(t), \tag{6.2}$$

where b_* is some other smooth function on $\mathbb{R}^{l+1}$ and W_* the $\mathfrak{F}_t$-martingale part of $X(t)$. By the Markoff property, $\mathfrak{P}_t$ and $\mathfrak{F}_t$ are conditionally independent given the present $\mathfrak{P}_t \cap \mathfrak{F}_t$.

In this case the sample paths are not differentiable (a.e.) and the two mean velocities (1.3) and (1.4) are different. In fact,

$$DX_t = b(X(t),t), \quad D_* X_t = b_*(X(t),t). \tag{6.3}$$

Let us consider again the Lagrangian (5.4). It follows from the stochastic Hamilton's principle that the dynamics of this Markoffian system will be defined by YEN equation (2.3),

$$\tfrac{1}{2}M(DD_* X(t) + D_* DX(t)) = -\nabla V(X(t),t), \tag{6.4}$$

which is clearly a generalization of Newton's law (5.5).

For noncontemporaneous variations, we need a (deterministic)time change for semimartingales.[19] It is defined by

$$T_\phi : X(u) \to X(t(u)) = \overline{X}(u), \tag{6.5}$$

where $t(u)$ satisfies for some unspecified but smooth ϕ

$$\frac{dt}{du} = \phi(t(u)) > 0. \tag{6.6}$$

One proves easily the existence of a $\mathfrak{P}_u$-martingale $\overline{W}(u) \equiv W(t(u))$ for the measure PT_ϕ^{-1} such that Eq. (6.1) is modified to

$$d\overline{X}(u) = b(\overline{X}(u),t(u)) \cdot \phi \cdot du + \sqrt{\hbar/M}\,\phi^{1/2}\, d\overline{W}(u). \tag{6.7}$$

If Eq. (6.6) can be uniquely solved for $t(u)$, for example, when ϕ is a bounded Borel measurable function Lipschitz continuous in t, Eq. (6.7) is well defined.

One follows the same procedure for Eq. (6.2). Thus, the relations (2.8) and (2.9) are satisfied since

$$D_u\overline{X} = b(\overline{X}(u),t(u)) \cdot \phi, \quad D_{*u}\overline{X} = b_*(\overline{X}(u),t(u)) \cdot \phi. \tag{6.8}$$

Since the partial differential operator

$$D^+ \equiv -\frac{\partial}{\partial t} - b^i\partial_i - \partial_i b^i + \tfrac{1}{2}(\hbar/M)\partial_i\partial_i$$

is hypoelliptic,[20] the process X_t has a smooth density on $\mathbb{R}^l \times I$ (with respect to the Lebesgue measure d^lX), $\rho(X,t)$ satisfying the forward equation

$$\frac{\partial\rho}{\partial t} = -\operatorname{div}(b\rho) + (\hbar/2M)\Delta\rho. \tag{6.9}$$

By the averaging with the other equation for ρ corresponding to the decomposition (6.2), Nelson shows that

$$\frac{\partial\rho}{\partial t} = -\operatorname{div}(v\rho), \tag{6.10}$$

where v is defined by $v = \frac{1}{2}(b + b_*)$. This function v is a gradient. Indeed, if one considers the variable end point problem associated with Eq. (2.2), it may be seen that[9]

$$(M/\hbar)v(X(t),t) = \operatorname{grad} S(X(t),t), \tag{6.11}$$

for some smooth function S. In these conditions, Nelson proved that

$$\psi(X,t) \equiv \rho^{1/2}(X,t)e^{iS(X,t)} \tag{6.12}$$

satisfies

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2M}\Delta\psi + V\psi, \tag{6.13}$$

that is, the Schrödinger equation for a particle with mass M in the given potential V (if $l = 3$). The constant $\hbar$ of Eqs. (6.1) and (6.2) is identified with Planck's constant over 2π, and it follows from Eq. (6.12) that $|\psi(X,t)|^2\, d^lX = P(X(t) \in d^lX)$.

The proof of the existence of the associated time symmetric semimartingales was given recently by Carlen.[21]

The conservation of energy used as far as noncontemporaneous variations are concerned [for example, Eq. (2.22) or more generally Eq. (4.8′)] is nothing else but the quantum form of this conservation law. Hence, for a quantum stationary state with energy eigenvalue E_n, we get

$$E[\mathscr{E}(X,DX,D_*X)] = E_n, \tag{6.14}$$

where $\mathscr{E}$ is given by (2.21),

$$\mathscr{E}(X,DX,D_*X) = \tfrac{1}{4}M|DX|^2 + \tfrac{1}{4}M|D_*X|^2 + V(X) \tag{6.15}$$

or, more explicitly, in terms of Eqs. (6.3),

$$\mathscr{E} = \tfrac{1}{4}M|b|^2 + \tfrac{1}{4}M|b_*|^2 + V. \tag{6.16}$$

Since for a stationary state[8,22]

$$DX(t) = -D_*X(t), \tag{6.17}$$

Eq. (6.16) is reduced to

$$\mathscr{E} = \tfrac{1}{2}M|b|^2 + V. \tag{6.16′}$$

Come back to the one-dimensional case for Theorem 6 examined in Sec. 5 within a deterministic frame. It follows from Theorem 2 that the reduced action can be written

$$I = \frac{1}{2}E\left[\mathfrak{P}\int_{t_a}^{t_a+\tau} p\cdot dX + \mathfrak{F}\int_{t_a}^{t_a+\tau} p_*\cdot dX\right]$$
$$= E\left[\int_{t_a}^{t_a+\tau} 2T\,dt\right] \tag{6.18}$$

where, by Eqs. (2.18), (2.19), and the definition (5.4) of the Lagrangian (conservative case)

$$p = M\,DX, \quad p_* = M\,D_*X. \tag{6.19}$$

Taking into account Eqs. (2.30) and (2.31),

$$I = E\left[\int_{t_a}^{t_a+\tau}\frac{M}{2}(DX)^2\,dt + \int_{t_a}^{t_a+\tau}\frac{M}{2}(D_*X)^2\,dt\right]. \tag{6.20}$$

For stationary states, by Eq. (6.17), the two integrals coincide. Then, it will be sufficient to compute the first expectation for a state

$$\psi(X,t) = \varphi_n(X)e^{-(i/\hbar)E_nt}, \tag{6.21}$$

where $\varphi_n \in L^2(\mathbb{R};\mathbb{C})$ is an eigenfunction of the stationary Schrödinger equation associated with Eq. (6.13) for the energy eigenvalue E_n. The Nelson process in this case is indeed a (strictly) stationary Markoff process whose invariant measure is $\rho(X)\,dX = |\varphi_n(X)|^2\,dX$.

We find, using an evident notation for Eq. (6.18),

$$I_n(\mathfrak{P}) = \tau\{E_n - E[V]\} = I_n(\mathfrak{F})$$
$$= \tau E[T]. \tag{6.22}$$

Now, by the virial theorem,[2] we have

$$2E[T] = E[X\cdot\nabla V]. \tag{6.23}$$

Provided that V is homogeneous of degree m, it follows from Euler's theorem that Eq. (6.23) is simplified to

$$E[T] = \tfrac{1}{2}mE[V]. \tag{6.24}$$

The comparison with Eq. (6.22) shows that

$$E[V] = [2/(m+2)]E_n \tag{6.25}$$

and then

$$I_n(\mathfrak{P}) = I_n(\mathfrak{F}) = [m/(m+2)]\tau E_n \tag{6.26}$$

or, by Eq. (6.20),

$$I = [2m/(m+2)]\tau E_n, \tag{6.27}$$

which yields for the harmonic case $m = 2$,

$$I = E_n\tau$$
$$= (n + \tfrac{1}{2})h, \quad n\in\mathbb{N}, \tag{6.28}$$

where the eigenvalue for $V(X) = \frac{1}{2}M\omega^2X^2$ and $\tau = 2\pi/\omega$ was used. Equation (6.28) for the reduced action (6.18) is the well-defined quantum analog of the historic recipe (5.16).

Observe that a direct (and naive) application of Theorem 6 for the variation between two stationary states of different energies in the limit of the correspondence principle and for periodic orbits gives

$$\delta I = \delta E[\mathscr{E}]\cdot\tau \tag{6.29}$$

or

$$\tau = \frac{\partial I}{\partial e}, \tag{6.30}$$

as for the classical case (5.14), if we put $e = E[\mathscr{E}]$. It is worthwhile to emphasize that Eqs. (6.27) and (6.28) are fully quantum mechanical expressions; they have nothing to do with a semiclassical limit.

7. CONCLUDING REMARKS

Let us indicate briefly that the proposed stochastic variational frame can be equally useful for the study of the semiclassical limit in quantum mechanics[23] and also for extensions of quantum mechanics in the domains where Hamiltonian or Lagrangian classical theories are powerless.[24,25] It is well adapted to the research of the possible dynamics in nonequilibrium statistical thermodynamics.[9] Furthermore, it can be formulated on any Riemannian manifold[9,26,27] and was used recently with success to investigate the Navier–Stokes equation.[28] A self-contained review on stochastic calculus of variations is given in Ref. 9.

ACKNOWLEDGMENTS

The author expresses his thanks to Professor E. Nelson, at Princeton, and P. A. Meyer, at Strasbourg, for their kind interest in stochastic calculus of variations and for their constructive criticism, to K. Yasue for many discusions on this subject (and on many others!), and to C. P. Enz for support.

[1]K. Yasue, J. Math. Phys. **22**, 1010 (1981).
[2]K. Yasue, J. Funct. Anal. **41**, 327 (1981).
[3]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).
[4]K. Itô, "Extension of Stochastic Integrals," in *Proceedings of the International Symposium of SDE, Kyoto 1976*, edited by K. Itô (Wiley, New York, 1978).
[5]C. Dellacherie and P. A. Meyer, *Probabilités et potentiel* (Hermann, Paris, 1975).
[6]K. Yasue, J. Math. Phys. **23**, 1577 (1982).
[7]E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1984).
[8]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[9]J.-C. Zambrini, "Stochastic Dynamics: A Review of Stochastic Calculus of Variations," Int. J. Theor. Phys. (to be published).
[10]W. A. Zheng and P. A. Meyer, "Quelques résultats de Mécanique Stochastique," in *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1983), Vol. 18.
[11]V. Arnold, *Méthodes mathématiques de la mécanique classique* (Mir, Moscow, 1976).
[12]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed.
[13]K. Itô and S. Watanabe, "Introduction to Stochastic Differential Equations," *Proceedings of the International Symposium of SDE, Kyoto 1976*, edited by K. Itô (Wiley, New York, 1978).
[14]E. T. Whittaker, *Analytical Dynamics* (Cambridge U. P., London, 1937), 4th ed.
[15]D. ter Haar, *The Old Quantum Theory* (Pergamon, Oxford, 1967).
[16]*Sources of Quantum Mechanics*, edited by B. L. Van der Waerden (Dover, New York, 1967).
[17]V. Maslov, *Théorie des perturbations et méthodes asymptotiques* (Dunod, Paris, 1972).
[18]M. C. Gutzwiller, "Path Integrals and the Relation Between Classical and Quantum Mechanics," in *Path Integrals*, edited by G. Papadopoulos and J. T. Devreese (Plenum, New York, 1978).
[19]S. R. S. Varadhan, *Diffusion Problems and Partial Differential Equations* (Tata Inst. of Fund. Research, Bombay, 1980).
[20]L. Hörmander, *Linear Partial Differential Operators* (Springer, New York, 1963).
[21]E. Carlen, "Conservative diffusion," Commun. Math. Phys. (to appear).
[22]E. Nelson, 'Connection between Brownian Motion and Quantum Mechanics," talk, Einstein Symposium, Berlin, 1979.
[23]J.-C. Zambrini and K. Yasue, Ann. Phys. (N.Y.) **143**, 54 (1982).
[24]K. Yasue, Ann. Phys. (N.Y.) **114**, 479 (1978).
[25]K. Yasue, J. Math. Phys. **19**, 1892 (1978).
[26]D. Dohrn and F. Guerra, Lett. Nuovo Cimento **22**, 121 (1978).
[27]P. A. Meyer, "Géométrie différentielle stochastique (bis),"in *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, 1982), Vol. 921, p. 165.
[28]K. Yasue, "A variational principle for the Navier–Stokes equation," J. Funct. Anal. **51**, 133 (1983).

A bounded convergence theorem for the Feynman integral

G. W. Johnson
Department of Mathematics and Statistics, University of Nebraska, Lincoln, Nebraska 68588-0323

(Received 15 April 1983; accepted for publication 22 July 1983)

We give a bounded convergence theorem for the Feynman integral for the class of bounded, measurable potentials.

PACS numbers: 02.50.Sk, 03.65.Db

1. INTRODUCTION

There have been many different approaches to the Feynman integral which have illuminated the subject in a variety of ways, but, to the best of our knowledge, there has never been any really satisfactory convergence theorem applying to any reasonably large class of potentials. The main theorem of this paper gives such a theorem for the class of bounded, measurable potentials. The setting is an approach to the Feynman integral introduced by Cameron and Storvick[1] and studied further by them[2–5] and others including Haugsby[6] and the author and Skoug.[7–9] Parts of Ref. 9 are especially relevant to the present paper.

The class of bounded potentials does not include by any means all the potentials of physical interest; indeed it fails to include such basic things as the harmonic oscillator and Coulomb potentials. However, a review of the physical literature shows that bounded potentials are of considerable interest in quantum mechanics. They are also of interest in connection with the Korteweg–de Vries (or KdV) equation. A beautiful relationship between the KdV equation and the Schrödinger equation of quantum mechanics has been discovered. The function giving the boundary data of the KdV equation becomes the potential in the related Schrödinger equation. The natural potentials that arise in this way are bounded. See Ref. 10 for information and references on this subject.

There has been considerable progress on the Feynman integral in recent years. However, the absence of satisfactory convergence theorems has certainly been one of the basic difficulties in the theory. The kind of limiting arguments so strongly associated with integration theories have not been possible. To take another point of view, the mathematical theories have not had very satisfactory stability properties even though the physical world apparently does. If $\{\theta_m\}$ is a sequence of potentials which converges pointwise, or even uniformly, to θ, there has been no assurance that the corresponding Feynman integrals converge.

Remark: The work below permits the potentials to be complex-valued (or C-valued). Such potentials seem to be of interest physically in connection with "open quantum systems." See Refs. 11–15 for references.

2. BACKGROUND

Let $t>0$ be fixed. $C_0[0,t]$ will denote one-dimensional Wiener space, that is, the set of all continuous paths on $[0,t]$ which vanish at 0. $C_0^v[0,t]$ will denote the product of v copies of $C_0[0,t]$. We will consider $C_0^v[0,t]$ as equipped with v-dimensional Wiener measure m which is just the product of v one-dimensional Wiener measures.

Given a C-valued function F on $C_0^v[0,t]$, $\lambda>0$, ψ in $L_2(\mathbb{R}^v)$, and an element ξ of $\mathbb{R}^v$, we consider the expression

$$(I_\lambda(F)\psi)(\xi) := \int_{C_0^v[0,t]} F(\lambda^{-1/2}X+\xi) \times \psi(\lambda^{-1/2}X(t)+\xi)\,dm(X). \tag{1}$$

This formula may define, for each $\lambda>0$, a bounded linear operator $I_\lambda(F)$ on $L_2(\mathbb{R}^v)$. If this is so, and if the operator-valued function $\lambda\to I_\lambda(F)$ has an analytic continuation to $C^+ := \{\lambda \text{ in } C: \operatorname{Re}\lambda>0\}$, we denote this analytic continuation $I_\lambda^{\mathrm{an}}(F)$ and call it the operator-valued, analytic Wiener integral of F with parameter λ. Finally given a real parameter q, $q\neq 0$, the operator-valued analytic Feynman integral of F with parameter q is denoted $J_q^{\mathrm{an}}(F)$ and is defined by

$$J_q^{\mathrm{an}}(F) := \lim_{\lambda\to -iq} I_\lambda^{\mathrm{an}}(F), \tag{2}$$

where the limit is taken in the strong operator topology and where λ approaches $-iq$ through C^+. These definitions were given by Cameron and Storvick in Ref. 1.

It is only for certain special classes of functions F for which the above definitions are relevant to quantum mechanics. These are primarily functions of the type that we are about to describe or certain variations of them.

Let θ be a C-valued, Lebesgue measurable function on $\mathbb{R}^v$, and let

$$F(X) := \exp\left\{\int_0^t \theta(X(s))ds\right\}. \tag{3}$$

In this paper, θ will be required to be bounded. As we continue, we will have a sequence of bounded functions $\{\theta_m\}$ such that $\theta_m\to\theta$ almost everywhere (a.e.). Let

$$F_m(X) := \exp\left\{\int_0^t \theta_m(X(s))ds\right\}. \tag{4}$$

Remarks: (i) Regarding the Feynman integral as an operator rather than as a number is a natural point of view. If θ in Eq. (3) is thought of as the potential for the quantum system and if ψ is the initial probability amplitude, then $J_q^{\mathrm{an}}(F)\psi$ is the probability amplitude giving the state of the system at time t.

(ii) From this point on, q will be an arbitrary but fixed nonzero real number, and so we will drop the subscript q in the expression $J_q^{\mathrm{an}}(F)$. Further since, except for one remark, we will discuss only the operator-valued *analytic* Feynman

 0022-2488/84/051323-04$02.50

integral, we will write simply $J(F)$ rather than $J_q^{\rm an}(F)$.

The following is a special case of part of Ref. 9, Theorem 5.1. (In retrospect, it is clear that the exposition in Ref. 9 left something to be desired. We should have emphasized and stated more explicitly the special cases of our theorems which were likely to be of most interest.)

Theorem 1: Let θ be an essentially bounded, complex-valued, Lebesgue measurable function on $\mathbb{R}^v$ and let F be given by Eq. (3). Then $J(F)$ exists as a bounded linear operator on $L_2(\mathbb{R}^v)$ and, for any ψ in $L_2(\mathbb{R}^v)$, it is given by

$$\begin{aligned}(J(F)\psi)(\xi) = \sum_{n=0}^{\infty} \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} (-iq)^{v/2}[2\pi(t-s_n)]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_n)\exp\left[\frac{iq\|V_n-\xi\|^2}{2(t-s_n)}\right](-iq)^{v/2}[2\pi(s_n-s_{n-1})]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_{n-1})\exp\left[\frac{iq\|V_{n-1}-V_n\|^2}{2(s_n-s_{n-1})}\right]\times\cdots\times(-iq)^{v/2}[2\pi(s_2-s_1)]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_1)\exp\left[\frac{iq\|V_1-V_2\|^2}{2(s_2-s_1)}\right](-iq)^{v/2}[2\pi s_1]^{-v/2} \\ \times \int_{\mathbb{R}^v} \psi(V_0)\exp\left[\frac{iq\|V_0-V_1\|^2}{2s_1}\right] dV_0 dV_1 \cdot\cdots\cdot dV_n \cdot ds_1 \cdot\cdots\cdot ds_n, \end{aligned} \tag{5}$$

where the integrals over $\mathbb{R}^v$ are interpreted in the mean (as in the Fourier–Plancherel theory) and the integral over

$$\Delta_n := \{(s_1,\ldots,s_n): 0<s_1<s_2<\cdots<s_{n-1}<s_n<t\} \tag{6}$$

may be interpreted either as a Bochner integral (Ref. 16, pp. 71–89) of an $L_2(\mathbb{R}^v)$-valued function or as a Lebesgue integral.

The theorem as just stated gives the basic information, but we will need some additional notation and some further information from Ref. 9 connected with the terms of Eq. (5). Let

$$\begin{aligned}(B_n(F)(s_1,\ldots,s_n)\psi)(\xi) = (-iq)^{v/2}[2\pi(t-s_n)]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_n)\exp\left[\frac{iq\|V_n-\xi\|^2}{2(t-s_n)}\right](-iq)^{v/2}[2\pi(s_n-s_{n-1})]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_{n-1})\exp\left[\frac{iq\|V_{n-1}-V_n\|^2}{2(s_n-s_{n-1})}\right]\times\cdots\times(-iq)^{v/2}[2\pi(s_2-s_1)]^{-v/2} \\ \times \int_{\mathbb{R}^v} \theta(V_1)\exp\left[\frac{iq\|V_1-V_2\|^2}{2(s_2-s_1)}\right]\times(-iq)^{v/2}[2\pi s_1]^{-v/2} \\ \times \int_{\mathbb{R}^v} \psi(V_0)\exp\left[\frac{iq\|V_0-V_1\|^2}{2s_1}\right] dV_0 dV_1\cdots dV_n, \end{aligned} \tag{7}$$

and let

$$(A_n(F)\psi)(\xi) = \int_{\Delta_n} (B_n(F)(s_1,\ldots,s_n)\psi)(\xi)\,dS, \tag{8}$$

where $S=(s_1,\ldots,s_n)$. Note that $(A_n(F)\psi)(\xi)$ is just the nth term of the series in Eq. (5).

It is useful to think of $B_n(F)(s_1,\ldots,s_n)$ as the composition of a succession of alternating convolution and multiplication operators. Let

$$e_s(U) := (-iq)^{v/2}[2\pi s]^{-v/2}\exp[iq\|U\|^2/2s] \tag{9}$$

and let C_s be the operator of convolution by e_s; that is,

$$(C_s\psi)(\xi) = (-iq)^{v/2}[2\pi s]^{-v/2}\int_{\mathbb{R}^v} \psi(U)e_s(\xi-U)dU. \tag{10}$$

The convolution operators in (10) are well known (as is discussed in Ref. 9; proof of Lemma 1.1) to be unitary operators on $L_2(\mathbb{R}^v)$. If $M_\theta = M$ denotes the operator of multiplication by θ, then M is also a bounded linear operator on $L_2(\mathbb{R}^v)$ and $\|M\| = \|\theta\|_\infty$. Now, looking carefully at Eq. (7), we see that

$$\begin{aligned}B_n(F)(s_1,\ldots,s_n)\psi = (C_{t-s_n}\circ M\circ C_{s_n-s_{n-1}}\circ M\circ\cdots \\ \circ C_{s_2-s_1}\circ M\circ C_{s_1})\psi, \end{aligned} \tag{11}$$

and so

$$\|B_n(F)(s_1,\ldots,s_n)\psi\|_2 \leq \|\theta\|_\infty^n \|\psi\|_2. \tag{12}$$

The fact that $B_n(F)(s_1,\ldots,s_n)\psi$ is Bochner integrable over Δ_n comes out of the work in Ref. 9, and so it follows from a basic inequality for Bochner integrals (Ref. 16; Theorem 3.7.6, p. 82) that

$$\begin{aligned}\|A_n(F)\psi\| &\leq \int_{\Delta_n} \|B_n(F)(s_1,\ldots,s_n)\psi\|_2 dS \\ &\leq ((\|\theta\|_\infty t)^n/n!)\|\psi\|_2. \end{aligned} \tag{13}$$

Now since from (5), (7), and (8)

$$J(F)\psi = \sum_{n=0}^{\infty} A_n(F)\psi, \tag{14}$$

we have from (13),

$$\|J(F)\psi\|_2 \leq \|\psi\|_2 \sum_{n=0}^{\infty} (\|\theta\|_\infty t)^n/n! = \|\psi\|_2 \exp[\|\theta\|_\infty t]. \tag{15}$$

All of the above information is contained in Ref. 9 but in Ref. 9, much of the information is set in a more general framework and/or is somewhat hidden in proofs. Also we are using different notation here.

3. THE BOUNDED CONVERGENCE THEOREM

Theorem 2: (A bounded convergence theorem for the Feynman integral.) Let $\{\theta_m\}$ be a sequence of complex-valued, Lebesgue measurable functions on $\mathbb{R}^v$ all of which are essentially bounded by the number L. Suppose that $\theta_m \to \theta$ a.e. on $\mathbb{R}^v$. Then, of course, θ is also essentially bounded by L and, by Theorem 1, $J(F)$ and $J(F_m)$, $m = 1,2,...$, all exist where F is given by (3), and F_m is given by (4). Further, the analytic, operator-valued Feynman integral of F_m converges in the strong operator topology to the analytic, operator-valued Feynman integral of F as $m \to \infty$. In symbols,

$$J(F_m) \to J(F) \text{ in the strong operator topology as } m \to \infty. \tag{16}$$

The following lemma is certainly known, but its proof is simple, and the result plays a key role here, and so we include it.

Lemma: Let $\{\theta_m\}$ be a sequence of complex-valued, Lebesgue measurable functions on $\mathbb{R}^v$ all of which are essentially bounded by L. Suppose that $\theta_m \to \theta$ a.e. on $\mathbb{R}^v$. Then θ is essentially bounded by L, and the sequence of multiplication operators M_{θ_m} converges in the strong operator topology to M_θ as $m \to \infty$.

Proof: Let ψ be in $L_2(\mathbb{R}^v)$. We must show that

$$\int_{\mathbb{R}^v} |\theta_m(U)\psi(U) - \theta(U)\psi(U)|^2 dU \to 0 \quad \text{as } m \to \infty.$$

Now $\theta_m(U)\psi(U) \to \theta(U)\psi(U)$ a.e. since $\theta_m(U) \to \theta(U)$ a.e. Also $4L^2|\psi(U)|^2$ is in $L_1(\mathbb{R}^v)$ and is a dominating function for the sequence $|\theta_m(U)\psi(U) - \theta(U)\psi(U)|^2$. Simply apply the ordinary dominated convergence theorem for Lebesgue measure on $\mathbb{R}^v$ to finish the proof.

Proof (of the theorem): Fix ψ in $L_2(\mathbb{R}^v)$. Also, for now, we fix n. We first claim that for every $(s_1,...,s_n)$ in Δ_n,

$$\|B_n(F_m)(s_1,...,s_n)\psi - B_n(F)(s_1,...,s_n)\psi\|_2 \to 0 \quad \text{as } m \to \infty. \tag{17}$$

This follows form the Lemma, formula (11), and the fact that the composition of operators is jointly continuous in the strong operator topology provided one of the operators is restricted to lie in a bounded subset of the space of bounded linear operators on $L_2(\mathbb{R}^v)$.

Now applying Eq. (12) to the functions F_m and using the dominated convergence theorem for Bochner integrals (Ref. 16, Theorem 3.7.9, p. 83), we get

$$\|A_n(F_m)\psi - A_n(F)\psi\|_2 \to 0 \quad \text{as } m \to \infty. \tag{18}$$

We now know that each term of the series (14) for $J(F_m)\psi$ converges in $L_2(\mathbb{R}^v)$ norm to the corresponding term of the series for $J(F)\psi$. It remains to show that

$$\|J(F_m)\psi - J(F)\psi\|_2 \to 0 \quad \text{as } m \to \infty. \tag{19}$$

Given $\epsilon > 0$, let N_0 be so large that

$$\sum_{n=N_0+1}^{\infty} \|\psi\|_2 \frac{(tL)^n}{n!} < \frac{\epsilon}{4}. \tag{20}$$

Now, using (18), let N be so large that for $m \geq N$,

$$\sum_{n=0}^{N_0} \|A_n(F_m)\psi - A_n(F)\psi\|_2 < \frac{\epsilon}{2}. \tag{21}$$

Now let $m \geq N$, then using (14), (21), (13), and (20) we can write

$$\begin{aligned}
&\|J(F_m)\psi - J(F)\psi\|_2 \\
&= \left\| \sum_{n=0}^{\infty} A_n(F_m)\psi - \sum_{n=0}^{\infty} A_n(F)\psi \right\|_2 \\
&= \left\| \sum_{n=0}^{N_0} [A_n(F_m)\psi - A_n(F)\psi] + \sum_{n=N_0+1}^{\infty} A_n(F_m)\psi \right. \\
&\quad \left. - \sum_{n=N_0+1}^{\infty} A_n(F)\psi \right\|_2 \\
&\leq \sum_{n=0}^{N_0} \|A_n(F_m)\psi - A_n(F)\psi\|_2 \\
&\quad + \sum_{n=N_0+1}^{\infty} \|A_n(F_m)\psi\|_2 + \sum_{n=N_0+1}^{\infty} \|A_n(F)\psi\|_2 \\
&< \frac{\epsilon}{2} + \sum_{n=N_0+1}^{\infty} \frac{(tL)^n}{n!} \|\psi\|_2 + \sum_{n=N_0+1}^{\infty} \frac{(tL)^n}{n!} \|\psi\|_2 \\
&< \epsilon/2 + \epsilon/4 + \epsilon/4 = \epsilon
\end{aligned}$$

as desired.

We describe a simple corollary of Theorem 2: If $\{D_m\}$ is a sequence of operators in a normed linear space such that $D_m \to D$ in the strong operator topology and if $\|X_m - X\| \to 0$, then $\|D_m X_m - DX\| \to 0$ as $m \to \infty$. This easily proved, and well-known fact (Ref. 17, Problem 3.10, p. 151), combines with Theorem 2 to show that if $\theta_m \to \theta$ a.e. and $\|\psi_m - \psi\|_2 \to 0$, then $\|J(F_m)\psi_m - J(F)\psi\|_2 \to 0$ as $m \to \infty$. In the language of quantum mechanics and speaking a bit loosely, if both the potential θ and the initial probability amplitude ψ are perturbed slightly, then the probability amplitude at time t is changed only slightly.

We finish this paper with a series of remarks. The first of these remarks is crucial in understanding why the proof of Theorem 2 works.

Remarks: 1. The use of the dominated convergence theorem for Bochner integrals is essential to the proof of Theorem 2. An attempt to apply the ordinary dominated convergence theorem putting absolute values inside the integrals defining the B_n's simply doesn't work; the functions need not even be integrable. The use of the dominated convergence theorem for Bochner integrals allows us to take the norm inside the integral with respect to Δ_n but still to take advantage of the canceling effects of the integrals over $\mathbb{R}^v$.

2. In some earlier work[1,2,7] on bounded potentials θ,θ was assumed to be continuous a.e.. This assumption was reduced to measurability in Haugsby's thesis[6], and his improvement carried over to the setting of Ref. 9 and, in particular, to Theorem 1 above. While continuity a.e. is probably good enough to include all the bounded potentials of physical interest, it is useful in this paper to have the weaker measurability assumption. Because of it, we know immediately that the a.e. limit of the θ_m's of Theorem 2 is again a

potential to which Theorem 1 applies.

3. We have worked in this paper only with the operator-valued analytic Feynman integral $J^{an}(F)$. In fact, under the hypotheses of Theorem 1, Cameron and Storvick's operator-valued sequential Feynman integral $J^{seq}(F)$ also exists and equals $J^{an}(F)$ (Ref. 9; Theorem 4.6, p. 124). [See Ref. 1 or 7 for the definition of $J^{seq}(F)$.] Hence the results of this paper hold for $J^{seq}(F)$ just as well as for $J^{an}(F)$.

4. It seems likely that the arguments above can be extended to time dependent potentials and to certain classes of unbounded potentials of the type treated in Ref. 9. Further, the arguments may well extend to "Feynman-type integrals" based on Gaussian–Markov processes other than the Wiener process.

Note added in proof: Motivated by a preprint of this paper and discussions with the author, Michel Lapidus has recently proved a nice convergence theorem for his "modified Feynman integral."

[1]R. H. Cameron and D. A. Storvick, "An operator valued function space integral and a related integral equation," J. Math. Mech. **18**, 517–552 (1968).

[2]R. H. Cameron and D. A. Storvick, "An integral equation related to the Schroedinger equation with application to integration in function space," in *Problems in Analysis* (Princeton U. P., Princeton, NJ, 1970), pp. 175–193.

[3]R. H. Cameron and D. A. Storvick, "An operator valued function space integral applied to integrals of functions of class L_2," J. Math. Anal. Appl. **42**, 330–372 (1973).

[4]R. H. Cameron and D. A. Storvick, "An operator valued function space integral applied to integrals of functions of class L_1," Proc. London Math. Soc. **27**, 345–360 (1973).

[5]R. H. Cameron and D. A. Storvick, "An operator valued function space integral applied to multiple integrals of functions of class L_1," Nagoya Math. J. **51**, 91–122 (1973).

[6]Bernard O. Haugsby,"An operator valued integral in a function space of continuous vector valued functions," thesis, University of Minnesota, 1972.

[7]G. W. Johnson and D. L. Skoug, "A Banach algebra of Feynman integrable functionals with application to an integral equation formally equivalent to Schroedinger's equation," J. Functional Analysis **12**, 129–152 (1973).

[8]G. W. Johnson and D. L. Skoug, "The Cameron–Storvick function space integral: The L_1 theory," J. Math. Anal. Appl. **50**, 647–667 (1975).

[9]G. W. Johnson and D. L. Skoug, "The Cameron–Storvick function space integral: an $L(L_p,L_{p'})$ theory," Nagoya Math. J. **60**, 93–137 (1976).

[10]P. L. Bhatnagar, *Nonlinear Waves in One-dimensional Dispersive Systems* (Oxford U. P., Oxford, 1979).

[11]J. Blank, P. Exner, and M. Havlíček, "Quantum-mechanical pseudo-Hamiltonians," Czech, J. Phys. B. **29**, 1325–1341 (1979).

[12]E. B. Davies, *Quantum Theory of Open Systems* (Academic, New York, 1976).

[13]P. Exner, "Complex-potential description of the damped harmonic oscillator, I. The propagator, and II. The one-dimensional case," Preprints E2-81-608 and E2-81-609 of the Joint Institute for Nuclear Research, Dubna, U.S.S.R., 1981.

[14]P. Exner and G. I. Kolerov, "Uniform product formulae with application to the Feynman-Nelson integral for open systems," Preprint E2-81-186 of the Joint Institute for Nuclear Research, Dubna, U.S.S.R., 1981.

[15]P. Exner and G. I. Kolerov, "Path-integral expressions of dissipative dynamics," Phys. Lett. **83**, 203–206 (1981).

[16]E. Hillie and R. S. Phillips, *Functional Analysis and Semi-groups, American Mathematical Society Colloquium Publications*, No. 31 (Amer. Math. Soc., Providence, RI, 1957).

[17]T. Kato, *Perturbation Theory for Linear Operators*, 1st ed. (Springer-Verlag, New York, 1966).

Scattering by impurities in a solvable model of a three-dimensional crystal

S. Albeverio
Université d'Aix-Marseille II, Centre de Physique Theorique, CNRS, Marseille, France and Mathematisches Institut, Ruhr-Universität, Bochum, West Germany

R. Hoegh-Krohn
Université de Provence, Marseille, Centre de Physique Théorique, CNRS, Marseille, France and Matematisk Institutt, Universitet i Oslo, Oslo, Norway

M. Mebkhout
Université d'Aix-Marseille II, Centre de Physique Theorique, CNRS, Marseille, France

(Received 2 June 1983; accepted for publication 14 October 1983)

We compute explicitly the scattering amplitude for a quantum mechanical particle scattered by a discrete set of impurities in a three-dimensional crystal with point interactions.

PACS numbers: 02.70. + d, 03.65.Nk, 61.70.Rj

1. INTRODUCTION

The importance of exactly solvable models in theoretical physics can hardly be overestimated. Such models give orientation in situations where the analytic tools at our disposal are too weak to cope with the complex physical situation at hand. Moreover, they often give answers which are correct in certain controllable approximations. In solid-state physics as well as in the many-body problem, in the study of low energetical nuclear reactions, and in the study of electromagnetic phenomena, idealized point interactions (also called Fermi pseudopotentials or zero range interactions) have been introduced and studied extensively, particularly for the reason that they provide a very good approximation at low energy[1–25] and they are exactly solvable. Early work using such interactions has been done already in the thirties, by L. H. Thomas and continued by H. Bethe, R. Peierls, and others, in nuclear physics, by K. Huang, C. N. Yang, J. M. Luttinger, T. D. Lee, T. T. Wu, and others, in the fifties, for the many-body problem, and by R. de L. Kronig, W. G. Penney (1931) and continued by H. L. Frisch, S. P. Lloyd, R. E. Borland, and others, for solid-state physics. For references see Refs. 1–22. For applications to electromagnetic theory see Refs. 23–25.

In recent years, much activity concerning point interactions has been developed, along three main lines:

(a) Give a mathematical definition of point interactions suitable for the case of finitely and infinitely many centers. Here new techniques have been found, including methods of Dirichlet forms[26] and of nonstandard analysis,[12] see Refs. 1–29 and references therein. In particular, Hamiltonians of the form $H = -\Delta + \Sigma_{\alpha\in A}\lambda_\alpha\delta(x-x_\alpha) + W(x)$ with Δ the Laplacian in $L^2(\mathbb{R}^d, dx)$, $d = 1, 2, 3$ $(x_\alpha, \alpha\in A)$ a discrete (finite or countable) set of fixed "sources" and $W(x)$ a sum of one-and-two-body "nice" potentials (including Coulomb ones) have been defined.

(b) Compute the resolvent,[1–19] the spectrum (eigenvalues, resonances),[1–5,8,9,14–16,20–22,29] and the scattering quantities[1–11,14,17–19] of such models.

(c) Prove that the point interactions give the leading terms in low energy expansions for given "nice" potentials and find analytic expansions in a low energy parameter "around point interactions," for the resolvent,[1–9] the eigenvalues and resonances,[1–4,8,9,28] and scattering quantities.[1–9]

In the present paper we shall give yet another contribution to point (b), namely we shall compute the scattering by impurities in a three-dimensional model of a solid with point interactions (a three-dimensional version of the Kronig–Penney model).

Besides Einstein–Debye's harmonic crystal, the only exactly solvable model for the motion of a quantum mechanical particle in a d-dimensional ($d\leqslant 3$) crystal is the one where the crystal consists of fixed centers which act as sources of point interactions. This model has been defined mathematically and studied originally (starting from the construction of a resolvent for finitely many points in Ref. 12) in Refs. 1, 8, 9, 15, and 25.

In particular the resolvent was given, and proven to be the low energy limit of the one for a crystal with nice potentials. In Ref. 15, scattering quantities were also computed, in the case of an infinite straight polymer and in the case of a monomolecular layer. The present paper is concerned with a quantum mechanical particle (electron or neutron or, alternatively, a scalar electromagnetic wave or acoustic wave) moving in a crystal consisting as above of fixed centers creating point interactions and having a discrete set of impurities, given by point interactions centered at points not coinciding with any of the crystal sizes.

The scattering quantities (wave and scattering operators) for the scattering by impurities, where the crystal interactions are bounded periodic and the impurities are in suitable L^p-spaces, have been studied in Refs. 30–44. In our case, neither the crystal interactions nor the impurities interactions belong to these classes. Yet from the approximation results of point interactions by smooth interactions mentioned above,[1–9] we can define such quantities also in our model of scattering by impurities, the main advantage of our model being then that all quantities can be computed explicitly.

In Secs. 2 and 3 we recall the results on the Hamiltonian for crystals with point interactions.

In Sec. 4 we study the scattering quantities describing the scattering by impurities on a crystal with point interactions. We study both the scattering between given Bloch

 0022-2488/84/051327-08$02.50

waves and the one between given plane waves.

2. THE OFF-SHELL SCATTERING MATRIX FOR THE POINT INTERACTIONS

As discussed in the introduction, the Schrödinger operator for a particle moving in the potential given by a finite or discrete set of point interactions has been investigated in several publications.[1-28] Such an operator is given formally by

$$H = H_0 - \sum_{a\in A} \lambda_{\alpha_a} \delta(x-a), \tag{2.1}$$

where $H_0 = -\Delta$, Δ being the Laplacian in $L^2(\mathbb{R}^3)$. A is some finite or discrete subset of $\mathbb{R}^3$, δ is the Dirac's function, and λ_{α_a} are suitable coefficients. As discussed in Refs. 12 and 13, in order to give a rigorous sense to (2.1) as an Hamiltonian different from H_0, one must choose the λ_{α_a} as infinitesimal quantities in the sense of nonstandard analysis, depending on real parameters α_a. H can also be described in terms of standard analysis, as the self-adjoint operator in $L^2(\mathbb{R}^3)$ having as resolvent kernel

$$\begin{aligned}(H-E)^{-1}(x,y) &= G_E(x-y) + \sum_{a,b\in A} [(\alpha_a - \sqrt{-E}/4\pi)\delta_{ab} \\ &\quad - \tilde{G}_E(a-b)]^{-1}_{a,b} G_E(x-a) G_E(y-b),\end{aligned} \tag{2.2}$$

where

$$\begin{aligned}G_E(x-y) &\equiv (-\Delta - E)^{-1}(x,y) \\ &= (4\pi|x-y|)^{-1}\exp(-\sqrt{-E}\,|x-y|).\end{aligned}$$

$\delta_{ab} = 1$ if $a=b$, $\delta_{ab}=0$ if $a\neq b$, while $\tilde{G}_E(x-y) \equiv G_E(x-y)$ if $x\neq y$, $\tilde{G}_E(x-y)=0$ if $x=y$. $[\ ]^{-1}_{ab}$ is the a,b matrix element of the inverse of the operator $[\ ]$ in $l^2(A)$, with a, b matrix elements

$$(\alpha_a - \sqrt{-E}/4\pi)\delta_{ab} - \tilde{G}_E(a-b).$$

α_a are lower bounded, uniformly in a and H is lower bounded.

In Refs. 1, 2, 4, and 9 it was proven that H is the limit in the strong resolvent sense of the operators H_ϵ as $\epsilon \to 0$, where

$$H_\epsilon \equiv H_0 + \sum_{a\in A} \frac{\lambda_a(\epsilon)}{\epsilon^2} V_a\left(\frac{1}{\epsilon}(x-a)\right), \tag{2.3}$$

λ_a being smooth functions in a neighborhood of the origin, with $\lambda_a(0)=1$, and V_a being potentials such that $V_a(x)$ has a zero energy resonance, in the sense of Refs. 1, 2, 4, and 9. The relation between V_a, λ_a in (2.3) and α_a in (2.2) is such that α_a only depends on V_a and the derivative $\lambda'_a(0)$ of λ_a at the origin, see Refs. 1, 2, 4, and 9 for details. In this paper we shall need the resolvent formula corresponding to (2.2) for the case where H is given by (2.1) but with H_0 replaced by an operator different from $-\Delta$. This more general formula was derived in Ref. 14, so that from that reference we have the following.

Theorem 2.1: Let A be a subset of $\mathbb{R}^3$ without finite accumulation points. Let H_0 be a self-adjoint positive operator on $L^2(\mathbb{R}^3)$, with resolvent kernel

$$K_E(x,y) \equiv (H_0 - E)^{-1}(x,y),$$

such that for $a\neq b$, $K_E(a,b)$ is finite and continuous in a neighborhood of a and b, $K'_E(a,a)$ is finite and continuous in a neighborhood of (a,a) (K'_E being the derivative of K_E), and one has $K_E(a,\cdot)\in L^2(\mathbb{R}^3)$ (a, b run over the points of A).

Let α_a be bounded from below, uniformly for $a\in A$. Then the Schrödinger operator (2.1) is well defined as the self-adjoint operator in $L^2(\mathbb{R}^3)$ with resolvent kernel

$$\begin{aligned}(H-E)^{-1}(x,y) &= K_E(x,y) + \sum_{a,b\in A}\left[\left(\alpha_a - \int_{-1}^{E} K'_{\tilde{E}}(a,a)d\tilde{E}\right)\delta_{ab}\right. \\ &\quad \left. - \tilde{K}_E(a,b)\right]^{-1}_{a,b} K_E(a,x)K_E(b,y),\end{aligned}$$

where $\tilde{K}_E(a,b) \equiv K_E(a,b)$ if $a\neq b$, and $\tilde{K}_E(a,b)\equiv 0$ if $a=b$. ∎

Remark. The relation between the coefficients λ_{α_a} in (2.1) and the quantities α_a in Theorem 2.1 is given formally by $\lambda_{\alpha_a}^{-1} = K_{-1}(a,a) + \alpha_a$. This shows in particular that if $K_{-1}(a,a)$ is not finite (which is the case if, e.g., $H_0 = -\Delta$) then λ_a is infinitesimal. See Refs. 12 and 13.

In the following we shall first recall a few notions of scattering theory for Schrödinger operators of the form

$$H = H_0 + V \tag{2.4}$$

with suitable H_0, having a nontrivial absolutely continuous part, and potentials V, say bounded. Let $R(E) \equiv (H-E)^{-1}$ be the resolvent of H. In scattering theory one defines the corresponding T operator by

$$T(E) \equiv V - VR(E)V, \tag{2.5}$$

see, e.g., Refs. 45–47.

We remark that this is equivalent to

$$T(E) = (H_0 - E)(R_0(E) - R(E))(H_0 - E), \tag{2.6}$$

where $R_0(E) = (H_0 - E)^{-1}$, as seen by an iteration of the resolvent formula (first for $|\mathrm{Im}\, E|$ sufficiently large, and then by analytic continuation).

Associated with the absolutely continuous part H_0^{ac} of H_0 there is a space P which supports the spectral decomposition of H_0^{ac}. Let $\psi_{n,p}(x)$ be a complete set of generalized eigenfunctions for the absolutely continuous part of the spectrum of H_0, i.e., we have

$$H_0\psi_{n,p} = E_n(p)\psi_{n,p}, \tag{2.7}$$

for any $p\in P$, $n\in\mathbb{N}$, with the orthonormality relations

$$\int_{\mathbb{R}^3} \bar{\psi}_{n,p}(x)\psi_{m,q}(x)dx = \delta_{nm}\delta(p-q)$$

and the completeness relation for the absolutely continuous part, namely the property that

$$\sum_n \int_P dp\, \psi_{n,p}(x)\bar{\psi}_{n,p}(y)$$

is the projection onto the absolutely continuous part of H_0.

The off-shell scattering matrix $\tilde{S}_E$ is defined in general, given the T-operator $T(E)$ and generalized eigenfunctions $\psi_{n,p}$ for the absolutely continuous part of H_0, by

$$\tilde{S}_E(p,n;q,m) \equiv (\psi_{n,p}, T(E)\psi_{m,q}), \tag{2.8}$$

where $(\ ,\ )$ is the $L^2(\mathbb{R}^3)$ scalar product.

By definition the on-shell scattering matrix is given by

$$\tilde{S}(p,n;q,m)\equiv\tilde{S}_{E_n(p)}(p,n;,q,m)\delta(E_n(p)-E_m(q)). \tag{2.9}$$

The scattering operator S is by definition an operator on the subspace $P_0L^2(\mathbb{R}^3)$ of $L^2(\mathbb{R}^3)$ on which H_0 is absolutely continuous. (P_0 is the projection onto the absolutely continuous subspace of H_0, see Refs. 45–47).

The kernel of S is given by

$$S(p,n;q,m)\equiv\delta(p-q)\delta_{n,m}+\tilde{S}(p,n;q,m). \tag{2.10}$$

The relation between the scattering matrix (2.8), the wave operators,

$$W_{\pm}=s-\lim_{t\to\pm\infty}e^{-itH}e^{itH_0}P_0, \tag{2.11}$$

and the S matrix

$$S=W_-^*W_+ \tag{2.12}$$

is well known, see, e.g., Refs. 45–47 (where sufficient conditions for all quantities to exist and all relations to hold are given).

In Refs. 2 and 18 it is shown that (2.6) gives a suitable definition of the off-shell T matrix also for point interactions, in the one-center problem. Using the results of Refs. 1, 4, and 9 one can extend this to the case of a discrete number of centers, arriving at the following result for the point interactions of Theorem 2.1:

Theorem 2.2: Under the same assumptions as in Theorem 2.1 the off-shell scattering matrix is given for any complex E in the complement of the spectrum of H_0 by

$$\tilde{S}_E(p,n;q,m)$$
$$=\sum_{a,b\in A}\psi_{n,p}(a)\bar{\psi}_{m,q}(b)\times\left[\left(\alpha_a-\int_{-1}^{E}K_{\tilde{E}}(a,a)d\tilde{E}\right)\delta_{ab}-\tilde{K}_E(a,b)\right]^{-1}_{a,b},$$

where $\psi_{n,p}(x)$ is a complete orthonormal system of generalized eigenfunctions corresponding to the absolutely continuous part of H_0.

3. SCATTERING BY POINT INTERACTIONS

Let us now consider an operator H of the form (2.1) with A finite, i.e.,

$$H=-\Delta-\sum_{a\in A}\lambda_{\alpha_a}\delta(x-a). \tag{3.1}$$

The kernel $(H-E)^{-1}(x,y)$ of its resolvent is then given by (2.2).

From Theorem 2.2, observing that here $(2\pi)^{-3/2}e^{-ipa}$, $a\in\mathbb{R}^3$ play the role of the ψ_n, we then have that the off-shell scattering matrix is given by

$$\tilde{S}_E(p,q)=(2\pi)^{-3}\sum_{a,b\in A}e^{ipa-iqb}\times\left[(\alpha_a-\sqrt{-E}/4\pi)\delta_{ab}-\tilde{G}_E(a-b)\right]^{-1}_{a,b}, \tag{3.2}$$

where

$$\tilde{G}_E(a-b)\equiv\begin{cases}G_E(a-b) & \text{if } a-b\neq 0\\ 0 & \text{if } a-b=0\end{cases}$$

with

$$G_E(a-b)\equiv(-\Delta-E)^{-1}(a,b)=(4\pi|a-b|)^{-1}\exp(-\sqrt{-E}\,|a-b|).$$

In this case we see that the on-shell scattering matrix $\tilde{S}(p,q)$ exists and is given by

$$\tilde{S}(p,q)=(2\pi)^{-3}\sum_{a,b\in A}e^{i(pa-qb)}\left[\left(\alpha_a-\frac{i|p|}{4\pi}\right)\delta_{ab}-\tilde{G}_{p^2}(a-b)\right]^{-1}_{a,b}\delta(p^2-q^2). \tag{3.3}$$

Hence we have the following:

Theorem 3.1: The on-shell scattering matrix for the Hamiltonian in $L^2(\mathbb{R}^3)$ obtained by perturbing $-\Delta$ by a finite number of point scatterers a of strengths α_a is given by

$$\tilde{S}(p,q)=(2\pi)^{-3}\sum_{a,b\in A}e^{i(pa-qb)}\left[\left(\alpha_a-\frac{i|p|}{4\pi}\right)\delta_{ab}-\tilde{G}_{p^2}(a-b)\right]^{-1}_{a,b}\delta(p^2-q^2),$$

while the corresponding off-shell scattering matrix is given by

$$\tilde{S}_E(p,q)=(2\pi)^{-3}\sum_{a,b\in A}e^{i(pa-qb)}\left[\left(\alpha_a-\frac{\sqrt{-E}}{4\pi}\right)\delta_{ab}-\tilde{G}_E(a-b)\right]^{-1}_{a,b}.$$

Remark 1: In the case of a one-point scatterer at the origin we have simply

$$\tilde{S}(p,q)=(2\pi)^{-3}(\alpha-i|p|/4\pi)^{-1}\delta(p^2-q^2)$$

and

$$\tilde{S}_E(p,q)=(2\pi)^{-3}(\alpha-\sqrt{-E}/4\pi)^{-1}.$$

Remark 2: In the case of two-point scatterers at a resp. b we have with $s(p,q;a,b)\equiv\exp[i(pa-qb)]+\exp[i(qa-pb)]$:

$$\begin{aligned}\tilde{S}(p,q)=(2\pi)^{-3}&[(\alpha_a-i|p|/4\pi)(\alpha_b-i|p|/4\pi)-G_{p^2}(a-b)^2]^{-1}[e^{i(p-q)b}(\alpha_a-i|p|/4\pi)\\&+e^{i(p-q)a}(\alpha_b-i|p|/4\pi)+G_{p^2}(a-b)s(p,q;a,b)]\delta(p^2-q^2)\end{aligned}$$

and

$$\begin{aligned}\tilde{S}_E(p,q)=(2\pi)^{-3}&[(\alpha_a-\sqrt{-E}/4\pi)(\alpha_b-\sqrt{-E}/4\pi)-G_E(a-b)^2]^{-1}[e^{i(p-q)b}(\alpha_a-\sqrt{-E}/4\pi)\\&+e^{i(p-q)a}(\alpha_b-\sqrt{-E}/4\pi)+G_E(a-b)s(p,q;a,b)],\end{aligned}$$

where we recall that

$$G_E(x-y)=(4\pi|x-y|)^{-1}\exp(-\sqrt{-E}\,|x-y|).$$

In a similar way, using Theorems 2.1 and 2.2 one can compute the off-shell scattering matrix for $-\Delta$ perturbed by an infinite number of point scatterers (without accumulation points), obtaining the following:

Theorem 3.2: Let A be any subset of $\mathbb{R}^3$ without finite accumulation points and let α_a uniformly lower bounded

real-valued functions for $a\in A$. Then

$$H = -\Delta - \sum_{a\in A} \lambda_{\alpha_a}\delta(x-a)$$

exists for a suitable choice of infinitesimal λ_{α_a}, depending on α_a, as a self-adjoint lower bounded operator in $L^2(\mathbb{R}^3, dx)$, given by its resolvent kernel (2.2). The corresponding off-shell scattering matrix also exists and is given by

$$\tilde{S}_E(p,q) = (2\pi)^{-3}\sum_{a,b\in A} e^{i(pa-qb)} \times \left[(\alpha_a - \sqrt{-E}/4\pi)\delta_{ab} - \tilde{G}_E(a-b)\right]^{-1}_{a,b}. \quad \blacksquare$$

Let us now consider the particularly interesting case where both the set A of points creating the point interaction and the strengths α_a of the point interactions are invariant under a discrete subgroup Λ^ν of $\mathbb{R}^3$ with ν independent generators, where ν can take any of the values 1, 2, 3. Decomposing $\mathbb{R}^3$ as $\mathbb{R}^\nu\times\mathbb{R}^{3-\nu}$ we can look upon Λ^ν as a subgroup of $\mathbb{R}^\nu$. Since A is invariant under Λ^ν we have that A is of the form $A = \Lambda^\nu + C$, with C a discrete subset of $\mathbb{R}^3$, in the sense that every point $a\in A$ can be written as $a = \lambda + c$, with $\lambda\in\Lambda^\nu$, $c\in C$. Moreover we have $\alpha_a = \alpha_c$, because of the Λ^ν invariance of the strengths, hence with α_c independent of λ. In this case the formula for the resolvent in Theorem 2.1 becomes

$$(H-E)^{-1}(x,y) = G_E(x-y) + \sum_{c,c'}\sum_{\lambda,\lambda'}\left[\left(\alpha_c - \frac{\sqrt{-E}}{4\pi}\right)\delta_{cc'},\delta_{\lambda\lambda'} - \tilde{G}_E(c-c'+\lambda-\lambda')\right]^{-1}_{\substack{c,c'\\ \lambda,\lambda'}} \times G_E(c+\lambda-x)G_E(c'+\lambda'-y). \tag{3.4}$$

For the corresponding off-shell scattering matrix we get

$$\tilde{S}_E(p,q) = (2\pi)^{-3}\sum_{c,c'}\sum_{\lambda,\lambda'}\exp[i(pc-qc')] \times\exp[i(p\lambda-q\lambda')]\left[(\alpha_c - \sqrt{-E}/4\pi)\delta_{cc'}\delta_{\lambda\lambda'} - \tilde{G}_E(c-c'+\lambda-\lambda')\right]^{-1}_{\substack{c,c'\\ \lambda,\lambda'}}. \tag{3.5}$$

We can simplify (3.4) and (3.5) by using the Fourier transformation with respect to the discrete abelian group Λ^ν. Define for $k\in\mathbb{R}^\nu$, $x\notin\Lambda^\nu$:

$$g_E(x,k) = \sum_{\lambda'\in\Lambda^\nu} G_E(x-\lambda')e^{ik\lambda'}, \tag{3.6}$$

where $k\lambda$ is the inner product in $\mathbb{R}^\nu$, and define for $\lambda\in\Lambda^\nu$

$$g_E(\lambda,k) = \sum_{\substack{\lambda'\neq\lambda\\ \lambda'\in\Lambda^\nu}} G_E(\lambda-\lambda')e^{ik\lambda'} + \frac{\sqrt{-E}}{4\pi}. \tag{3.7}$$

Then $g_E(x,k)$ is periodic in k with periodicity Γ_ν, where Γ_ν is the lattice orthogonal to Λ^ν, i.e., if $\Lambda^\nu = \{\Sigma_{i=1}^\nu m_i a_i, m_i\in\mathbb{Z}\}$ where the fixed vectors $a_i\in\mathbb{R}^\nu$ are the generators of Λ^ν, then $\Gamma_\nu = \{\Sigma_{i=1}^\nu n_i b_i, n_i\in\mathbb{Z}\}$, with b_i satisfying

$$a_i b_j = 2\pi\delta_{ij}. \tag{3.8}$$

The periodicity of g_E is expressed by

$$g_E(x,k+\gamma) = g_E(x,k), \quad \forall\gamma\in\Gamma_\nu. \tag{3.9}$$

Moreover $x\to g_E(x,k)$ satisfies the k-boundary conditions, i.e., for $\gamma\in\Gamma_\nu$, $\lambda\in\Lambda^\nu$, $x\notin\Lambda^\nu$ we have

$$g_E(x+\lambda,k) = e^{-ik\lambda}g_E(x,k). \tag{3.10}$$

Observing that $e^{ik\lambda}$ are all the characters of Λ^ν we get a natural identification of the dual group $\hat{\Lambda}^\nu$ of Λ^ν with $\mathbb{R}^\nu/\Gamma_\nu$. Using the Plancherel formula for Λ^ν the summations over Λ^ν in (3.4) and (3.5) may be rewritten as integrals over the dual group $\mathbb{R}^\nu/\Gamma_\nu$. Since H is invariant under translations by λ for $\lambda\in\Lambda^\nu$ we have the direct integral decomposition

$$H = \int_{\mathbb{R}^\nu/\Gamma_\nu} H(k)dk, \tag{3.11}$$

where dk is the Lebesgue measure on the torus $\mathbb{R}^\nu/\Gamma_\nu$, and $H(k)$ is the reduced Hamiltonian. From (3.4) we get then

$$[H(k)-E]^{-1}(x,y) = g_E(x-y,k) + \sum_{c,c'}\left[\alpha_c\delta_{cc'} - g_E(c-c',k)\right]^{-1}_{c,c'} \times g_E(x-c,k)g_E(y-c',k). \tag{3.12}$$

Let $\tilde{S}^k_E(p,q)$ be the corresponding reduced off-shell scattering matrix, so that

$$\tilde{S}_E(p,q) = \int_{\mathbb{R}^\nu/\Gamma_\nu}\tilde{S}^k_E(p,q)dk. \tag{3.13}$$

From (3.6) we have then

$$\tilde{S}^k_E(p,q) = (2\pi)^{-3}\sum_{c,c'} e^{i(pc-qc')} \times\left[\alpha_c\delta_{cc'} - g_E(c-c',k)\right]^{-1}_{c,c'} \times\sum_{\gamma,\gamma'\in\Gamma_\nu}\delta_\nu(p_\nu-k-\gamma)\delta_\nu(q_\nu-k-\gamma'), \tag{3.14}$$

where δ_ν is the Dirac δ-function in $\mathbb{R}^\nu$ and p_ν, q_ν are the projections of p and q onto $\mathbb{R}^\nu$ in the decomposition $\mathbb{R}^3 = \mathbb{R}^\nu\times\mathbb{R}^{3-\nu}$. Hence we have the following:

Theorem 3.3: Let H be the Hamiltonian for a point interaction in $\mathbb{R}^3$ invariant under a discrete subgroup Λ^ν of $\mathbb{R}^\nu$, for some fixed $\nu = 1, 2, 3$. The resolvent kernel is given by

$$(H-E)^{-1}(x,y) = G_E(x-y) + \sum_{c,c'\in C}\sum_{\lambda,\lambda'\in\Lambda^\nu}\left[\left(\alpha_c - \frac{\sqrt{-E}}{4\pi}\right)\delta_{c,c'}\delta_{\lambda,\lambda'} - \tilde{G}_E(c-c'+\lambda-\lambda')\right]^{-1}_{\substack{c,c'\\ \lambda,\lambda'}} \times G_E(c+\lambda-x)G_E(c'+\lambda'-y),$$

the support of the point interactions being $\Lambda^\nu + C$, with C a fixed discrete subset of $\mathbb{R}^3$.

We have the direct integral decomposition $H = \int_{\mathbb{R}^\nu/\Gamma_\nu} H(k)dk$, where Γ_ν is the orthogonal lattice to Λ^ν in $\mathbb{R}^\nu$, defined by (3.8). The resolvent kernel for the reduced Hamiltonian $H(k)$ is given by

$$[H(k)-E]^{-1}(x,y) = g_E(x-y,k)$$

$$+\sum_{c,c'\in C}[\alpha_c\delta_{c,c'}-g_E(c-c',k)]^{-1}_{c,c'}$$
$$\times g_E(x-c,k)g_E(y-c',k),$$

with g_E defined by (3.6) and (3.7).

The corresponding reduced off-shell scattering matrix is given by

$$\tilde{S}^k_E(p,q)=(2\pi)^{-3}\sum_{c,c'\in C}e^{i(pc-qc')}$$
$$\times[\alpha_c\delta_{c,c'}-g_E(c-c',k)]^{-1}_{c,c'}$$
$$\times\sum_{\gamma,\gamma'\in\Gamma_\nu}\delta_\nu(p_\nu-k-\gamma)\delta_\nu(q_\nu-k-\gamma'),$$

where p_ν, q_ν are the projections of p resp. q onto $\mathbb{R}^\nu$ in the decomposition $\mathbb{R}^3=\mathbb{R}^\nu\times\mathbb{R}^{3-\nu}$, and δ_ν is the δ-distribution in $\mathbb{R}^\nu$.

It is easy to see that for $\nu=1,2$ the on-shell scattering matrix exists as a limit of the off-shell scattering matrix and we have

Corollary 3.4: Under the assumptions of Theorem 3.3 with $\nu=1$ or $\nu=2$ the reduced on-shell scattering matrix exists and is given by

$$\tilde{S}^k(p,q)=(2\pi)^{-3}\sum_{c,c'\in C}e^{i(pc-qc')}[\alpha_c\delta_{c,c'}$$
$$-g_{p^2}(c-c',k)]^{-1}_{c,c'}\sum_{\gamma,\gamma'\in\Gamma_\nu}\delta_\nu(p_\nu-k-\gamma)$$
$$\times\delta_\nu(q_\nu-k-\gamma')\delta(p^2-q^2).$$

The on-shell scattering matrix is given by

$$\tilde{S}(p,q)=\int_{\mathbb{R}^\nu/\Gamma_\nu}\tilde{S}^k(p,q)dk.$$

Rem.: For $\nu=3$ the on-shell scattering matrix does not exist.

4. IMPURITY SCATTERING IN SIMPLE POINT CRYSTALS

In this section we shall consider perturbations of Hamiltonians given by point interactions. More precisely we consider a situation like in (2.1) but with H_0 replaced by the Hamiltonian of a periodic point interaction as given in Theorem 3.3, with $\nu=3$ and C consisting of a simple point, taken at the origin. We write Λ instead of Λ^ν. Let H_0 be the operator with resolvent kernel

$$(H_0-E)^{-1}(x,y)$$
$$=G_E(x-y)+\sum_{\lambda,\lambda'\in\Lambda}\left[\left(\alpha-\frac{\sqrt{-E}}{4\pi}\right)\delta_{\lambda,\lambda'}\right.$$
$$\left.-\tilde{G}_E(\lambda-\lambda')\right]^{-1}_{\lambda,\lambda'}G_E(\lambda-x)G_E(\lambda'-y). \tag{4.1}$$

Using the direct integral decomposition $H_0=\int_{\mathbb{R}^3/\Gamma}H(k)dk$, where Γ stands for Γ_ν, i.e., the orthogonal lattice to $\Lambda=\Lambda^\nu$, and Theorem 3.3 we have also

$$(H_0-E)^{-1}(x,y)=\int_{\mathbb{R}^3/\Gamma}[H_0(k)-E]^{-1}(x,y)dk,$$

with

$$[H_0(k)-E]^{-1}(x,y)$$
$$=g_E(x-y,k)+[\alpha-g_E(0,k)]^{-1}g_E(x,k)g_E(y,k). \tag{4.2}$$

In this case we have from (3.7), using the Poisson summation formula, as in Ref. 14 $|\Lambda|$, with $|\hat{\Lambda}|$ the volume of the Brillouin zone:

$$g_E(0,k)=\int_0^E\frac{dg_{\tilde{E}}(0,k)}{d\tilde{E}}d\tilde{E}$$
$$=|\hat{\Lambda}|(2\pi)^{-3}\int_0^E\sum_{\gamma\in\Gamma}[|k-\gamma|^2-\tilde{E}]^{-2}d\tilde{E}. \tag{4.3}$$

This shows that $g_E(0,k)$ is, for any fixed k, monotone strictly increasing in E with a first-order pole of residue $(2\pi)^{-3}$ at each point $E=|k-\gamma|^2$, $\gamma\in\Gamma$. Hence there is exactly one solution E of the equation

$$\alpha-g_E(0,k)=0 \tag{4.4}$$

in each of a sequence of bounded open intervals

$$I^k_n\equiv(A^k_n,B^k_n),\ A^k_n<B^k_n$$
$$<A^k_{n+1}<B^k_{n+1}<\ldots,\ n=1,2,\ldots,$$

such that

$$\mathbb{R}-\{|k-\gamma|^2,\ \gamma\in\Gamma\}=I^k_0\cup\left(\bigcup_{n=1}^{\infty}I^k_n\right),$$

with I^k_0 of the form $(-\infty,x^k_0)$, for some $x^k_0\in\mathbb{R}$. Moreover (4.4) has also a unique solution in I^k_0, since $g_E(0,k)\to-\infty$ as $E\to-\infty$. Let us denote by $E_\gamma(k)$ the solutions of (4.4) s.t. $E_\gamma(k)\in I^k_n$, with $I^k_n=(A^k_n,B^k_n)$, $B^k_n=|k-\gamma|^2$. In addition to these values $E_\gamma(k)$, in order to study the eigenvalues of $H_0(k)$, we should also look at the poles of $g_E(0,k)$. In the case where there is a number $m>1$ of points, $\gamma_1,\ldots,\gamma_m\in\Gamma$ s.t. $|k-\gamma_1|^2=\cdots=|k-\gamma_m|^2$, then we call $E_{\gamma_1}(k)$ the solution of (4.4) in the interval I^k_n with $B^k_n=|k-\gamma_1|^2$, and we set $E_{\gamma_2}(k)\equiv E_{\gamma_3}(k)\equiv\ldots\equiv E_{\gamma_n}(k)\equiv|k-\gamma_1|^2$.

Then $(E_\gamma(k),\gamma\in\Gamma)$ are the eigenvalues for the reduced Hamiltonian $H_0(k)$ with correct multiplicities, the bottom of the spectrum of $H_0(k)$ being the unique $E_\gamma(k)$ with $E_\gamma(k)\in I^k_0$ and the spectrum of $H_0(k)$ being pure point spectrum, see Refs. 30 and 14 [where also additional information on the dependence of $E_\gamma(k)$ on k is given].

We shall call H_0 *the Hamiltonian for a simple point crystal.*

We shall now study the perturbation of H_0 by point *impurities* located at some finite subset A of $\mathbb{R}^3$. Let us call H the Hamiltonian of the perturbed system, i.e., H is obtained from $-\Delta$ by first inserting point interactions at the points of $C+\Lambda$ obtaining H_0 and then inserting additional point interactions with support on the set A of impurities. By Sec. 2 we know that the resolvent kernel of H is given by

$$(H-E)^{-1}(x,y)$$
$$=(H_0-E)^{-1}(x,y)+\sum_{a,b\in A}\left[\left(\alpha_a-\frac{\sqrt{-E}}{4\pi}\right)\delta_{ab}\right.$$
$$\left.-K_E(a,b)\right]^{-1}_{a,b}K_E(a,x)K_E(b,y), \tag{4.5}$$

where $(H_0-E)^{-1}(x,y)$ is given by (4.1) and

$$K_E(a,b) \equiv \tilde{G}_E(a-b) + \int_{\mathbb{R}^3/\Gamma} [\alpha - g_E(0,k)]^{-1} \times g_E(a,k) g_E(b,k) dk, \tag{4.6}$$

with dk the Lebesgue measure on $\mathbb{R}^3/\Gamma$ and

$$\tilde{G}_E(a-b) \equiv G_E(a-b) \text{ if } a-b \neq 0$$

and

$$\tilde{G}_E(a-b) \equiv 0 \text{ if } a-b=0.$$

Let $E_\gamma(k)$ be the solution of $\alpha = g_E(0,k)$ in the interval $I_n^k = (A_n^k, B_n^k)$, with $B_n^k = |\gamma - k|^2$. Then $E_\gamma(k)$ is a simple eigenvalue of $H(k)$, i.e., a simple pole of the resolvent kernel $[H_0(k) - E]^{-1}(x,y)$ as a function of E. $g_E(0,k)$ is an analytic function of E in a neighborhood of $E_{\gamma(k)}$, thus we have

$$g_E(0,k) = \alpha + (E - E_\gamma(k)) g'_{E_\gamma(k)}(0,k) + O((E - E_\gamma(k))^2), \tag{4.7}$$

with

$$g'_{E_\gamma(k)} \equiv \frac{d}{dE} g_E(0,k)|_{E=E_\gamma(k)} = (2\pi)^{-3} |\hat{\Lambda}| \sum_{\gamma \in \Gamma} \frac{1}{[|k-\gamma|^2 - E]^2}. \tag{4.7}$$

From (4.2) we have then that the residue of $[H_0(k) - E]^{-1}(x, y)$ at the pole $E_\gamma(k)$ is

$$[g'_{E_\gamma(k)}(0,k)]^{-1} g_{E_\gamma(k)}(x,k) g_{E_\gamma(k)}(y,k). \tag{4.8}$$

Let $\psi_\gamma(x,k)$ be the eigenfunction of $H_0(k)$ to the eigenvalue $E_\gamma(k)$, normalized such that

$$\int_{\mathbb{R}^3/\Lambda} |\psi_\gamma(x,u)|^2 dx = 1. \tag{4.9}$$

$\psi_\gamma(x,k)$ is called a Bloch wave.

We must have

$$[H_0(k) - E]^{-1}(x,y) = \sum_{\gamma \in \Gamma} [E_\gamma(k) - E]^{-1} \psi_\gamma(x,k) \psi_\gamma(x,k) \tag{4.10}$$

and comparing the residue at the pole $E_\gamma(k)$ from (4.10) with (4.7) we get

$$\psi_\gamma(x,k) = [g'_{E_\gamma(k)}(0,k)]^{-1/2} g_{E_\gamma(k)}(x,k). \tag{4.11}$$

$g'_{E_\gamma(k)}$ is given by (4.7) or, equivalently, using Fourier transforms:

$$g'_{E_\gamma(k)}(0,k) = \sum_{\lambda \in \Lambda} \frac{1}{4\pi\sqrt{-E}} e^{-\sqrt{-E}|\lambda|} e^{ik\lambda}. \tag{4.12}$$

Inserting this information into Theorem 2.2 we get for the corresponding off-shell scattering matrix at energy E, for scattering off a finite set A of impurities of a particle moving in a simple point crystal, with incoming resp. outcoming waves described at fixed k resp. k' by the eigenfunctions $\psi_\gamma(x,k)$ resp. $\psi_{\gamma'}(x,k')$ of $H_0(k)$:

$$\tilde{S}_E(k,\gamma;k',\gamma') = [g'_{E_\gamma(k)}(0,k) g'_{E_{\gamma'}(k')}(0,k')]^{-1/2} \times \sum_{a,b \in A} g_{E_\gamma}(a,k) g_{E_{\gamma'}(k')}(b,k') \times [(\alpha_a - \sqrt{-E}/4\pi)\delta_{ab} - K_E(a,b)]^{-1}. \tag{4.13}$$

To compute the corresponding on-shell scattering matrix we observe that $E_\gamma(k) = E_{\gamma'}(k)$ implies that either $\gamma = \gamma'$ or $|\gamma - k|^2 = |\gamma' - k|^2$ or both. From this and (2.8) we compute the corresponding on shell scattering matrix to be

$$\tilde{S}(k,\gamma;k',\gamma') = [g'_{E_\gamma(k)}(0,k)]^{-1} \delta_{\gamma,\gamma'} \delta(E_\gamma(k) - E_\gamma(k')) \times \sum_{a,b \in A} g_{E_\gamma(k)}(a,k) g_{E_{\gamma(k')}}(b,k') \times \left[\left(\alpha_a - \frac{\sqrt{-E_\gamma(k)}}{4\pi}\right)\delta_{ab} - K_{E_\gamma(k)}(a,b)\right]^{-1}_{a,b}.$$

We formulate these results in the following:

Theorem 4.1: Let H_0 be the Hamiltonian for a simple point crystal, i.e., its resolvent kernel is given by

$$(H_0 - E)^{-1}(x,y) = G_E(x-y) + \sum_{\lambda,\lambda' \in \Lambda} \left[\left(\alpha - \frac{\sqrt{-E}}{4\pi}\right)\delta_{\lambda\lambda'} - \tilde{G}_E(\lambda - \lambda')\right]^{-1}_{\lambda,\lambda'} G_E(\lambda - x) G_E(\lambda' - y),$$

where Λ is a discrete subgroup of $\mathbb{R}^3$ with three independent generators. Let H be the Hamiltonian for the point crystal perturbed by a finite number of point impurities at $A \subset \mathbb{R}^3$, disjoint from Λ, of strength α_a, $a \in A$. Then the resolvent kernel for H is given by

$$(H - E)^{-1}(x,y) = (H_0 - E)^{-1}(x,y) + \sum_{a,b \in A} \left[\left(\alpha_a - \frac{\sqrt{-E}}{4\pi}\right)\delta_{ab} - K_E(a,b)\right]^{-1}_{a,b} K_E(a,x) K_E(b,y),$$

where K_E is given in (4.6).

The corresponding impurity on-shell scattering matrix describing scattering from impurity from the channel described by a Bloch wave $\psi_\gamma(x,k)$ given by the (quasi) momentum k and the energy $E_\gamma(k)$ of the simple point crystal into the channel described by a Bloch $\psi_{\gamma'}(x,k')$, given by momentum k' and energy $E_{\gamma'}(k')$, is given by

$$\tilde{S}(k,\gamma;k',\gamma') = \delta_{\gamma,\gamma'} \delta(E_\gamma(k) - E_\gamma(k')) [g_{E_\gamma(k)}(0,k)]^{-1} \times \sum_{a,b \in A} g_{E_\gamma(k)}(a,k) g_{E_{\gamma'}(k')}(b,k') \times \left[\left(\alpha_a - \frac{\sqrt{-E_\gamma(a)}}{4\pi}\right)\delta_{ab} - K_{E_\gamma(k)}(a,b)\right]^{-1}_{a,b}. \quad \blacksquare$$

Remark: The scattering matrix $\tilde{S}_{E_\gamma(k)}(k,\gamma;k',\gamma')$ expresses the scattering in the simple point crystal of the Bloch wave $\psi_\gamma(x,k)$ in (4.11) into the Bloch wave $\psi_{\gamma'}(x,k')$, the scattering being caused by the finite set A of impurities. In an actual scattering experiment the crystal is of finite size and the incident and final waves are plane waves, rather than Bloch waves. (See, e.g., Ref. 48.) For this reason, as a first step towards the complicated realistic situation, we shall now give the expression of the scattering operator in plane

waves rather than Bloch waves.

The impurity scattering operator S corresponding to the impurity scattering matrix of Theorem 4.1 is given by

$$S = 1 + \sum_{\gamma,\gamma'} \int_{\mathbb{R}^3|\Gamma} \int_{\mathbb{R}^3|\Gamma} \tilde{S}(k, \gamma; k', \gamma') \times \psi_\gamma(\cdot, k)\psi_{\gamma'}(\cdot, k')dk\, dk', \tag{4.14}$$

where $\psi_\gamma(\cdot, k)$ is understood as the element in $L^2(\mathbb{R}^3, dx)$, for fixed k, γ, given by (4.9). The scattering amplitude $\hat{S}(p, q)$ for the scattering of an incoming plane wave $(2\pi)^{-3/2} e^{ipx}$ to an outgoing plane wave $(2\pi)^{-3/2} e^{iqx}$ is by definition

$$\hat{S}(p, q) = (2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} e^{i(px - qy)} S(x, y)dx\, dy, \tag{4.15}$$

$S(x, y)$ being the kernel of (4.14). Inserting in (4.14), (4.15) the expressions given by Theorem 4.1 we arrive at the following:

Theorem 4.2: The impurity scattering operator for the scattering of a particle moving asymptotically with Hamiltonian H_0 of a simple point crystal under the influence of a finite set of impurities, is given by (4.14) together with the formula for the Bloch-waves scattering amplitude $\tilde{S}$ given in Theorem 4.1. The plane waves scattering amplitude $\hat{S}(p, q)$ is defined as

$$\hat{S}(p, q) = (2\pi)^{-3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} e^{i(px - qy)} S(x, y)dx\, dy,$$

with $S(x, y)$ the kernel of the scattering operator in configuration space. The explicit expression of $\hat{S}(p, q)$ is

$$\hat{S}(p, q) = \delta(p - q) + (2\pi)^{-3} \sum_{\gamma} \delta(E_\gamma(p) - E_\gamma(q)) \times \left[g'_{E_\gamma(p)}(0, p) g'_{E_\gamma(p)}(0, q) \right]^{-1}$$

$$\times \left[p^2 - E_\gamma(p) \right]^{-1} \left[q^2 - E_\gamma(q) \right]^{-1} \sum_{a,b \in A} g_{E_\gamma(p)}(a, p) \times g_{E_\gamma(p)}(b, q) \left[\left(\alpha_a - \frac{\sqrt{-E_\gamma(p)}}{4\pi} \right) \delta_{a,b} - K_{E_\gamma(p)}(a, b) \right]^{-1}_{a,b}.$$

Note added in proof: For new developments see also H. Holden, R. Høegh-Krohn, M. Mebkhout (Marseille Preprint, in prep.) and S. Albeverio, F. Gesztesy, H. Holden, R. Høegh-Krohn (book in preparation).

ACKNOWLEDGMENTS

It is a pleasure to thank Prof. Dr. Fritz Gesztesy, Helge Holden, and Steinar Johannesen for interesting discussions. The hospitality of the Mathematics Departments of the Universities of Bochum and Oslo, of the Physics Department of Bielefeld, and of the Université d'Aix-Marseille II and the Centre de Physique Théorique, CNRS, Marseille–Luminy is gratefully acknowledged. This work has been partially supported by the Norwegian Research Council for Science and the Humanities and the Deutsche Forschungsgemeinschaft. We are grateful to Mrs. Mischke and Mrs. Richter for their skillful typing.

[1]S. Albeverio and R. Høegh-Krohn, "Point interactions as limits of short-range interactions," J. Operator Theory **6**, 313–339 (1981).
[2]S. Albeverio, F. Gesztesy, and R. Høegh-Krohn, "The low energy expansion in nonrelativistic scattering theory," Ann. Inst. H. Poincaré A **37**, 1–28 (1982).
[3]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and L. Streit, "Charged particles with short range interactions," Ann. Inst. H. Poincaré **38**, 263–293 (1983).
[4]H. Holden, R. Høegh-Krohn, and S. Johannesen, "The short range expansion," Oslo Preprint 1982, to appear in Adv. Appl. Math.
[5]S. Albeverio, D. Bollé, F. Gesztesy, R. Høegh-Krohn, and L. Streit, "Low energy parameters in non-relativistic scattering theory," Ann. Phys. (NY) **148**, 308–326 (1983).
[6]D. Bollé and S. F. J. Wilk, "Low energy sum rules for two-particle scattering," KUL-Preprint-TF-82/09, Univ. Leuven, 1982; J. Math. Phys. **24**, 1555–1564 (1983).
[7]S. Albeverio, L. S. Ferreira, F. Gesztesy, R. Høegh-Krohn, and L. Streit, "On the model dependence of Coulomb corrected scattering lengths," Bielefeld Preprint, 1982, Phys. Rev. C, Jan. 1984.
[8]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and W. Kirsch, "On point interactions in one dimension," Bochum Preprint, 1982, J. Operator Theory 1984.
[9]H. Holden, R. Høegh-Krohn, and S. Johannesen, "The short range expansion in solid state" Oslo Preprint 1983, to appear in Ann. Inst. H. Poincaré.
[10]L. E. Thomas, "Scattering from point interactions" in *Lecture Notes in Physics* **130** (Springer-Verlag, Berlin, 1980), pp. 159–162.
[11]G. Flamand, "Mathematical theory of non-relativistic two- and three-particle systems with point interactions," edited by F. Lurgat, in *Cargèse Lectures in Theoretical Physics* (Gordon and Breach, New York, 1967), pp. 247–287.
[12]S. Albeverio, J. E. Fenstad, and R. Høegh-Krohn, "Singular perturbations and non standard analysis," Trans. Am. Math. Soc. **252**, 275–295 (1979).
[13]S. Albeverio, J. E. Fenstad, R. Høegh-Krohn, and T. Lindstrøm, *Nonstandard analysis methods in probability theory and mathematical physics* (book in preparation).
[14]A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, "The one-particle theory of periodic point interactions," Commun. Math. Phys. **77**, 87–110 (1980).
[15]A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, "A class of explicitly soluble, local, and many-center Hamiltonians for one-particle quantum mechanics in two and three dimensions I," J. Math. Phys. **21**, 2376–2385 (1980).
[16]L. E. Thomas, "Birman–Schwinger bounds for the Laplacian with point interactions," J. Math. Phys. **20**, 1848–1850 (1979).
[17]V. I. Karloukovski, "Scattering on δ-function potentials," Rpt. Math. Phys. **10**, 87–105 (1976).
[18]J. Zorbas, "Perturbation of self-adjoint operators by Dirac distributions," J. Math. Phys. **21**, 840–847 (1980).
[19]F. Gesztesy and L. Pittner, "Two-body scattering for Schrödinger operators involving zero-range interactions," Rep. Math. Phys. (to be published).
[20]S. Albeverio, R. Høegh-Krohn, W. Kirsch, and F. Martinelli, "The spectrum of the three-dimensional Kronig–Penney model with random point defects," Adv. Appl. Math. **3**, 435–440 (1982).
[21]H. Holden, R. Høegh-Krohn, and F. Martinelli, "The spectrum of defect periodic point interactions", Lett. Math. Phys. **7**, 221–228 (1983).
[22]S. Albeverio, R. Høegh-Krohn, and T. T. Wu, "A class of exactly solvable three-body quantum mechanical problems and the universal low energy behavior," Phys. Lett. **83A**, 105–109 (1981).
[23]S. Albeverio, "Analytische Lösung eines idealisierten Stripping—oder Beugungsproblems," Helv. Phys. Acta **40**, 135–184 (1967).
[24]M. Gaudin, "Sur le problème de deux on trois electrons en presence d'un moment localisé," J. Phys. **39**, 1143 (1978).
[25]A. Grossmann and T. T. Wu, "A class of potentials with extremely narrow resonances I. Case with discrete rotational symmetry," CNR-CPT Marseille Preprint, 1981.
[26]S. Albeverio, R. Høegh-Krohn, and L. Streit, "Energy forms, Hamiltonians, and distorted Brownian paths," J. Math. Phys. **18**, 907–917 (1977).
[27]T. Brasche, "Störungen von Schrödinger Operatoren durch Masse" Diplomarbeit Bielefeld 1983.
[28]S. Albeverio and R. Høegh-Krohn, "Perturbations of resonances in quantum mechanics," Bochum Preprint, 1982, to appear in J. Math. Anal. Appl.
[29]R. Høegh-Krohn and M. Mebkhout, "The multiple well problem. Asymptotic behaviour of the eigenvalues and resonances," Marseille Preprint, 1982.
[30]J. Avron, A. Grossmann, and R. Rodriguez, "Hamiltonians in the one-

electron theory of solids," Rep. Math. Phys. **5**, 113–120 (1974).
[31]L. E. Thomas, "Time-dependent approach to scattering from impurities in a crystal," Commun. Math. Phys. **33**, 335–343 (1973).
[32]G. M. Troianiello, "Scattering theory for Schrödinger operators with L^∞ potentials and distorted Bloch waves," J. Math. Phys. **15**, 2048–2052 (1974).
[33]F. Bentosela, "Scattering from impurities in a crystal," Comm. Math. Phys. **46**, 153–166 (1976).
[34]M. Reed and B. Simon, *Methods of modern mathematical physics, IV. Analysis of operators* (Academic, New York, 1978).
[35]G. Scharf, "Das Blochsche Theorem für unendliche Systeme," Helv. Phys. Acta **39**, 556–560 (1966).
[36]F. Odeh and J. B. Keller, "Partial differential equations with periodic coefficients and Bloch waves in crystal," J. Math. Phys. **5**, 1499–1504 (1964).
[37]J. Callaway, "Theory of scattering in solids," J. Math. Phys. **5**, 783–798 (1964).
[38]S. T. Kuroda, "A stationary method of scattering and some applications," in *Proceedings of the International Conference on Function Analysis and Related Topics* (Univ. of Tokyo Press, Tokyo, 1970).
[39]E. Davies and B. Simon, "Scattering theory for systems with different spatial asymptotics on the left and right," Commun. Math. Phys. **63**, 277–301 (1978).
[40]J. E. Avron, A. Grossmann, and R. Høegh-Krohn, "The reflection from a semi-infinite crystal of point scattering," preprint, Calif. Inst. Technol.
[41]L. W. MacMillan and T. A. Osborn, "Transit time measures in scattering theory," Ann. Phys. (NY) **126**, 1–19 (1980).
[42]R. F. Alvarez-Estrada, "Interaction, diffraction, and scattering of a two-particle system by a thin crystal," Ann. Phys. (NY) **93**, 1–40 (1975).
[43]R. F. Alvarez-Estrada and M. E. Villalon, "On the applicability of the Lippmann-Schwinger approach to the scattering of particles by semiinfinite crystals," Ann. Phys. (NY) **111**, 239–250 (1978).
[44]A. M. Berthier, *Spectral theory and wave operators for the Schrödinger equation*, Research Notes in Mathematics **71** (Pitman, Boston, 1982).
[45]R. Newton, *Scattering theory of waves and particles* (McGraw-Hill, New York, 1982), 2nd ed.
[46]M. Reed and B. Simon, *Methods of mathematical physics, III: Scattering theory* (Academic, New York, 1979).
[47]W. Amrein, J. Jauch, and K. Sinha, *Scattering theory in quantum mechanics* (Benjamin, Reading, PA 1977).
[48]J. M. Ziman, *Principles of theory of solids* (Cambridge U. P., Cambridge, 1964).

On a class of nonintegrable equations in 1 + 1 dimensions with factorized associated linear operators

Henri Cornille[a)]
Division de la Physique Theorique, CEN-Saclay, F-91191 Gif-sur-Yvette Cedex, France

(Received 16 November 1982; accepted for publication 29 July 1983)

We consider a class of nonintegrable nonlinear equations with powerlike nonlinearities $K^N, \partial_x K^N$ (K being the solutions and $N \geqslant 2$, N integer) building explicitly their exponential type bisolitons. The denominators of the bisolitons have no soliton couplings, and the linear differential operators of the linear part of the equations are factorized operators. We extend our study to a larger class of nonlinearities: polynomial nonlinearities which are linear combinations of powerlike nonlinearities. We study the two extreme possibilities. Either the bisoliton is specific of a mixed nonlinearity, not being a solution of any component nonlinearity, or the bisoliton is common to all components. Different properties occur depending whether the components are K^N or $\partial_x K^N$. For K^N nonlinearities, in order to understand the origin of the factorization of the linear operators, we give a criterion which is easily checked at an almost entirely linear level of constraints. We conjecture that all possible bisolitons are of the type studied here. Finally for K^N and $\partial_x K^N$ we enlarge the class of bisolitons found in previous works.

PACS numbers: 02.90. + p, 02.30.Jr

1. INTRODUCTION

Although the nonintegrable nonlinear equations are interesting to study (see Ref. 1), there are actually very few theoretical investigations in this field. This is partly due to the lack of methods to solve these equations but also to the difficulty of defining classes of nonintegrable equations with specific properties. Maybe a tool to classify these equations could be provided by the study of nontrivial solutions such as the bisolitons (in the paper my use of this word is a working definition). Recently,[2–5] we have introduced a class of nonintegrable equations in 1 + 1 dimensions sharing common features.

(i) They contain monomial powerlike nonlinearities $K^N, K^{N-1}K_x$ (K being the solution and N integer, $N \geqslant 2$).

(ii) Let L_{qN} be a qth differential operator in 1 + 1 dimensions with constant coefficients, associated with the linear part $L_{qN}K$ of the equation. Then, either L_{qN} is a factorized operator or, in the $K^{N-1}K_x$ case, L_{NN} is a germ differential operator which becomes a factor of L_{qN} when $q > N$.

(iii) *Let us define bisolitons as solutions with only two exponential variables* $\omega_i = \exp(\gamma_i x + \rho_i t)$ (which can be rewritten $\omega_i = \exp x_i$ by linear transformations of the coordinates) *such that there exists some powers of the solutions which are rational functions.* Then, if they are not "trivial bisolitons," their denominators are functions of $\Delta = 1 + \omega_1 + \omega_2$ without the coupling terms const $\omega_1\omega_2$. The trivial bisolitons associated with K_N nonlinearities are direct product of solitons solutions, and their denominators contain products of the functions $\Delta_j = 1 + \omega_j, j = 1,2$.

(iv) There exists a direct constructive method of simultaneously building both the factorized L_{qN} and the appropriate power of the solution. The results were obtained with simple examples and for any integer N value.

Our class can be characterized by two properties which must simultaneously be present: bisolitons without soliton couplings and factorization of the associated linear operators. We build the class in a constructive way. It appears that the most efficient tool is the one which builds simultaneously L_{qN} and the nonlinearity solution. Maybe other methods could be tried. For K^2 where it was proved[3,4] for rational bisoliton functions that the denominators depend only on Δ, we write K as a polynomial $\Sigma\omega_2^l F_l(\Delta)$, solve the (F_l) equations, and only at the end verify that L_{q2} is factorized. Even for $N = 2$ this method is very cumbersome and not convenient for a generalization in the $N > 2$ case where the bisolitons are no more rational functions.

Can we obtain a larger class of nonlinearities, polynomial nonlinearities being linear combinations of powerlike nonlinearities such that the two above properties hold? For a nonlinearity sum of different terms, many situations can occur. Either the solution is common to all the components, or to some of them, or does not belong to any one of them. Here, for simplicity, we limit our study to the two extreme possibilities: either common to all the components or distinct.

In Sec. 2, we seek intrinsic bisolitons to the mixed nonlinearity $\Sigma_i \lambda_i \partial_{x^i} K^N$, with the same N integer value and which do not belong to any component K^N or K_x^N or K_{xx}^N. We find that this occurs only for well-defined linear combintions of the monomial nonlinearities: for instance, $(-N/(N+1) + \partial_x)K^N$. In all cases, the linear operator associated with the mixed nonlinearity is factorized. Further, using trivial transforms, we show that the solutions can be associated with a less number of nonlinearities but the coefficients in L_{qN} being not constant values. This means that our class contains linear differential operators with variable dependent coefficients, a property not seen in previous works.

In all other sections we look at bisolitons common to all components of the nonlinearity. In Sec. 4, we give, for $L_{q_iN_i}K$

[a)] Chercheur au C.N.R.S.

 0022-2488/84/051335-13$02.50

$= K^{N_i}$, a criterion, in order that our ansatz solution necessarily corresponds to a factorized $L_{q_iN_i}$. If the ansatz is a solution, the factor, being N_i independent, is common to all $L_{q_iN_i}$. In Secs. 3 and 5 for "true bisolitons" (denominators depending only on $\Delta = 1 + \Sigma\omega_i$) and "trivial bisolitons" (on $\Delta_j = 1 + \omega_j$), we effectively build this common factor of the set $L_{q_iN_i}$. We find it has in general more terms than the simple factor predicted by the criterion of Sec. 4. When we introduce an ansatz solution into the nonlinear equation, a part of $L_{q_iN_i}$ called $L_{\mathrm{I}q_iN_i}$ must kill terms present in the linear part and absent in the nonlinear one, whereas $L_{\mathrm{II}q_iN_i}$, the remaining operator, $L_{q_iN_i} = L_{\mathrm{I}q_iN_i}L_{\mathrm{II}q_iN_i}$, reconstructs exactly the nonlinearity. The common factor is the intersection $\cap L_{\mathrm{I}q_iN_i}$ $\forall N_i$ and is reduced to the $L_{\mathrm{I}q_iN_i}$ associated with the smallest N_i integer. In Sec. 6, we consider $\partial_\lambda K^{N_i}$ or equivalently $L_{q_iN_i}G = (G_x)^{N_i}$ for the potentials $G_x = K$, $G = G(\Delta)$. We introduce germs differential operators $\tilde{l}_N$ and $L_{q_iN_i}$ $= \tilde{l}_N(x,t)l_{q_i-N_i}(t)$ and the remaining problem is to solve $\partial_\Delta^{N_i}{}_{N_i} l_{q_i-N_i}G = (G_\Delta)^{N_i}$. We find two classes: Either (i) the common solution is a polynomial with Δ^{-1} power terms, $(\Delta = 1 + \Sigma\omega_i)$; in which case necessarily the $L_{q_iN_i}$ have a common factor $l_{\mathrm{I}q_1-N_1}$ if $N_1 < N_i$ or (ii) G includes a $\log\Delta$ term and we exhibit counterexamples to the common factorization property.

Can we understand, in a simple way, the necessity of factorized L_{qN} associated with K^N? In Sec. 4 we give criterion: let us assume mainly that K is a bisoliton without soliton couplings and $K \simeq \omega_2^{m_2}$ when $\omega_2 \to 0$, ω_1 fixed, then $K^N \simeq \omega_2^{Nm_2}$ whereas $L_{qN}K \simeq \text{const}\,\omega_2^{m_2}$. The constant, which depends on the coefficients of L_{qN}, must vanish, thus leading to factorized L_{qN}. This very important result is established almost entirely at a linear level of constraints. Unfortunately, it is not possible to remove the assumptions of the criterion; from simple examples we know that otherwise we must take into account a great number of nonlinear constraints. *Consequently, as a result of our study, we conjecture that all bisolitons of K^N satisfy the assumptions of the criterion (or can be reduced to it by trivial transforms), i. e., they are without soliton couplings and have factorized L_{qN}.*

In the monomial cases K^N and $(G_x)^N$ can we enlarge the class of solutions and L_{qN} previously[2,3] obtained? In the different sections we explain, quite generally, how to build separately the two above-defined operators $L_{\mathrm{I}qN}$, $L_{\mathrm{II}qN}$. In Sec. 3, we study the true bisolitons of $L_{qN}K = K^N$ and give a solution with an arbitrary number of Δ^{-1} terms. In Sec. 5 we introduce the "trivial bisolitons" not considered in Refs. 2 and 3. They are products of two solitons $K(\omega_1,\omega_2)$ $= K(\omega_1)K(\omega_2)$, and the denominators can contain $\Delta_1\Delta_2 = 1 + \omega_1 + \omega_2 + \omega_1\omega_2$ with a soliton couplings 1, independently of the parameters values of the solitons. By construction, their L_{qN} are factorized, being the product of the two-soliton linear operators. In Sec. 6, for $(G_x)^N$ nonlinearities we build general classes of bisolitons $G(\Delta)$ and associated l_{q-N} operators. For instance, for Δ^{-1} polynomials G solutions we obtain different families with an arbitrary number of terms.

For a study of nonlinear equations classes there exist generic equations. For the K^N nonlinearity it is the Boltzmann equation (Sec. 7) and for $(G_x)^N$ they are the equations associated with the germs operators $\tilde{l}_N$: Burgers equation and generalization.

In the tables (I–V) we quote both the nonlinear equations with factorized linear operators and their solutions.

2. INTRINSIC BISOLITONS TO MIXED NONLINEARITIES $(\Sigma_0^{i\max}\lambda_i\partial_x^i)K^N$

A. General considerations

In this section we seek exponential type bisolitons which are specific of a mixed nonlinearity and do not belong to any one of the components, K^N alone or $\partial_x K^N$ alone or $\partial_{x^2}^2 K^N$ alone... . Let $L_{qN}(x,t)$ be a qth differential operator with constant coefficients and consider a polynomial nonlinearity $L_{qN}K = c_{qN}(\Sigma_{i=0}^{i\max}\lambda_i\partial_{x^i}^i)K^N$ *with the same N integer power*. Practically, we shall mainly discuss a sum of two nonlinearities $(\lambda_0 + \lambda_1\partial_x)K^N$.

In order to simplify the formalism, by a linear transformation $(x,t) \to (x_1,x_2)$ we symmetrize the nonlinearity $\partial_x \to \partial_{x_1} + \partial_{x_2}$ and $\omega_i \to \exp x_i$. We assume that the denominators of the bisolitons are functions of $\Delta = 1 + \Sigma_1^2\omega_i$, and look at the class of nonlinear equations

$$L_{qN}K = \sum_{i+j=0}^{i+j=q} a_{ij}\partial_{x_1^i x_2^j}^{i+j}K(\omega_1,\omega_2) = c_{qN}\sum_{i=0}^{2}\lambda_i(\partial_{x_1}+\partial_{x_2})^i K^N. \tag{2.1}$$

These specific bisolitons exist only for well-defined linear combinations of the nonlinearities, and, as we will explain, this is due to a balance between the contributions of both the linear part and the nonlinear one.

(i) First, we consider $K = \omega_2^m\Delta^{-p}$ and a single nonlinearlity K^N. L_{qN} is a qth differential operator such that when $K \to L_{qN}K$, then $\Delta^{-p} \to \Delta^{-Np}$ with $Np = p + q$ whereas $\omega_2^m \to \omega_2^{Nm}$, $Nm \leqslant q + m$. This is the key property and is due to the fact that differential operators necessarily raise the power of the denominator but not necessarily the one of the numerator. If $m < p$, we find $mN \leqslant m + (N-1)p < pN$, and there is no contradiction; on the contrary, if $m > p$, then $mN \leqslant m + (N-1)p < mN$ and this is not possible.

Now we look at a mixed nonlinearity that we write $(-NQ + \partial_{x_1} + \partial_{x_2})K^N$ with Q arbitrary. We find

$$K = \omega_2^m\Delta^{-p} \to (-NQ + \partial_{x_1} + \partial_{x_2})K^N = (\omega_2^m\Delta^{-p})^N[N(m-p-Q) + pN\Delta^{-1}].$$

A part of the operator L_{qN} must build $\omega_2^{mN}\Delta^{-pN}$. This part is a k th operator such that $\Delta^{-p} \to \Delta^{-Np}$, $q(N-1) = k$ and $\omega_2^m \to \omega_2^{m+M}$, $M \leqslant k$. Applying the previous argument, if $m > p$, then $m \to M \leqslant m + k = m + p(N-1) < mN$. This is not possible, and, consequently, the term $(\omega_2^m\Delta^{-p})^N$ must disappear and $Q = m - p$. For this example, when $K \to L_{qN}K$, then $\Delta^{-p} \to \Delta^{-(Np+1)}$ and $p = (q-1)(N-1)^{-1}$.

(ii) Second, we consider an ansatz more complicated

$$K=\omega_2^m\Delta^{-p}\left(1+\sum_1^k b_l\Delta^{-l}\right)\to(NQ+\partial_{x_1}+\partial_{x_2})K^N$$
$$=(\omega_2^m\Delta^{-p})^N(N(m-p-Q)+O(\Delta^{-1})).$$

The same argument as above shows that if $m>p$ (i.e., intrinsic bisolitons of the mixed nonlinearity), then necessarily $Q=m-p$.

B. $K=\omega_2^m\Delta^{-p}$, $m>p$, and a nonlinearity sum of two terms

We want to find the factorized differential operator L_{qN} of order q such that

$$L_{qN}K=c_{qN}[N(p-m)+\partial_{x_1}+\partial_{x_2}]K^N$$
$$=c_{qN}pN\omega_2^{Nm}/\Delta^{pN+1}. \tag{2.2}$$

Necessarily, we have $p+q=Np+1$, and we define a raising ω_2^m and Δ^{-p} first-order differential operator

$$l_{x_i}=\left(1-\frac{\partial_{x_i}}{M_i}\right),\quad l_{x_i}\frac{\omega_i^{M_i}\omega_j^{M_j}}{\Delta^Q}=\frac{Q}{M_i}\frac{\omega_i^{M_i+1}\omega_j^{M_j}}{\Delta^{Q+1}}$$

and apply it q times to K for $x_i=x_2$ (Table I).

C. $K=\omega_2^{m_2}\omega_1^{m_1}\Delta^{-p}$, $m_1+m_2>p$, and a nonlinearity sum of two terms

Applying the previous argument in order to determine the necessary linear combination of K^N and $(\partial_{x_1}+\partial_{x_2})K^N$, we find $q=(N-1)p+1$,

$$L_{qN}K=c_{qN}[N(p-m_1-m_2)+\partial_{x_1}+\partial_{x_2}]K^N$$
$$=\frac{pN}{\Delta}\left(\frac{\omega_1^{m_1}\omega_2^{m_2}}{\Delta^p}\right)^N \tag{2.3}$$

and we apply the l_{x_i} operators (Table I).

D. $K=\omega_2^m\Delta^{-p}$, $m>p$, and a nonlinearity sum of three terms

Starting with a nonlinearity $[\alpha_0N^2+\alpha_1N(\partial_{x_1}+\partial_{x_2})+(\partial_{x_1}+\partial_{x_2})^2]K^N$ and substituting the K ansatz, we find $p+q=N+2$ and

$$[N^2(m-p)(m-p-1/N)+(2N(p+m)+1)(\partial_{x_1}+\partial_{x_2})+(\partial_{x_1}+\partial_{x_2})^2]\omega_2^{mN}/\Delta^{pN}$$
$$=pN(pN+1)\omega_2^{mN}/\Delta^{pN+2}. \tag{2.4}$$

It remains to build a linear operator L_{qN} which when applied to K gives $\omega_2^{mN}\Delta^{-(pN+2)}$. We still use a product of l_{x_i} order operators (Table I).

E. $K=(\omega_2^m/\Delta^p)(1+b\Delta^{-1})$ and a sum of two nonlinearities

Following the general considerations of 2.1, the mixed nonlinearity gives a contribution

TABLE I. Intrinsic bisolitons to $\Sigma_0^2\lambda_j(\partial_{x_1}+\partial_{x_2})^jK^N$, $\Delta=1+\Sigma_1^2\omega_i$, $\omega_i=e^{x_i}$.

$$L_{qN}K=c_{qN}\left(-\frac{N}{N-1}+\partial_{x_1}+\partial_{x_2}\right)K^N$$

$K=(\omega_2\Delta^{-1})^{(q-1)/(N-1)}\omega_2^{-1/(N-1)}$, $L_{qN}=\prod_{s=q/(N-1)}^{Nq/(N-1)-1}(1-s^{-1}\partial_{x_2})$, $c_{qN}=N^{-1}\frac{\Gamma(q/(N-1))\Gamma((qN-1)/(N-1))}{\Gamma(qN/(N-1))\Gamma((q+N-2)/(N-1))}$, $q\geqslant 2$

$K=\left[\frac{\omega_1^{M_1}\omega_2^{M_2}}{\Delta^{q-1}}\right]^{1/(N-1)}$, M_i integers >0, $M_1+M_2=q$, $q\geqslant 2$

$L_{qN}=\prod_{j=1,2}\prod_{s_j=M_j/(N-1)}^{(N/(N-1))M_j-1}\left(1-\frac{\partial_{x_j}}{s_j}\right)$, $c_{qN}=N^{-1}\frac{\Gamma((qN-1)/(N-1))}{\Gamma((q+N-2)/(N-1))}\prod_{j=1,2}\frac{\Gamma(M_j/(N-1))}{\Gamma(NM_j/(N-1))}$

$K=\frac{\omega_2^m}{\Delta^p}(1+b\Delta^{-1})$, $m=\frac{q-N+1}{N-1}$, $p=\frac{q-N}{N-1}$, $L_1=\prod_{(q-N+1)/(N-1)}^{[N(q-N)+1]/(N-1)}\left(1-\frac{\partial_{x_2}}{s}\right)$, $\mathscr{L}_{qN}=L_1L_{II}$, $L_1K=\tilde{K}$, $q\geqslant N+1$

$\mathscr{L}_{II}=1+\sum_1^{N-1}\Omega_j(\partial_{x_1}+\partial_{x_2})^j$, $\mathscr{L}_{II}\tilde{K}=c_{qN}\omega_2^{mN}X(\Delta^{-1})$, $\tilde{K}=\left[\prod_{s=0}^{(N-1)p}\left(\frac{p+s}{p+1/(N-1)+s}\right)\right]\frac{\omega_2^{mN}}{\Delta^{Np+1}}\left(1+\frac{b}{p}\frac{(Np+1)}{\Delta}\right)$

$N=2$, $K=\frac{\omega_2^{q-1}}{\Delta^{q-2}}\left(1+\frac{(q-2)^2}{(3q-4)}\Delta^{-1}\right)$, $L_{q2}=\prod_{q-1}^{2q-3}\left(1-\frac{\partial_{x_2}}{s}\right)\left(1+\frac{(q-2)^2}{7q^2-16q+8}(\partial_{x_1}+\partial_{x_2})\right)$, $c_{q2}=\frac{3q-4}{7q^2-16q+8}$, $q\geqslant 3$

$N=3$, $b^2-b\left(18p+21+\frac{6}{p}\right)+(6+8p)p=0$, $\Omega_2^{-1}=\frac{1}{4}-\frac{(3+4p)}{b}+\frac{3(3p+1)(3p+2)(p-b)}{pb^2}$, $q\geqslant 4$

$\Omega_1=\Omega_2\left(-1+\frac{6+8p}{b}\right)$, $c_{q3}=\Omega_2\frac{\Gamma(3p+3)}{\Gamma(3p+\frac{3}{2})}\frac{\Gamma(p+\frac{1}{2})}{\Gamma(p)b^2p}$,

$$L_{qN}K=c_{qN}\left[\frac{2N(N+1)}{(N-1)^2}-\frac{(3N+1)}{N-1}(\partial_{x_1}+\partial_{x_2})+(\partial_{x_1}+\partial_{x_2})^2\right]K^N$$

$K=(\omega_2\Delta^{-1})^{(q-2)/(N-1)}\omega_2^{2/(N-1)}$, $L_{qN}=\prod_{s=q/(N-1)}^{Nq/(N-1)-1}\left(1-\frac{\partial_{x_2}}{x}\right)$, $c_{qN}=\frac{\Gamma(q/(N-1))\Gamma(N(q-2)/(N-1))}{\Gamma((q-2)/(N-1))\Gamma(Nq/(N-1))}$

$$c_{qN}[\,-N(m-p)+\partial_{x_1}+\partial_{x_2}]K^N=c_{qN}\omega_2^{Nm}X(\Delta^{-1}),$$

$$X(\Delta^{-1})=\Delta^{-pN}\Bigg[\sum_{l=1}^{N}b^{l-1}\Delta^{-l}c_N^{l-1}[N(p-b)+(l-1)(b+1)]+b^N N(p+1)\Delta^{-(N+1)}\Bigg],$$

where X is a Δ^{-1} polynomial with well-defined p, b-dependent coefficients. Our aim is to find a factorized L_{qN} differential operator, of order q, which when applied to K reproduces exactly $c_{qN}\omega_2^{mN}X(\Delta^{-1})$. We must kill all $\omega_2^M\Delta^{-Q}$ terms having $M<mN$ and $Q<pN$; this will define a first operator $\mathscr{L}_{\rm I}$. The complementary operator $\mathscr{L}_{\rm II}$, $L_{qN}=\mathscr{L}_{\rm I}\mathscr{L}_{\rm II}$, must both preserve the ω_2^{mN} power and reconstruct exactly $X(\Delta^{-1})$ (Table I).

It was suggested by Brézin (private communication) that these solutions could be rewritten as solutions with *a less number of nonlinearities but* x_i-dependent linear operators $L_N\to\tilde{L}_{qN}$. For instance,

$$L_{qN}K=c_{qN}(-N(N-1)^{-1}+\partial_{x_2}+\partial_{x_2})K^N,$$
$$K=H\exp((x_1+x_2)/(2(N-1))),$$
$$\tilde{L}_{qN}H=c_{qN}(\partial_{x_1}+\partial_{x_2})H^N,$$
$$L_{qN}=e^{[-N/2(N-1)](x_1+x_2)}L_{qN}e^{[1/2(N-1)](x_1+x_2)}.$$

It is clear that $\tilde{L}_{qN}$ does not only depend on $\partial^{i+j}_{x_1^i x_2^j}$ with constant coefficients a_{ij} but also on x_1+x_2. For the solutions of Table I where the corresponding L_{qN} are written down, we can easily build $\tilde{L}_{qN}$. Similarly

$$L_{qN}K=c_{qN}\Bigg[\frac{2N(N+1)}{(N-1)^2}\frac{-(3N+1)}{N-1}(\partial_{x_1}+\partial_{x_2})+(\partial_{x_1}+\partial_{x_2})^2\Bigg]k^N,$$

$$K=H\exp\Big(\frac{\lambda^{\pm}(x_1+x_2)}{2}\Big),$$

$$\tilde{L}_{qN}H=c_{qN}(\partial_{x_1}+\partial_{x_2})(\pm 1+\partial_{x_1}+\partial_{x_2})H^N,$$

$$\tilde{L}_{qN}=\exp\Big(\frac{-\lambda^{\pm}N(x_1+x_2)}{2}\Big)L_{qN}\times\exp\Big(\frac{\lambda^{\pm}(x_1+x_2)}{2}\Big),$$

$$\lambda^{+}=2(N-1)^{-1},\quad \lambda^{-}=(N+1)(N(N-1))^{-1}.$$

3. "TRUE BISOLITONS" COMMON TO NONLINEARITIES K^{N_i} AND TO MIXED NONLINEARITIES $\Sigma_i\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}$

A. General remarks

In this section we work with the so-called "true bisolitons" of $L_{qN}K=K^N$, where $L_{qN}\equiv L_{qN}(x_1,x_2)$ are qth-order differential operators with constant coefficients acting on two independent variables x_1 and x_2. These bisolitons depend on two variables which for instance can be $\Delta=1+\Sigma_1^2\,\omega_i$ and one of the two $\omega_i=\exp x_i$, and their denominators are built with Δ.

Let us consider a mixed nonlinearity of the following type:

$$L_qK=\Big(\sum_i\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}L_{q_iN_i}\Big)K=\sum_i\lambda_i c_{q_iN_i}(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}\qquad(3.1)$$

and assume that K is a solution common to different K^{N_i} nonlinearities

$$L_{q_iN_i}K=c_{q_iN_i}K^{N_i},\quad i=1,2,\ldots,i_{\max}.$$

The property that we want to check is whether or not L_q in Eq. (3.1) is a factorized operator, in other words, if the set $\{L_{q_iN_i}\}$ have common factors.

Let us look at a family of solutions having $k+1$ different Δ^{-1} terms: $K=\omega_1^{m_1}\omega_2^{m_2}\Delta^{-p}(1+\Sigma_1^k b_l\Delta^{-l})$. We want to explain why the set $\{L_{q_iN_i}\}$ must have common factors if a solution exists for different q_i, N_i values. In an intermediate step of the building of the K^N nonlinearity, by the operator L_{qN} acting on K, we must necessarily go through a term like $(\omega_1^{m_1}\omega_2^{m_2})^N\Delta^{-(p+(N-1)(m_1+m_2))}$ multiplied by a Δ^{-1} polynomial. If N_1 is the smallest N_i value, then all the ith paths from K to K^N have in common the path from K to $(\omega_1^{m_1}\omega_2^{m_2})^{N_1}\Delta^{-(p+(N_1-1)(m_1+m_2))}$ multiplied by a Δ^{-1} polynomial. This common path, translated in terms of differential operators, gives the common factor to the $\{L_{q_iN_i}\}$.

B. $K=\omega_1^{m_1}\omega_2^{m_2}\Delta^{-p}$, $m_i\geqslant 0$

We first study the nonlinearity K^N alone and build the factorized L_{qN} operator. We must go from K to K^N. With the help of the two first-order differential operators

$$\Big(1-\frac{\partial_{x_j}}{M_j}\Big)\frac{\omega_j^{M_j}}{\Delta^Q}=\frac{Q}{M_j}\frac{\omega_2^{M_j+1}}{\Delta^{Q+1}},$$

$$\Big(1+\frac{1+\partial_{x_1}+\partial_{x_2}}{Q-M_1-M_2}\Big)\frac{\omega_1^{M_1}\omega_2^{M_2}}{\Delta^Q}=\frac{Q\omega_1^{M_1}\omega_2^{M_2}}{Q-M_1-M_2}\Delta^{-(Q+1)}$$

we follow the path

$$K\to\frac{(\omega_1^{m_1}\omega_2^{m_2})^N}{\Delta^{p+(N-1)(m_1+m_2)}}\to K^N$$

and we find L_{qN}, c_{qN} (Table II).

C. $K=\omega_2^m\Delta^{-p}(1+pk^{-1}\Delta^{-1})$, k integer

In order to reconstruct K^N we follow a path with two intermediate stages

$$K\to\text{const}\Big(\frac{\omega_2^{Nm}}{\Delta^{p+(N-1)m+k+1}}\Big)\to\text{const}\Big(\frac{\omega_2^m}{\Delta^p}\Big)^N\to K^N.$$

In order to reach the first stage we successively apply two

TABLE II. "True bisolitons" for K^N, $\Delta = 1 + e^{x_1} + e^{x_2} = 1 + \Sigma\omega_i$.

$$L_{qN}K = c_{qN}K^N, \quad L_{qN} = L_{\mathrm{I}qN}L_{\mathrm{II}qN}$$

$$L_qK = \left(\sum_i \lambda_i(\partial_{x_1} + \partial_{x_2})^{l_i}L_{qN_i}\right)K = \sum \lambda_i c_{qN_i}(\partial_{x_1} + \partial_{x_2})^{l_i}K^{N_i}, \quad N_1 < N_i$$

$$K = \frac{\omega_1^{m_1}\omega_2^{m_2}}{\Delta^p}, \quad p = \frac{q}{N-1}, \quad m_i = \frac{M_i}{N-1}, \quad 1 < M_1 + M_2 < q-1, \quad M_i \text{ integers} > 0$$

$$L_{qN} = \prod_{j=1,2}\prod_{m_j}^{Nm_j-1}\left(1 - \frac{\partial_{x_j}}{s_j}\right)\prod_{p-m_1-m_2}^{N(p-m_1-m_2)-1}\left(1 + \frac{\partial_{x_1}+\partial_{x_2}}{r}\right) c_{qN} = \frac{\Gamma(Np)\Gamma(p-m_1-m_2)}{\Gamma(p)\Gamma(N(p-m_1-m_2))}\prod_{j=1,2}\frac{\Gamma(m_j)}{\Gamma(Nm_j)}$$

L_q factorizes $L_{q_1N_1}$, $m_j = \dfrac{M_{ji}}{N_i - 1}$, $j = 1, 2, \forall i$, $p = \dfrac{q_i}{N_i - 1}\ \forall i$

$$K = \omega_2^m\Delta^{-p}(1 + pk^{-1}\Delta^{-1}), \quad p = \frac{q}{N-1} - 1, \quad m = \frac{M}{N-1}, \quad \mathscr{L}_{\mathrm{II}qN} = 1 + \sum_1^N \Omega_j(\partial_{x_1} + \partial_{x_2})^j$$

$$\mathscr{L}_{\mathrm{I}qN} \equiv \left\{\prod_m^{Nm-1}\left(1 - \frac{\partial_{x_2}}{s}\right)\prod_{r=p-m}^{p-m+k-1}\left(1 + \frac{\partial_{x_1}+\partial_{x_2}}{r}\right)\left[\prod_{p-m+k+1}^{N(p-m)-1}\left(1 + \frac{\partial_{x_1}+\partial_{x_2}}{r}\right)\right]\right. \quad \text{if } q > k + M + N + 1,$$

$[\ ] \equiv 1$ if $q = k + M + N$

$$c_{qN} = \nu k^{-1}\frac{\Gamma(Np)\Gamma(m)}{\Gamma(p)\Gamma(Nm)}\left[\prod_{\substack{s=0\\ s\neq k}}^{(N-1)(p-m)-1}(p - m + s)\right]^{-1}, \quad L_{\mathrm{I}qN} = \left(\frac{\omega_2^m}{\Delta^p}\right)^N\nu^{-1}c_{qN}$$

$$\mathscr{T}_j^l = \mathscr{T}_{j-1}^{l-1} + (N(m-p) - l + 1)\mathscr{T}_{j-1}^l\mathscr{T}_j^1 = (N(m-p))^{j-1}$$

$$1 = \nu^{-1} + \sum_1^N \Omega_j\nu^{-1}\mathscr{T}_{j+1}^1, \quad \sum_{j=s}^N \nu^{-1}\Omega_j\mathscr{T}_{j+1}^{s+1} = c_N^s\left(\frac{p}{k}\right)^s\Gamma(Np+s)[\Gamma(Np)]^{-1}$$

$$N = 2, \quad 1 = \nu^{-1} + \sum_1^2 2^j(m-p)^j\Omega_j\nu^{-1}, \quad k^{-1} = \Omega_1\nu^{-1} + \Omega_2\nu^{-1}(4(m-p) - 1), \quad \Omega_2\nu^{-1} = \frac{p}{2k^2(2p+1)}$$

$$N = 3, \quad 1 = \nu^{-1} + \sum_1^3 3^j(m-p)^j\Omega_j\nu^{-1}, \quad \nu k^{-1} = \Omega_1 + \Omega_2(4(m-p) - 1) + \Omega_3(27(m-p)^2 + 9(p-m) + 1)$$

$$\frac{\nu p}{k^2(3p+1)} = \Omega_2 + 3\Omega_3(3(m-p) - 1), \quad \frac{\nu p^2}{3k^3(3p+1)(3p+2)} = \Omega_3$$

L_q factorizes $\mathscr{L}_{\mathrm{I}q_iN_i}$, $m = \dfrac{M_i}{N_i - 1}$, $p = \dfrac{q_i}{N_i - 1}$, $q_i > k + N_i + M_i$

first-order differential operators

$$\left(1 - \frac{\partial_{x_2}}{M}\right)\frac{\omega_2^M}{\Delta^Q}(A + B\Delta^{-1}) = \frac{\omega_2^{M+1}}{M\Delta^{Q+1}}(AQ + B(Q+1)\Delta^{-1}),$$

$$\left(1 + \frac{1}{Q-M}(\partial_{x_1} + \partial_{x_2})\right)\frac{\omega_2^M}{\Delta^Q}(A + B\Delta^{-1}) = \frac{\omega_2^M}{(Q-M)\Delta^{Q+1}}(AQ - B + B(Q+1)\Delta^{-1}),$$

the first one up to $\omega_2^m \to \omega_2^{Nm}$ and the second increases the power of the two Δ^{-1} terms up to a stage where one Δ^{-1} term vanishes. This second stage defines a first operator $\mathscr{L}_{\mathrm{I}qN}$ of $L_{qN} = \mathscr{L}_{\mathrm{I}qN}\mathscr{L}_{\mathrm{II}qN}$ while $\mathscr{L}_{\mathrm{II}qN}$ is the Nth differential operator which when applied to $(\omega_2^m\Delta^{-p})^N$ reproduces exactly K^N (Table II)

$$\mathscr{L}_{\mathrm{II}qN} = 1 + \sum_1^N \Omega_j(\partial_{x_1} + \partial_{x_2})^j,$$

$$\left(\frac{\omega_2^m}{\Delta^p}\right)^{-N}\mathscr{L}_{\mathrm{II}}\left(\frac{\omega_2^m}{\Delta^p}\right)^N \equiv \nu\left(1 + \frac{p}{k\Delta}\right)^N. \tag{3.2}$$

In order to determine the operator $\mathscr{L}_{\mathrm{II}}$ we introduce the identity:

$$\left(\frac{\omega_2^m}{\Delta^p}\right)^{-N}(\partial_{x_1} + \partial_{x_2})^{j-1}\left(\frac{\omega_2^m}{\Delta^p}\right)^N = \mathscr{T}_j^1 + \sum_{l=2}^j \mathscr{T}_j^l\frac{\Gamma(Np + l - 1)}{\Gamma(Np)\Delta^{l-1}}$$

which defines the numbers $\mathscr{T}_j^l$ (Table II).

The identity (3.2) leads to a set of triangular relations for the (Ω_i/ν), $1/\nu$. Consequently, these parameters are recursively determined, and we find both ν^{-1} and the set (Ω_i) (Table II).

D. $K = (\omega_2/\Delta)^p(1 + \Sigma_1^k(b_m/\Delta^m))$, $k = 1,2,\ldots$

In order to go from K to K^N, we stop at an intermediate stage $\tilde{K} = (\omega_2/\Delta)^{Np}$ multiplied by a Δ^{-1} polynomial. The path $K \to \tilde{K}$ defines a first operator $\mathscr{L}_{\mathrm{I}}$ whereas $\tilde{K} \to K^N$ defines the second operator $\mathscr{L}_{\mathrm{II}}$ and $L_{qN} = \mathscr{L}_{\mathrm{I}qN}\mathscr{L}_{\mathrm{II}qN}$. With the help of the first-order operator

$$\left(1 - \frac{\partial_{x_2}}{M}\right)\frac{\omega_2^M}{\Delta^Q}\left(1 + \sum_1^k \frac{B_m}{\Delta^m}\right) = \frac{\omega_2^{M+1}}{\Delta^{Q+1}}\left(\frac{Q}{M} + \sum_1^k \frac{B_m}{\Delta^m}\frac{\Gamma(Q+m+1)}{\Gamma(Q+1)}\right),$$

we build $\mathscr{L}_{\mathrm{I}qN}$ and define $\tilde{K}$, $\mathscr{L}_{\mathrm{II}qN}$

$$\mathscr{L}_{\mathrm{I}qN} = \prod_{p}^{Np-1}\left(1-\frac{\partial_{x_2}}{s}\right),$$

$$\mathscr{L}_{\mathrm{I}qN}K \equiv \tilde{K} = \left(\frac{\omega_2}{\Delta}\right)^{Np}\left(1+N\sum_{m=1}^{k}\frac{b_m}{\Delta^m}\prod_{s=1}^{m-1}\left(\frac{Np+s}{p+s}\right)\right), \tag{3.3}$$

$$\mathscr{L}_{\mathrm{II}qN} = 1+\sum_{1}^{k(N-1)}\Omega_j(\partial_{x_1}+\partial_{x_2})^j,$$

$$\mathscr{L}_{\mathrm{II}}\tilde{K} \equiv K^N.$$

$\mathscr{L}_{\mathrm{I}}$ having $(N-1)p$ terms and $\mathscr{L}_{\mathrm{II}}$ being of order $k(N-1)$, it follows that $p = q/(N-1) - k$. Let us assume that, for the same k and p values, we have found a solution (b_m) common to different $N_i q_i$ values. This means at least $p = q_i/(N_i - 1) - k$, $\forall i$. For the set of equations $(L_{q_iN_i}K = K^{N_i})$ it is clear that if $N_1 < N_i$, $\forall i$, then

$$\mathscr{L}_{\mathrm{I}q_iN_i} = \mathscr{L}_{\mathrm{I}q_1N_1}\prod_{N_1p-1}^{N_ip-1}\left(1-\frac{\partial_{x_2}}{s}\right).$$

Now we want to find both $\mathscr{L}_{\mathrm{II}qN}$ [or (Ω_j)] and K. We must require that the coefficients of any Δ^{-1} power are the same on both the lhs and rhs of (3.3). For Δ^{-1} and Δ^{-2} this leads to the following possibilities: either (i) $b_1 = p$ and $(b_2 - p(p+1)/2)\,(\Sigma 2^j(-1)^j\Omega_j + p(N-1)/(pN+1)) = 0$ or (ii) $\Sigma(-1)^j\Omega_j = 0$ and

$$\sum(-1)^j\Omega_j 2^j\left(\frac{p}{2}-b_1+\frac{b_2}{p+1}\right) - \frac{(N-1)}{Np+1}\left(\frac{b_1^2}{2}-\frac{b_2p}{p+1}\right) = 0.$$

At this stage three ways are open: $b_1 = p$, $b_2 = p(p+1)/2$; only $b_1 = p$ and a relation between the Ω_j's in (i); or in (ii) we have only two relations and neither b_1 nor b_2 are determined.

TABLE III. Bisolitons $K = (\omega_2/\Delta)^p(1+\Sigma_{m=1}^{k}[\Gamma(p+m)/m!\Delta^m\Gamma(p)])$, $\Delta = 1+\Sigma_1^2\omega_i$, $\omega_i = e^{x_i}$.

$L_{qN}K = K^N$, $p = \dfrac{q}{N-1} - k$, k integer

$L_{qN} = \mathscr{L}_{\mathrm{I}qN}\mathscr{L}_{\mathrm{II}qN}$, $\displaystyle\prod_{p}^{Np-1}\left(1-\frac{\partial_{x_2}}{s}\right) = \mathscr{L}_{\mathrm{I}qN}$, $\displaystyle 1+\sum_{1}^{k(N-1)}\Omega_j(\partial_{x_1}+\partial_{x_2})^j = \mathscr{L}_{\mathrm{II}qN}$

$r_{m,k}$ def: $\displaystyle\left[1+\sum_{1}^{k}\frac{\Gamma(p+m)}{\Delta^m m!\Gamma(p)}\right]^N = 1+\sum_{1}^{k}\frac{\Gamma(Np+m)}{m!\Delta^m\Gamma(Np)}+\sum_{k+1}^{kN}\frac{r_{m,k}}{\Delta^m}$

general k case: $\mathscr{S}_j^l = \mathscr{S}_{j-1}^{l-1} - (k+l)\mathscr{S}_{j-1}^l$, $1 \leqslant l \leqslant j$, $\mathscr{S}_j^1 = (-(k+1))^{j-1}$, $\mathscr{S}_j^j = 1$

$\displaystyle\sum_{1}^{k(N-1)}\Omega_l\mathscr{S}_l^1 = r_{k+1,k}$, $\displaystyle\sum_{l=s}^{k(N-1)}\Omega_l\mathscr{S}_l^s = r_{k+s,k}\,k!\Gamma(Np)[\Gamma(Np+k+1)]^{-1}$, $s = 2, 3, \ldots, k(N-1)$

$k=1$, $K = \left(\dfrac{\omega_2}{\Delta}\right)^p\left(1+\dfrac{p}{\Delta}\right)$, $N=2$, $p=q-1$, $q\geqslant 2$, $\Omega_1 = \dfrac{p}{2(2p+1)}$, $N=3$, $p=\dfrac{q}{2}-1$, $q\geqslant 3$, $\Omega_1 - 2\Omega_2 = \dfrac{p}{3p+1}$

$\Omega_2 = \dfrac{p^2}{3(3p+1)(3p+2)}$, $N=4$, $p=\dfrac{q}{3}-1$, $q\geqslant 4$, $\displaystyle\sum_{1}^{3}(-2)^{l-1}\Omega_l = \frac{3p}{2(4p+1)}$, $\Omega_2 - 5\Omega_3 = \dfrac{p^2}{(4p+1)(4kp+2)}$

$\Omega_3 = \dfrac{p^3}{4(4p+1)(4p+2)(4p+3)}$, $N=5$, $p=\dfrac{q}{4}-1$, $q\geqslant 5$, $\displaystyle\sum_{1}^{4}(-2)^{l-1}\Omega_l = \frac{2p}{5p+1}$, $\Omega_2 - 5\Omega_3 + 19\Omega_4 = \dfrac{2p^2}{(5p+1)(5p+2)}$

$\Omega_3 - 9\Omega_4 = \dfrac{p^3}{(5p+1)(5p+2)(5p+3)}$, $\Omega_4 = \dfrac{p^4}{5(5p+1)(5p+2)(5p+3)}$

$k=2$, $K = \left(\dfrac{\omega_2}{\Delta}\right)^p\left(1+\dfrac{p}{\Delta}+\dfrac{p(p+1)}{2\Delta^2}\right)$, $N=2$, $p=q-2$, $q\geqslant 3$, $\Omega_1 - 3\Omega_2 = \dfrac{p}{2(2p+1)}$, $\Omega_2 = \dfrac{p(p+1)}{8(2p+1)(2p+3)}$,

$N=3$, $p=\dfrac{q}{2}-2$, $q\geqslant 5$, $\displaystyle\sum_{1}^{4}(-3)^{l-1}\Omega_l = \frac{2p(4p+3)}{3(3p+1)(3p+2)}$, $\Omega_2 - 7\Omega_3 + 37\Omega_4 = \dfrac{p}{6(3p+2)}$

$\Omega_3 - 12\Omega_4 = \dfrac{p^2(p+1)}{6(3p+1)(3p+2)(3p+4)}$, $\Omega_4 = \dfrac{p^2(p+1)^2}{36(3p+1)(3p+2)(3p+4)(3p+5)}$

$k=3$, $N=2$, $p=q-3$, $q\geqslant 4$, $\Omega_1 - 4\Omega_2 + 16\Omega_3 = \dfrac{(7p+11)}{12}p^2(p+1)$, $\Omega_2 - 9\Omega_3 = \dfrac{p(p+1)}{8(2p+1)(2p+3)}$

$\Omega_3 = \dfrac{p(p+1)(p+2)}{48(2p+1)(2p+3)(2p+5)}$

$L_qK = \left(\sum_i\lambda_i(\partial_{x_1}+\partial_{x_2})\right)^{l_i}L_{q_iN_i}K = \sum\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}$, $N_1 < N_i$

$\dfrac{q_i}{N_i-1} = p+k$, L_q factorizes $L_{1q_1N_1}$

Investigating the $\Delta^{-3}, \Delta^{-4}, \cdots$ coefficients, other ways can be open. In the remaining of the subsection, for simplicity we choose the most simple solution:

$$
\begin{aligned}
&K=\left(\frac{\omega_2}{\Delta}\right)^p F,\\
&F=1+\sum_1^k \Gamma(p+m)(\Gamma(p)\Delta^m m!)^{-1},\\
&\tilde{K}=\left(\frac{\omega_2}{\Delta}\right)^{Np}\tilde{F},\\
&\tilde{F}=1+\sum_1^k \Gamma(Np+m)(\Gamma(Np)\Delta^m m!)^{-1},\\
&F^N=\tilde{F}+R,\\
&R=\sum_{k+1}^{kN}\frac{r_{m,k}}{\Delta^m},\quad R=O\left(\frac{1}{\Delta^{k+1}}\right),
\end{aligned}
\tag{3.4}
$$

$$
\tag{3.5}
$$

which defines the parameters $r_{m,k}$. Substituting this identity, (3.5) into Eq. (3.3), we find after trivial algebra that we must have the condition:

$$
\left(\frac{\omega_2}{\Delta}\right)^{-Np}\sum_1^{k(N-1)}\Omega_j(\partial_{x_1}+\partial_{x_2})^{j-1}(\partial_{x_1}+\partial_{x_2})\left(\frac{\omega_2}{\Delta}\right)^{Np}\tilde{F}
\equiv R\equiv O\left(\frac{1}{\Delta^{k+1}}\right). \tag{3.6}
$$

The ansatz solution, Eq. (3.20), is really a solution if independently of the the values of the set (Ω_j), the lhs of Eq. (3.23) is also of order $O(1/\Delta^{k+1})$. In fact, with the help of

$$
(\partial_{x_1}+\partial_{x_2})\frac{\omega_2^M}{\Delta^Q}=\frac{\omega_2^M}{\Delta^Q}\left(\frac{M-Q}{\Delta^Q}+\frac{Q}{\Delta^{Q+1}}\right)
$$

we find

$$
\left(\frac{\omega_2}{\Delta}\right)^{-Np}(\partial_{x_1}+\partial_{x_2})\left(\frac{\omega_2}{\Delta}\right)^{Np}\tilde{F}=\frac{\Gamma(Np+k+1)}{k!\Delta^{k+1}\Gamma(Np)},
$$

and this proves that the ansatz (3.4) is necessarily a solution, we find

$$
\left(\frac{\omega_2}{\Delta}\right)^{-Np\,k(N-1)-1}\sum_0\Omega_{j+1}(\partial_{x_1}+\partial_{x_2})^j\left(\frac{\omega_2}{\Delta}\right)^{Np}\frac{1}{\Delta^{k+1}}
=\frac{\Gamma(Np)}{\Gamma(Np+k+1)}\sum_{m=k+1}^{kN}\frac{r_{m,k}}{\Delta^m}. \tag{3.7}
$$

In Eq. (3.7) we must apply the differential operator $(\partial_{x_1}+\partial_{x_2})^j$. In the general case we essentially find a Δ^{-1} polynomial where the coefficients S_j^m are some kind of generalized Stirling numbers[6] (Table III)

$$
\left(\frac{\omega_2}{\Delta}\right)^{-Np}(\partial_{x_1}+\partial_{x_2})^j\left(\frac{\omega_2}{\Delta}\right)^{Np}\frac{1}{\Delta^{k+1}}
=\frac{\mathscr{S}_{j+1}^1}{\Delta^{k+1}}+\sum_{m=k+1}^{k+j-1}\frac{\mathscr{S}_{j+1}^{m-k}}{\Delta^m}\frac{\Gamma(Np+m)}{\Gamma(Np+k+1)}.
$$

For each Δ^{-1} power we can now identify the coefficient in both sides of (3.7);

$$
\begin{aligned}
&\sum_0^{k(N-1)}\Omega_l\mathscr{S}_l^1=r_{k+1,k},\\
&\sum_{l=s}^{k(N-1)}\Omega_l\mathscr{S}_l^s=r_{k+s,k}\Gamma(Np)[\Gamma(Np+k+s)^{-1}]k!,\\
&s=2,3,\ldots,k(N-1),\quad r_{k+s,k}\in[r_{k+2,k},r_{kN,k}].
\end{aligned}
\tag{3.8}
$$

We set that Eq. (3.8) gives a set of triangular relations for the (Ω_l) which, for any k, as large as we want, can be explicitly determined.

4. A SUFFICIENT CONDITION FOR THE EXISTENCE OF FACTORIZED L_{qN}

We give a criterion such that an ansatz K corresponds to a factorized L_{qN}.

4.1 Theorem: Let us assume $\omega_i=\exp x_i$, $i=1,2$, and

(i) $K(\omega_1,\omega_2)=\omega_2^{m_2}A(Z=\omega_1+\omega_2,\omega_1,\omega_2)$, $m_2>0$,

(ii) $\partial_{Z^{l+l_1+l_2}\omega_1^{l_1}\omega_2^{l_2}}^{l+l_1+l_2}A(Z,\omega_1,\omega_2)_{\omega_2=0}$ exists $\forall l\geqslant 0$, $\forall l_1\geqslant 0$, $\forall l_2\geqslant 0$,

(iii) $\sum_0^q\lambda_l\left(\omega_1\frac{d}{d\omega_1}\right)^l A(\omega_1,\omega_1,0)=0\Rightarrow\lambda_l=0$, $l=0,\ldots,q$. (4.1)

Then in $L_{qN}K=c_{qN}K^N$, $N>1$, the operator $L_{qN}=\Sigma_{i+j=0}^{i+j=q}a_{ij}\,\partial_{x_1^i x_2^j}^{i+j}$ factorizes $(1-\partial_{x_2}/m_2)$. For the proof we use, when $\omega_2\to 0$ a balance between linear vs nonlinear contributions. For the nonlinear part we find $K^N\simeq_{\omega_2\to 0}\omega_2^{m_2}\times[A(\omega_1,\omega_1,0)]^N$. For K we get, after trivial algebra,

$$
\partial_{x_1^i x_2^j}^{i+j}K\underset{\omega_2\to 0}{\simeq}\omega_2^{m_2}m_2^j\left(\omega_1\frac{d}{d\omega_1}\right)^i A(\omega_1,\omega_1,0),
$$

with the meaning

$$
\left(\omega_1\frac{d}{d\omega_1}\right)^l=\left(\omega_1\frac{d}{d\omega_1}\right)\left(\omega_1\frac{d}{d\omega_1}\right)^{l-1},
$$

and consequently for the linear part

$$
L_{qN}K\underset{\omega_2\to 0}{\simeq}\omega_2^{m_2}\sum_{i+j=0}^{i+j=q}a_{ij}m_2^j\left(\omega_1\frac{d}{d\omega_1}\right)^i A(\omega_1,\omega_1,0).
$$

Compatible behaviors when $\omega_2\to 0$ requires that the coefficient of $\omega_2^{m_2}$ be zero, and from (iii) in (4.1) we find

$$
\sum_{j=0}^{q-i}a_{ij}m_2^j=0,\quad i=0,1,\ldots,q
$$

or

$$
L_{qN}=\left(1-\frac{\partial_{x_2}}{m_2}\right)^{i+k=q-1}\sum_{i+k=0}m_2^{-k}\partial_{x_1^i x_2^k}^{i+k}\sum_{j=0}^k a_{ij}m_2^j. \tag{4.2}
$$

First, we remark that the conditions (4.1) do not ensure the existence of a solution. Secondly, for $N=2$ and larger conditions, the decomposition (4.2) was obtained.[4] Thirdly, the crucial assumption (iii) in (4.1) means that the functions of the set $\{(\omega_1 d/d\omega_1)^l A(\omega_1,\omega_1,0), l=0,\ldots,q\}$ are linearly independent. Finally we note that this result is obtained mainly from linear constraints.

Corollary: Let us assume $K=\omega_1^{m_1}\omega_2^{m_2}\times B(Z=\omega_1+\omega_2,\omega_1,\omega_2)$, where both $\omega_i^{m_i}B=A$, $i=1,2$, satisfy (ii) in (4.1) and both $\omega_1^{m_1}B(\omega_1,\omega_1,0)$ and $\omega_2^{m_2}B(\omega_2,0,\omega_2)$ satisfy (iii) for the set $(\omega_i d/d\omega_i)^l\omega_i^{m_i}B$, then L_{qN} factorizes both $(1-\partial_{x_i}/m_i)$, $i=1,2$.

From the theorem we find $\Sigma_{j=0}^{q-i}a_{ij}m_2^j=0$, $\Sigma_{i=0}^{q-j}a_{ij}m_1^j=0$, or

$$
L_{qN}=\left(1-\frac{\partial_{x_1}}{m_1}\right)\left(1-\frac{\partial_{x_2}}{m_2}\right)\sum_{\substack{p=0\\ i+j=p}}^{q-2}a_{ij}\partial_{x_1^i}\partial_{x_2^j}
\times\left(\sum_{l_1+l_2=0}^{q-2-p}m_1^{-l_1}m_2^{-l_2}\partial_{x_1^{l_1}x_2^{l_2}}^{l_1+l_2}\right). \tag{4.3}
$$

4.2 Applications: (i) If

$$K=\frac{\omega_1^{m_1}\omega_2^{m_2}}{\Delta^{p}}\sum_{1}^{k}\frac{b_m}{\Delta^{m}},\quad \Delta=1+\omega_1+\omega_2,\ m_i>0,$$

the functions $(\omega_i\, d/d\omega_i)^l(\omega_i^{m_i}\Sigma_1^k b_m/(1+\omega_i)m+p)$ are linearly independent because they have poles of order $m+p+l$ at $\omega_i=-1$, and, consequently, for K solutions of $L_{qN}K=c_{qN}K^N$, the operators L_{qN} have at least the factorization (4.2) if $m_2\neq 0$, $m_1=0$, and (4.3) if $m_1\neq 0$, $m_2\neq 0$. We note that the solutions found in Sec. 3 (see Tables II, III) are of this type, but, of course, the complete account of nonlinearity constraints gives further factorization.

(ii) The essential assumption in Eq. (4.1) is that K itself factorizes at least one ω_i with some powers. For $N=2$, $q\leqslant 3$, we know that the only existing bisolitons are either of this type or can be reduced to it using transforms like $K\to K-a_{00}$ or $(\omega_i^{-1},\omega_i^{-1}\omega_j)\to(\omega_i,\omega_j)$. The assumption (i) in (4.1) must be always understood modulo these transforms. For instance, for $N=2$ the ansatz $K=\Sigma_{l=0}^{q-1}\omega_2^l\Sigma_{m=l+1}^{q}b_{l,m}(1+\omega_1+\omega_2)^{-m}$ violates assumption (i) because in the sum for $l=0$, ω_2 is not factorized. However, K can be rewritten $\Sigma_{l=0}^{q-1}\omega_2^{-(m-l)}\Sigma_{m=l+1}^{q}b_{l,m}(1+\omega_2^{-1}+\omega_1\omega_2^{-1})$ factorizing ω_2^{-1}.

(iii) *Conjecture*: All the bisolitons (as defined in Sec. 1) satisfy the assumptions (4.1). Only for $N=2$, $q\leqslant 3$ has this conjecture been proved. Can we prove the conjecture without investigating all (or a lot of) nonlinear constraints? Equivalently can we prove in a simple manner that any ansatz solution violating (4.1) (or not reducing to it by transforms) cannot lead to bisolitons? Unfortunately, the answer is no. Even for the most simple case, $N=2$, $q=2$, we cannot, without detailed calculations (see Appendix, Ref. 4) rule out all the ansatz violating Eq. (4.1).

5. TRIVIAL BISOLITONS COMMON TO NONLINEARITIES K^{N_i} AS SOLUTIONS OF MIXED NONLINEARITIES $\Sigma_i\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}$

A. General considerations

In this section we work with the so-called "trivial bisolitons" of $L_{qN}K=K^N$. They are characterized by various factorization properties:

(i) $K(\omega_1,\omega_2)=K_1(\omega_1)K_2(\omega_2)$, $\omega_i=\exp x_i$, $L_{qN}=l_{q-m}\times(x_1)l_m(x_2)$, m integer $\leqslant q$.

(ii) $l_{qi}(x_i)K_i(\omega_i)=[K_i(\omega_i)]^N$, $i=1,2$, $q_i=q,q-m$.

(iii) Their denominators are the product $\Delta_1^{-p_1}\Delta_2^{-p_2}$ (p_1 and p_2 not necessarily integers for $N\geqslant 3$) where $\Delta_j=1+\omega_j$.

From (ii) we see that $K_i(\omega_i)$ are, in fact, solitons solutions associated to linear operators of order $q-m$ and m. We will choose a "trivial bisoliton" common to different $L_{q_iN_i}K=c_{q_iN_i}K^{N_i}$ equations, and it will be a solution of the mixed nonlinearity,

$$\sum\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}L_{q_iN_i}K$$
$$=L_qK(\omega_1,\omega_2)=\sum c_{q_iN_i}\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}. \tag{5.1}$$

B. $K=\Pi_{j=1,2}\Delta_j^{-p_j}$, $\Delta_j=1+\omega_j$, $\omega_j=e^{x_j}$

We introduce a first-order differential operator

$$\left(1+\frac{\partial_{x_j}}{Q_j}\right)\Delta_j^{-Q_j}=\Delta_j^{-(Q_j+1)} \tag{5.2}$$

TABLE IV. Trivial bisolitons: $K=K_1(\omega_1)K_2(\omega_2)$, $\Delta_j=1+\omega_j$, $\omega_j=e^{x_j}$.

$L_{qN}K=c_{qN}K^N$
$K=(\Delta_1^{P_1}\Delta_2^{P_2})^{-1/(N-1)}$, $1\leqslant P_j$ integers, $P_1+P_2=q$, $c_{qN}=1$ $L_{qN}=\prod_{j=1,2}\prod_{s_j=P_j/(N-1)}^{[N/(N-1)]P_j-1}\left(1+\frac{\partial_{x_j}}{s_j}\right)$
$K=[\omega_2^M\Delta_1^{-P_1}\Delta_2^{-P_2}]^{1/(N-1)}$, M integer $<P_2$, P_j integers, $P_1+P_2=q$ $L_{qN}=\prod_{P_1/(N-1)}^{[N/(N-1)]P_1-1}\left(1+\frac{\partial_{x_1}}{s}\right)\prod_{M/(N-1)}^{NM/(N-1)-1}\left(1-\frac{\partial_{x_2}}{u}\right)\prod_{(P_2-M)/(N-1)}^{[N/(N-1)](P_2-M)-1}\left(1+\frac{\partial_{x_2}}{v}\right)$, $c_{qN}=\frac{\Gamma(NP_2/(N-1))\Gamma(M/(N-1))\Gamma((P_2-M)/(N-1))}{\Gamma(NM/(N-1))\Gamma(N(P_2-M)/N-1))\Gamma(P_2/(N-1))}$
$L_qK=\left(\sum_i\lambda_i(\partial_{x_1}+\partial_{x_2})^{l_i}L_{q_iN_i}\right)K=\sum\lambda_ic_{q_iN_i}(\partial_{x_1}+\partial_{x_2})^{l_i}K^{N_i}$ $N_1<N_i$, $i>1$, $q_1<q_i$, $i>1$
$K=\Delta_1^{-P_1}\Delta_2^{-P_2}=(\Delta_1^{-P_{1i}}\Delta_2^{-P_{2i}})^{1/(N_i-1)}$, $\frac{P_{ji}}{N_i-1}=P_j\ \ \forall i$ L_q factorizes $L_{q_1N_1}$
$K=\omega_2^m\Delta_2^{-P_2}\Delta_1^{-P_1}=[\omega_2^{M_i}\Delta_2^{-P_{2i}}\Delta_1^{-P_{1i}}]^{1/(N_i-1)}$, $\frac{M_i}{N_i-1}=m$, $\frac{P_{ji}}{N_i-1}=p_j\ \ \forall i$, $P_{1i}+P_{2i}=q_i$ L_q factorizes $L_{q_1N_1}$

that we apply on K with Q_i varying from p_j up to $Np_j - 1$ (Table IV).

For instance, in the particular case $P_1 = P_2 = q/2$,

$$K = (1 + \omega_1 + \omega_2 + \omega_1\omega_2)^{-q/2(N-1)},$$

$$L_{qN} = \prod_{j=1,2} \prod_{q/2(N-1)}^{Nq/2(N-1)} \left(1 + \frac{\partial_{x_j}}{m_j}\right), \quad q \geqslant 2.$$

The factorization of L_{qN} into the product of two operators acting on two independent variables is clear because

$$\prod_{P_j/(N-1)^{-1}}^{NP_j/(N-1)-1} \left(1 + \frac{\partial_{x_j}}{m_j}\right)(\Delta_j)^{-P_j/(N-1)} = \Delta_j^{-P_jN/(N-1)}, \quad \Delta_j = 1 + \omega_j;$$

therefore, $\Delta_j^{-P_j/(N-1)}$ is a soliton associated to a linear operator of order P_j.

Now we choose a particular solution $K = \Delta_1^{-P_1}\Delta_2^{-P_2} = [\Delta_1^{P_{1i}}\Delta_2^{P_{2i}}]^{-1/(N_i-1)}$ common to different $L_{q_iN_i} = K^{N_i}$ equations, all the $L_{q_iN_i}$ as well as L_q in Eq. (5.1) factorize $L_{q_1N_1}$ if $N_1 < N_i$ (see Table IV).

C. $K = \omega_2^m \Pi_{j=1,2} \Delta_j^{-P_j}$

We introduce two new first-order differential operators

$$\left(1 - \frac{\partial_{x_2}}{M}\right)\frac{\omega_2^M}{\Delta_2^Q} = \frac{Q}{M}\frac{\omega_2^{M+1}}{\Delta_2^{Q+1}},$$

$$\left(1 + \frac{\partial_{x_2}}{P-M}\right)\frac{\omega_2^M}{\Delta_2^P} = \frac{P}{P-M}\frac{\omega_2^M}{\Delta_2^{P+1}}, \tag{5.3}$$

and with the help of (5.2) and (5.3) we build the nonlinearity K^N (Table IV).

For a common solution $\omega_2^m(\Delta_1^{P_1}\Delta_2^{P_2})^{-1} = [\omega_2^{M_i}\Delta_1^{-P_1}\Delta_2^{-P_{2i}}]^{1/N_i - 1}$ to different $L_{q_iN_i}K = c_{q_iN_i}K^{N_i}$, $N_1 < N_i$, then $L_{q_iN_i}$ factorizes $L_{q_1N_1}$, which is the common factor on the lhs of Eq. (5.1).

6. BISOLITONS COMMON TO DIFFERENT NONLINEARITIES $(G_x)^{N_i}$ AND THE PROBLEM OF THE EXISTENCE OF A COMMON FACTOR FOR THE ASSOCIATED $(L_{q_iN_i})$

A. General considerations

In this section we study the bisolitons of $L_{qN} = \text{const}\, K^{N-1}K_x$, still assuming that their denominators are functions of $\Delta = 1 + \Sigma\omega_i$, without coupling term const $\omega_1\omega_2$. Instead of K, we work with the potentials $G_x = K$ and consider the class of nonlinear equations:

$$L_{qN} = \sum_{i+j=1}^{i+j=q} a_{ij}\,\partial_{x^it^j}^{i+j}, \quad L_{qN}G = c_{qN}(G_x)^N. \tag{6.1}$$

Previous results[3–5] were obtained. Our first aim is to generalize these results in the arbitrary N case. Our second motivation is to look at G solutions common to different integer N_i values which will be solutions of

$$L_qG = \left(\sum_i \lambda_i\,\partial_{x^{l_i}}^{l_i} L_{q_iN_i}\right)G = \sum_i \lambda_i c_{q_iN_i}\,\partial_{x^{l_i}}^{l_i}(G_x)^N. \tag{6.2}$$

As in previous sections, depending on whether the set $(L_{q_iN_i})$ has a common factor or not, then L_q will necessarily be a factorized operator or not.

We recall the general scheme in order to discover the bisolitons.[3]

(i) We define the germ linear differential operator $\tilde{l}_N(x,t)$ of order N and of the type L_{qN} in Eq. (6.1) (with a_{ij} replaced by $\tilde{a}_{ij}$), such that for any differentiable $F(\Delta)$ and any differentiable $\Delta(x,t)$ we have the property

$$[\tilde{l}_N(x,t) - \tilde{a}_{N0}(\Delta_x)^N\partial_\Delta^N]F(\Delta) = 0. \tag{6.3}$$

We further restrict Δ to be

$$\Delta = 1 + \sum_1^2 \omega_i, \quad \omega_i = \exp(t + \gamma_i x),$$

$$\gamma_1 \neq \gamma_2, \quad \Delta_t = \Delta - 1, \tag{6.4}$$

and determine the conditions on the $\tilde{l}_N$ parameters and γ_i in order that both Eqs. (6.3) and (6.4) are satisfied.

(ii) We introduce $l_{q-N}(t)$ as a linear differential operator of order N in $\partial_{t^j}^j$, with constant coefficients and define L_{qN} as the product $\tilde{l}_N l_{q-N}$. Moreover, we assume $G \equiv G(\Delta)$ with Δ written down in Eq. (6.4). From Eq. (6.4), $l_{q-N}(t)$ can be written in terms of Δ, ∂_Δ, and, consequently, $l_{q-N}(G(\Delta))$ depends only on Δ. From Eq. (6.3) we find

$$L_{qN}G = \tilde{a}_{N0}(\Delta_x)^N\partial_\Delta^N(l_{q-N}G(\Delta)) \tag{6.5}$$

(iii) we assume that both $G(\Delta)$ and the operator $l_{q-N}(t)$ satisfy the NLODE

$$\partial_\Delta^N l_{q-N}(t)G(\Delta) = \nu_{qN}(\partial_\Delta G)^N, \tag{6.6}$$

whence the factorized L_{qN} is associated with Eq. (6.1):

$$L_{qN}G = l_{q-N}\tilde{l}_N G = \nu\tilde{a}_{N0}(\Delta_x G_\Delta)^N = c_{qN}(G_x)^N, \quad c_{qN} = \nu_{qN}\tilde{a}_{N0}. \tag{6.7}$$

The resolution of Eq. (6.1) is reduced to the determination of compatible l_{q-N}, G satisfying Eq. (6.6). Equation (6.6), in this section, is thus the key equation. Due to Eq. (6.4) we have for an arbitrary $F(\Delta)$ function $(\partial_t - (\Delta - 1)\partial_\Delta)F(\Delta) = 0$. Iterating, we find the operator identity $\partial_{t^j}^j \equiv \Sigma_{l=1}^j \mathscr{C}_j^l(\Delta - 1)^l\partial_\Delta^l$ where the $\mathscr{C}_j^l$ are the Stirling numbers of the second kind.[6]

B. Germs $\tilde{l}_N(x,t)$

First we start with

$$\tilde{l}_N = \prod_{i+j=1}^{i+j=N-1} \tilde{a}_{ij}\partial_{x^it^j}^{i+j} + \tilde{a}_{N0}\partial_{x^N}^N,$$

$$\tilde{l}_N F(\Delta) = \sum_{p=1}^{N-1} \frac{1}{p!}(\partial_\Delta^p F(\Delta)) \sum_{m=0}^{p-1} (-1)^m c_p^m \Delta^m \tilde{l}_N \Delta^{p-m} + \tilde{a}_{N0}(\Delta_x)^N\partial_\Delta^N F(\Delta). \tag{6.8}$$

A sufficient condition in order that Eq. (6.3) holds is that the coefficients of $F_{\Delta_p}, p = 1,\ldots,N-1$, vanish, or $\tilde{l}_N\Delta^j = 0$, $j = 1,2,\ldots,N-1$. It is trivial to verify that this condition is necessary by assuming $F = \Delta^j, j = 1, \ldots, N-1$.

(i) $N = 2$ and $\tilde{l}_2\Delta = \partial_t + a_{10}\partial_x + a_{20}\partial_{x^2}^2$. The condition $\tilde{l}_2\Delta = 0$ gives $\Delta_t + \tilde{a}_{10}\Delta_x + \tilde{a}_{20}\Delta_{xx} = 0$ or $1 + \tilde{a}_{10}\gamma_i + \tilde{a}_{20}\gamma_i^2 = 0$ and we find a Burgers family of $\tilde{l}_2$ operators with two arbitrary parameters $\tilde{a}_{10}, \tilde{a}_{20}, \tilde{a}_{10}^2 \neq 4\tilde{a}_{20}$.

(ii) $N=3$ and for simplicity, we put $\tilde{a}_{02}=0$,

$$\tilde{l}_3=\partial_t+\frac{9\tilde{a}_{30}}{2\tilde{a}_{20}}\partial_{xt}+\left(\frac{2\tilde{a}_{20}^2}{9\tilde{a}_{30}}-\frac{3\tilde{a}_{30}}{\tilde{a}_{20}}\right)\partial_x+\tilde{a}_{20}\partial_{x^2}^2+\tilde{a}_{30}\partial_{x^3}^3,$$

$\tilde{a}_{20}^3\neq 54\tilde{a}_{20}^2$,

and the two different γ_i values are solutions of a second-degree algebraic equation $\frac{9}{2}\tilde{a}_{30}/\tilde{a}_{20}+\tilde{a}_{20}\gamma_i+3\tilde{a}_{30}\gamma_i^2=0$.

(iii) $N=4$ and for simplicity we put $\tilde{a}_{03}=\tilde{a}_{12}=0$. We start with a seven-parameter family $\tilde{l}_4=\partial_t+\tilde{a}_{02}\partial_{t^2}^2+\tilde{a}_{11}\partial_{xt}^2+\tilde{a}_{21}\partial_{x^2t}^3+\Sigma_{i=0}^4\tilde{a}_{i0}\partial_{x^i}^i$. The conditions $\tilde{l}_4\Delta=\tilde{l}_4\Delta^2=\tilde{l}_4\Delta^3=0$ give, on the Δ derivatives, two linear and one quadratic relation

$$\sum\tilde{a}_{ij}\Delta_{x^it^j}=0,\quad \tilde{a}_{30}\Delta_x+6\tilde{a}_{40}\Delta_{xx}+\tilde{a}_{21}\Delta_t=0,$$

$$\Delta_x(\tilde{a}_{20}\Delta_x+4\tilde{a}_{40}\Delta_{xxx}+\tilde{a}_{11}\Delta_t+2\tilde{a}_{21}\Delta_{xt})+\Delta_{xx}(3\tilde{a}_{30}\Delta_x+3\tilde{a}_{40}\Delta_{xx}+\tilde{a}_{21}\Delta_t)+\tilde{a}_{02}\Delta_t^2=0.$$

For the quadratic relation we require that the coefficients of both $\Sigma\omega_i^2$ and $\omega_1\omega_2$ are zero. After some algebra we find that we have still finally a two-parameter family of $\tilde{l}_4$.

A counting argument can give this number of constraints on the parameters $\tilde{a}_{ij}$. Let us briefly explain for $N=5$. We start with the ansatz provided by Eq. (6.8) supplemented by $\tilde{a}_{22}=\tilde{a}_{13}=\tilde{a}_{04}=0$, and we have 11 $\tilde{a}_{ij}$ parameters. On the other hand, $\tilde{l}_5\Delta=\tilde{l}_5\Delta^4$ give two $\Sigma\omega_i$ equations, $\tilde{l}_5\Delta^2$ as previously gives $\Sigma\omega_i^2$ and $\omega_1\omega_2$, whereas $\tilde{l}_5\Delta^3$ leads to $\Sigma\omega_i^3$ and $\Sigma\omega_i^2\omega_j$. We find six constraints on the $\tilde{a}_{ij}$ from the relations $\Sigma\omega_i$, $\Sigma\omega_i^2$, $\Sigma\omega_i^3$ whereas $\omega_1\omega_2$ and $\Sigma\omega_i^2\omega_j$ lead to three other ones. Finally we obtain nine constraints and still two $\tilde{a}_{ij}$ parameters remain free with two independent γ_i values.

As a final remark, comparing $N=2$ and $N>2$, we see that only in the Burgers $N=2$ case is $\tilde{l}\Delta=0$ a linear differential equation on Δ.

C. Is it always true that the existence of a common solution implies a common factor for the set of associated linear operators?

We recall that our definition of bisolitons for $K=G_x$ is the existence of powers of K which are rational functions of ω_i, $i=1,2$. This means that, for G, a $\log\Delta$ term is allowed, and we must consider two classes of solutions

$$G=\Delta^{-p}\sum_0^k b_m\Delta^{-m}\quad\text{class I},$$

$$G=b_0\log\Delta+\sum_1^k b_m\Delta^{-m}\quad\text{class II}.$$

As we shall see the presence or the absence of a $\log\Delta$ term is the crucial point for the factorization property.

First, if $L_{qN}\equiv\tilde{l}_N$, Eq. (6.6) reduces to $\partial_\Delta^N G=\nu_{qN}(G_\Delta)^N$. For $N=2$ (Burgers case) the only solution is $G=$ const $\log\Delta$, and this too is one of the solutions for higher N values. This means that const $\log\Delta$ is the only common solution to all N integer values. Nevertheless, the corresponding set of germs operators $(\tilde{l}_{N_i})$ does not necessarily have a common factor.

Secondly, $G=\Delta^{-p}$ is a common solution of Eq. (6.6) for different N integer values

$$l_{q-N}=\prod_{s=p}^{Np-1}(1+s^{-1}\partial_t),\quad p=\frac{q-N}{N-1},\quad G=\Delta^{-p},$$

and, if $N_1<N_i$, the set $(l_{q_i-N_i})$ has a common factor $l_{q_1-N_1}$.

As we shall see in the two next sections, the existence or not of a common factor for a set $(l_{q_i-N_i})$ is directly linked to a balance property between lhs and rhs of Eq. (6.6). When substituted into Eq. (6.6), classes I and II lead essentially to Δ^{-1} polynomials on both sides. The factorization property exists when some of the smallest Δ^{-1} power terms present on the lhs are missing on the rhs. This factorization property is explicity obtained by killing these terms. This defines a first operator $\mathscr{L}_{\rm I}$ which acts on the G solution as far as the smallest Δ^{-1} power terms are the same in both sides. The complementary operator $\mathscr{L}_{\rm II}$ reconstructs exactly the nonlinearity and $l_{q-N}=\mathscr{L}_{\rm I}\mathscr{L}_{\rm II}$. As we shall see, $\mathscr{L}_{\rm I}$ disappears for class II, and, consequently, there is no necessary common factorization property.

D. $\Delta^pG(\Delta)$ is a Δ^{-1} polynomial: $G=\Delta^{-p}\Sigma_0^k b_m\Delta^{-m}$, k arbitrary, $b_0=1$

We rewrite Eq. (6.6) in an equivalent form which is more convenient for our study. For that purpose we introduce parameters $r_{m,k}$ by the identity

$$\left[1+\sum_1^k b_m(m+p)(p\Delta^m)^{-1}\right]^N\equiv\sum_0^{kN}r_{m,k}\Delta^{-m},$$

$$r_{0,k}=1,\quad r_{1,k}=Nb_1(p+1)p^{-1},$$

$$r_{2,k}=\frac{N}{2}(b_2(2+p)p^{-1}+(N-1)(b_1p^{-1}(p+1))^2),\ldots,r_{kN,k}=(b_k(p+k)p^{-1})^N.$$

$$l_{q-N}(t)G(\Delta)=\nu_{qN}\Delta^{-pN}p^{-N}X(\Delta^{-1}),$$

$$X(\Delta^{-1})=\sum_0^{kN}\frac{r_{m,k}\Gamma(Np+m)}{\Delta^m\Gamma((N+1)p+m)}.\tag{6.6'}$$

The rhs of Eq. (6.6') is a Δ^{-1} polynomial from Δ^{-Np} up to $\Delta^{-N(p+k)}$ (Np not necessarily integer). $l_{q-N}(t)$ must be an operator $l_{q-N}=1+\Sigma_1^{q-N}\phi_i\partial_{t^j}^i$ such that the lhs of Eq. (6.6') has also the same Δ^{-1} powers with the same coefficients. As a first result we determine the possible p values. Taking into account the relation $\partial_t\Delta^{-Q}=Q\Delta^{-Q}(-1+\Delta^{-1})$ and iterating up to $\partial_{t^j}^j\Delta^{-Q}$, we see that the lhs of (6.6') is a Δ^{-1} polynomial from Δ^{-p} up to $\Delta^{-(p+k+N-q)}=\Delta^{-N(p+k)}$. It follows that $p=(N-q)(N-1)^{-1}-k$.

1. $(1+\partial_t/p)$ is necessarily a factor of l_{q-N}

The power Δ^{-p} present on the lhs of (6.6') is absent on the rhs and the operator l_{q-N} must necessarily kill this term. We note that $\partial_{t^j}^jG=(-p)^j\Delta^{-p}+O(\Delta^{p-1})$ and a first balance between lhs and rhs of Eq. (6.6') gives $(1+\Sigma\phi_j(-p)^j)\Delta^{-p}+O(\Delta^{-(p+1)})\equiv O(\Delta^{-Np})$. It follows that $1+\Sigma\phi_j(-p)^j=0$ or

$$l_{q-N}=\left(1+\frac{\partial_t}{p}\right)\left(1+\sum_1^{q-N-1}\tilde{\phi}_j\partial_{t^j}^j\right).\tag{6.9}$$

This result is very important if there exists a common G solution, $L_{qN}G = c_{qN}(G_x)^N$, to different integer N values, i.e., if the set (b_m) is q,N-independent. In that case, necessarily all the L_{q,N_i} as well as the L_q in Eq. (6.2) have at least this factor $(1+p^{-1}\partial_t)$ in common.

2. l_{q-N} as the product of two operators $\mathscr{L}_{\rm I}$ and $\mathscr{L}_{\rm II}$

We can go on with this balance between lhs and rhs of Eq. (6.6'). Either $p+1=Np$ or $p+1<Np$. In the second case, if l_{q-N} given by Eq. (6.9), when applied to G gives a $\Delta^{-(p+1)}$ term, then $1+\Sigma\tilde{\phi}_j\partial_t^j$ must be such that this term disappears and so on as long as exists in the lhs of (6.6') terms Δ^{-l} with $l<Np$. In this way l_{q-N} becomes the product of two operators. First $\mathscr{L}_{\rm I}$: $\mathscr{L}_{\rm I}G = O(\Delta^{-Np})$ which kill all Δ^{-1} power terms with powers less than Np. Secondly, $\mathscr{L}_{\rm II}$ will be the complementary operator which when applied to $\mathscr{L}_{\rm I}G$ reconstructs exactly the rhs of (6.6'). One can have a great number of different possibilities. In the general case we write

$$\mathscr{L}_{\rm I}G = \Delta^{-Np}\sum_0^M \alpha_m\Delta^{-m}, \quad M=0,1,2,\ldots\leqslant kN,$$
$$\mathscr{L}_{\rm II} = 1+\sum_1^{kN-M}\Omega_j\partial_t^j. \tag{6.10}$$

3. Building of the operator $\mathscr{L}_{\rm II}$

We assume that $\mathscr{L}_{\rm I}G$ has the most general representation written down in Eq. (6.11), and we want to find the relations for the parameters Ω_j of $\mathscr{L}_{\rm II}$.

After some algebra we find in the general case

$$\alpha_j + \sum_{j_1+j_2=j}\frac{\alpha_{j1}}{\Gamma(Np+j_1)}\sum_{l=j_2}^{kN-M}\overline{\Omega}_l\Gamma(Np+l+j_1)(-1)^{l-j_2}c_l^{j_2}$$
$$= \nu_{qN}p^{-N}r_{j,k}\frac{\Gamma(Np+j)}{\Gamma((N+1)p+j)} \tag{6.11}$$

with the constraints

$$\overline{\Omega}_0 = 0, \quad \overline{\Omega}_l = \sum_{j=l}^{kN-M}\Omega_j\mathscr{C}_j^l, \quad \alpha_{j_1}=0 \quad \text{if} \quad j_1>M,$$
$$\overline{\Omega}_l = 0 \quad \text{if} \quad l>kN-M, \quad M<kN.$$

If the set (b_m) is known, then $r_{j,k}$ and α_{j_1} are also known and Eq. (6.11) gives both the corresponding values of $\overline{\Omega}_l$ (or Ω_l) and ν_{qN}. If some of the (b_m) are unknown, we first express $\overline{\Omega}_l$, $r_{j,k}$, α_{j_1} in terms of these unknown parameters. Compatibility condition give the possible (b_m) values and coming back the $\overline{\Omega}_l$ (or Ω_l) values.

The most simple case corresponds to $\alpha_m=0$, except α_0 that we rewrite a α_{qN} in the following subsections. In that case the lhs of (6.11) is simplified, and we find

$$\delta_{j,0} + \sum_{l=j}^{kN}(-1)^{l-j}c_l^j\overline{\Omega}_l\Gamma(Np+l)$$
$$= \frac{\nu_{qN}(\alpha_{qN})^{-1}p^{-N}r_{j,k}\Gamma(Np)}{\Gamma((N+1)p+j)}. \tag{6.12}$$

4. Building of the operator $\mathscr{L}_{\rm I}$ and determination of the set (b_m) for different general classes of solutions

We start with an arbitrary solution having $k+1$ terms

$$G = \Delta^{-p}\sum_0^k\frac{\Gamma(p+m)\tilde{b}_m}{\Gamma(p)m!\Delta^m}, \quad \tilde{b}_0=1,$$

where instead of the set (b_m) we work with an equivalent more convenient set.

We want to describe a general method such that $\mathscr{L}_{\rm I}G$ has only one Δ^{-Np} term, two terms Δ^{-Np} and $\Delta^{-(Np+1)}$, three terms····.

From the general result of subsection 6D1, we know that necessarily $1+p^{-1}\partial_t$ is a factor of $\mathscr{L}_{\rm I}$. We define G_1:

$$\left(1+\frac{\partial_t}{p}\right)G = G_1 = \Delta^{-(p+1)}\sum_0^k\frac{\Gamma(p+1+m)}{\Gamma(p+1)m!}\frac{\tilde{b}_{m,1}}{\Delta^m},$$
$$\tilde{b}_{m,1} = \tilde{b}_m - \tilde{b}_{m+1}. \tag{6.13}$$

For instance, if $\tilde{b}_{m,1}=0$ for $m\neq k$ or equivalently $\tilde{b}_m=1$, we obtain a G solution such that $G_1 = \Delta^{-(p+k+1)}\times\Gamma(p+k+1)(\Gamma(p+1))^{-1}$ and all the $\tilde{b}_m$ (or b_m) are known. With the help of the differential operator $(1+Q^{-1}\partial_t)\Delta^{-Q} = \Delta^{-(Q+1)}$ we can build $\mathscr{L}_{\rm I}$ and obtain a first family of solutions ($k_0=0$ in Table V).

In order to discover other general families of solutions, we successively iterate k_0 times the application $G\to G_1$ in the following way, $p\to p+1$; then $G_1\to G_2,\ldots,p+k_0-1\to -1+k_0$ and then $G_{k_0}\to G_{k_0+1}$. We find

$$\prod_{s=0}^{k_0}\left(1+\frac{\partial_t}{s+p}\right)G = G_{k_0+1}$$
$$= \Delta^{-(p+1+k_0)}\sum_0^k\frac{\Gamma(p+k_0+1+m)}{\Gamma(p+k_0+1)m!\Delta^m}\tilde{b}_{m,k_0+1}, \tag{6.14}$$
$$k_0=0,1,2,\ldots, \quad \tilde{b}_{m,k_0+1} = \sum_{s=0}^{k_0+1}\tilde{b}_{m+s}(-1)^m c_{k_0+1}^m.$$

(i) For instance, if $\tilde{b}_{m,k_0+1}=0$ for $m\neq k$, we still find that G_{k_0+1} has only one $\Delta^{-(p+k+k_0+1)}$ term and from Eq. (6.14) we get the solutions $k_0\neq 0$ of Table V.

(ii) Let us assume that $\tilde{b}_m=1$, except $\tilde{b}_k$ is unknown. G_1 becomes

$$\left(1+\frac{\partial_t}{p}\right)G = G_1 = \frac{\Gamma(p+k)}{\Delta^{p+k}\Gamma(p+1)(k-1)!}$$
$$\times\left(1+\tilde{b}_k+\tilde{b}_k\frac{(p+k)}{k\Delta}\right).$$

We introduce the operator

$$\left(1+\frac{\partial_t}{Q}\right)\Delta^{-Q}(A+B\Delta^{-1})$$
$$= \Delta^{-(Q+1)}(A-BQ^{-1}) + \frac{Q+1}{Q}B\Delta^{-(Q+2)}$$

that we apply k_0 times to G_1. We find

$$\left(1+\frac{\partial_t}{p}\right)\prod_{s=0}^{k_0-1}\left(1+\frac{\partial_t}{p+k+s}\right)G$$
$$= \frac{\Delta^{-(p+k+k_0)}}{\Gamma(p+1)(k-1)!}\Gamma(p+k)$$
$$\times\left[1-\frac{\tilde{b}_k(k+k_0)}{k}+\frac{\tilde{b}_k(p+k+k_0)}{\Delta}\right].$$

TABLE V. Bisolitons $G=\Delta^{-p}(1+\Sigma_1^k b_m/\Delta^m)$, $\Delta=1+\Sigma_1^2\omega_i$, $\omega_i=e^{t+\gamma_i x}$.

$L_{qN}G=c_{qN}[G_x]^N$, $L_{qN}=\tilde{l}_N l_{q-N}\partial_{\Delta^N}^N l_{q-N}G=\nu_{qN}[G_\Delta]^N$, $c_{qN}=\nu_{qN}\tilde{a}_{N0}$

$$p=\frac{q-N}{N-1}-k,\quad l_{q-N}=\mathscr{L}_{\mathrm{I}}\mathscr{L}_{\mathrm{II}},\quad \mathscr{L}_{\mathrm{I}}G=\frac{\alpha_{qN}}{\Delta^{Np}},\quad \mathscr{L}_{\mathrm{II}}=1+\sum_1^{kN}\Omega_j\partial_t^j,$$

$$\text{def } r_{m,k}:\ \left[1+\sum_1^k\frac{b_m}{\Delta^m}\left(1+\frac{p}{m}\right)\right]^N=\sum_0^{kN}\frac{r_{m,k}}{\Delta^m},\quad \mathscr{C}_j^l=\mathscr{C}_{j-1}^{l-1}+l\mathscr{C}_{j-1}^l,\quad \sum_l^{kN}\Omega_j\mathscr{C}_j^l=\bar{\Omega}_l,\quad \bar{\Omega}_0=0$$

$$\delta_{j0}+\sum_{l=j}^{kN}(-1)^{l-j}c_l^j\bar{\Omega}_l\frac{\Gamma(Np+l)}{\Gamma(Np)}=\frac{\nu_{qN}}{\alpha_{qN}}p^{-N}r_{j,k}\frac{\Gamma(Np+j)}{\Gamma((N+1)p+j)},\quad j=0,1,\ldots,kN$$

$$b_m=\frac{\Gamma(p+m)\Gamma(k-m+1+k_0)k!}{\Gamma(p)\Gamma(k-m+1)m!(k_0+k)!},\quad k_0\text{ integer}\geqslant 0,\quad \alpha_{qN}=\frac{\Gamma(p+k+k_0+1)k_0!}{\Gamma(p+k_0+1)(k+k_0)!}$$

$$L_1\prod_{s=p}^{k_0+p}\left(1+\frac{\partial_t}{s}\right)\left[\prod_{u=k_0+p+k+1}^{Np-1}(1+u^{-1}\partial_t)\right]\quad\text{if } q\geqslant(k+1)N+k_0+2$$

$$\left[\prod\cdots\right]\equiv 1\quad\text{if } q=(k+1)N+k_0+1$$

$$b_m=\frac{\Gamma(p+m)}{\Gamma(p)m!}\quad\text{if } m\neq k,\quad b_k=\frac{\Gamma(p+k)}{(k_0+k)(k-1)!\Gamma(p)},\ k_0\text{ integer},\quad \alpha_{qN}=\frac{(p+k+k_0)\Gamma(p+k)}{(k_0+k)(k-1)!\Gamma(p+1)}$$

$$\mathscr{L}_1=\begin{cases}\left(1+\dfrac{\partial_t}{p}\right)\displaystyle\prod_{\substack{s=p+k\\ s\neq p+k+k_0}}^{Np-1}\left(1+\dfrac{\partial_t}{s}\right) & \text{if } q\geqslant(k+1)N+k_0+2\\ \left(1+\dfrac{\partial_t}{p}\right)^{p+k+k_0-1}\displaystyle\prod_{s=p+k}\left(1+\dfrac{\partial_t}{s}\right) & \text{if } q=(k+1)N+k_0+1\end{cases}$$

$L_qG=\sum_i\lambda_i\partial_{x^{j_i}}^{j_i}L_{q_iN_i}G=\sum_i\lambda_i c_{q_iN_i}\partial_{x^{j_i}}^{j_i}[G_x]^{N_i}$, $N_1<N_i$

In all cases L_q factorizes $\mathscr{L}_{\mathrm{I}q_1N_1}$

If $\tilde{b}_k=k\ (k_0+k)^{-1}$, we get the second family solutions of Table V.

E. $G(\Delta)$ mixing $\log\Delta$ and a Δ^{-1} polynomial; violation of the common factor property for a common solution

Here we consider a general solution including a $\log\Delta$ term.

$$G=b_0\log\Delta+\sum_1^k\frac{b_m}{\Delta^m},\quad k\text{ arbitrary}.$$

Both sides of (6.6) are Δ^{-1} polynomials. The rhs is a polynomial with terms from Δ^{-N} up to $\Delta^{-N(k+1)}$ while the lhs is running from Δ^{-N} up to $\Delta^{-(k+q)}=\Delta^{-N(k+1)}$. As a first result, l_{q-N} can be written

$$l_{q-N}(t)=1+\sum_{j=1}^{k(N-1)}\phi_j\partial_t^j,\quad q-N=k(N-1).$$

The second, very important result is that on both sides the smallest Δ^{-1} power is the same Δ^{-N}. Consequently, we no longer have the decomposition $l_{q-N}=\mathscr{L}_{\mathrm{I}}\mathscr{L}_{\mathrm{II}}$. $\mathscr{L}_{\mathrm{I}}$ was the operator killing the smallest Δ^{-1} power terms present on the lhs of Eq. (6.6) and absent on the rhs. The balance lhs vs rhs will not require general conditions independent of the explicit determination of the ϕ_j's. Consequently, a common factor for the set $l_{q_i-N_i}$ associated with a common G solution is not a necessary property, and we will exhibit counter examples to that property.

We rewrite Eq. (6.6) in a form more convenient for the discussion:

$$\text{def: } r_{m,k},\quad \left[1-\sum_1^k\frac{m}{\Delta^m}\frac{b_m}{b_0}\right]^N=\sum_0^{kN}\frac{r_{m,k}}{\Delta^m}=\left(\frac{\Delta}{b_0}\right)^N(G_\Delta)^N,$$

which can be calculated:

$$r_{0,k}=1,\quad r_{1,k}=-Nb_0^{-1}b_1,$$

$$r_{2,k}=Nb_0^{-2}[(N-1)b_1^2-4b_0b_2]\cdots,$$

$$\partial_\Delta^N\left\{G+\sum_0^{(N-1)k-1}\phi_{l+1}\partial_t^l\left(\frac{-(b_0+b_1)}{\Delta}+\sum_2^{k+1}\frac{(m-1)b_m-mb_m}{\Delta^m}\right)\right\}=\nu_{qN}(G_\Delta)^N,$$

$$b_{k+1}=0.\qquad(6.6'')$$

As we said above, both sides have a Δ^{-N} term: the identification fixes ν_{qN}, the scaling parameter, $\nu_{qN}b_0^{N-1}=(-1)^{N+1}(N-1)!$ The next power term is $\Delta^{-(N+1)}$, the identification of the coefficients in both sides lead to $(b_0+b_1)(\Sigma_1^{(N-1)k}(-1)^j\phi_j)=0$. There exists two open ways.

The first one $b_0 + b_1 = 0$, being (ϕ_j) independent does not, of course, require any factorization of l_{q-N}. The second way $\Sigma(-1)^j\phi_j = 0$, though giving relations on the l_{q-N} parameters, does not imply a common factor for a common solution. At the next step $\Delta^{-(N+2)}$ the identification of the coefficients in Eq. (6.6″) give

$$(b_1 - 2b_2)\left(\frac{1-N}{1+N} - \sum(-2)^j\phi_j\right) = 0 \quad \text{if} \quad b_0 + b_1 = 0, \tag{6.15}$$

$$(b_0 + 2b_1 - 2b_2)\sum(-2)^j\phi_j + \frac{1-N}{1+N}\left(2b_2 + \frac{b_1^2}{b_0}\right) = 0 \quad \text{if} \quad \sum(-1)^j\phi_j = 0.$$

If $b_0 + b_1 = 0$, we still have two ways. The first one is $2b_2 = b_1$. In fact, there exists a general solution $mb_m = b_1$ or $G = b_0(\log\Delta - \Sigma_1^k m^{-1}\Delta^{-m})$. The second way leads to an N-dependent relation between the ϕ_j's.

If $\Sigma(-1)^j\phi_j = 0$, then Eq. (6.6″) gives relations between b_0, b_1, b_2, and ϕ_j's. The analysis can be go on identifying all $\Delta^{-(N+m)}$ coefficients on both sides of Eq. (6.6″).

First, we look at the general common solution $mb_m = b_1$ for $k = 1$: $G = -\log\Delta + \Delta^{-1}$, $N = 2$, $l_1 = \partial_t/6$, $N = 3$, $l_2 = 1 + \frac{17}{60}\partial_t + \frac{1}{60}\partial_{t^2}^2$, $N = 4$, $l_3 = 1 + (6/7!)\partial_t(314 + 33\,\partial_t + \partial_{t^2}^2)$. With these examples we can verify that the set $(l_{q-N} = l_{N-2}) = l_1, l_2, l_3, \ldots$ has no common factor while G is a common solution.

Second, for the other way $\Sigma(-1)^j\phi_j = 0$ we still restrict our study to $k = 1$: $N = 3$, $G = \log\Delta + 7\Delta^{-1}$, $N = 4$, $G = \log\Delta + b_1\Delta^{-1}$, $2b_1^2 - 17b_1 - 75 = 0$, and we see that the two solutions are different.

F. Can we have more than bisolitons: $\Delta = 1 + \Sigma_1^{i_0}\omega_i$, $i_0 > 2$?

The answer is no. We recall that $\Delta_t = \Delta - 1$ and for $N = 2,3,4$ the germs $\tilde{l}_N$ are such Δ must satisfy linear relations $\Delta_t + \alpha\Delta_x + \beta\Delta_{xx}$ which forbid for $\omega_i = \exp(t + \gamma_i x)$ to have more than two γ_i values.

7. DISCUSSION

The factorization property or not (the existence of a common factor or not) can be explained by a balance between linear versus nonlinear contributions: We look at $L_{q_iN_i}K = K^{N_i}$ and $\partial_\Delta^{N_i}(l_{q_i-N_i}G) = (G_\Delta)^{N_i}$. Then always $L_{q_iN_i}$ or $l_{q_i-N_i}$ is built as a product of two operators $\mathscr{L}_{\mathrm{I}}\cdot\mathscr{L}_{\mathrm{II}}$. Let us substitute an ansatz possible solution K into the nonlinear part of the equation and compare with the linear contribution. Two situations occur.

(i) *Either* the linear part contains terms which *necessarily* cannot belong to the nonlinear one. It follows that necessarily $L_{q_iN_i}$ or $l_{q_i-N_i}$ (or a part called L_{I}) must kill these terms. Then $L_{\mathrm{I}}K$ has terms similar to those of the nonlinear part, but no terms which necessarily must disappear. Finally the complementary operator L_{II} reconstructs exactly the terms present in the nonlinear part. In this can we have necessarily a factorization property of the linear operator.

(ii) *Or* the linear part does not contain terms which necessarily must disappear in the balance and $L_{q_iN_i}$ or $l_{q_i-N_i}$ is reduced to L_{II}. In this case there is no necessary factorization property.

For a study of a class of nonlinear equations exists in general a basic equation which can be used as a laboratory tool in order to test intrinsic properties of the class; see, for instance, the role of KdV for the class of integrable equations. For K^N nonlinearities, the basic equation is the Boltzmann equation with Maxwell type of interaction and nonlinearity K^2. The most simple case is the so-called Bobylev–Krook–Wu model[7,8] which can be written[4] with a factorized second-order differential operator L_{22}. However, for any q value exist factorized L_{q2} operators, $L_{q2}K = c_{q2}K^2$, corresponding to models of the Boltzmann equation,[9]

$$(1 + \partial_t)\prod_{p=1}^{q-1}\left(1 + \frac{\partial_x}{q+p-2}\right)K(x,t) = K^2,$$

where K is essentially a generalized Laplace transform of the Boltzmann distribution function. For $(G_x)^N$ nonlinearity, the basic equation for the germ $\tilde{l}_2$ is the Burgers equation and the germs $\tilde{l}_N$ for higher N values represent some extension of that equation.

Finally let us recall that the bisolitons of the classical integrable cases, like KdV, are rational functions. In the class studied here for $N > 2$ *they develop other singularities because only some powers are rational functions.* We are trying to include these new features in a general formalism.

[1] V. G. Makhankov, Phys. Rep. C **35**, 1 (1978).
[2] H. Cornille, Phys. Lett. A **91**, 211 (1982).
[3] H. Cornille, J. Phys. Lett. A **15**, L529 (1982).
[4] H. Conrille and A. Gervois, Phys. Lett. A **83**, 121 (1981); Physica D **6**, 1, 50 (1982).
[5] H. Cornille and A. Gervois, Phys. Lett. A **90**, 329; J. Math. Phys. **24**, 2042 (1983).
[6] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* NBS Applied Mathematics Series 55, (Natl. Bur. Stand., Washington, D.C., 1964), p. 824.
[7] A. V. Bobylev, Dokl. Akad. Nauk SSSR **225**, 1296 (1975); **231**, 571 (1976), [Sov. Phys. Dokl. **20**, 822 (1975); **21**, 632 (1976)].
[8] K. M. Krook and T. T. Wu, Phys. Rev. Lett. **16**, 1107 (1976).
[9] Talk at the Conference on Nonlinear Waves and Integrable Systems, Greenville, North Carolina, June 1982.

Lee Hwa Chung theorem for presymplectic manifolds. Canonical transformations for constrained systems

J. Gomis, J. Llosa, and N. Román
Departament de Fisica Teorica, Universitat de Barcelona, Diagonal 647, Barcelona 28, Spain

(Received 25 April 1983; accepted for publication 9 December 1983)

We generalize the analogous of Lee Hwa Chung's theorem to the case of presymplectic manifolds. As an application, we study the canonical transformations of a canonical system (M, S, Ω). The role of Dirac brackets as a test of canonicity is clarified.

PACS numbers: 03.20. + i

1. INTRODUCTION

Degenerate or singular Hamiltonian systems were introduced in mathematical physics by Dirac.[1,2] Later, the interest in this formalism has increased[3,4] mainly because it provides the suitable framework to deal with many physical theories, either in the infinite-dimensional case[2,5] (electromagnetic, gravitational, and Yang–Mills fields) or in the finite-dimensional one (relativistic systems of directly interacting particles).[6]

The usual way of dealing with these systems (we shall restrict ourselves to the finite-dimensional case) starts from a phase space M with a symplectic structure Ω, i.e., a nondegenerate Poisson bracket. Then, some constraints are introduced in order to define a submanifold S representing the set of all possible states of the physical system (e.g., the mass shell constraints in a system of relativistic particles). According to the mutual Poisson brackets, these constraints are classified in first and second class; as is commonly accepted, the latter correspond to spurious degrees of freedom and the former can be considered as generating functions for gauge motions of the system. It has been noticed by Shanmugadhasan[7] that an adapted canonical set of variables can be selected such that the constraints assume their simplest form, i.e., the submanifold S is obtained by making some coordinates and momenta equal to zero. The symplectic form Ω on M then induces a presymplectic structure ω on S, which is degenerate in most cases.

The invariance of the Poincaré–Cartan integral[8] turns out to be a sound principle to establish the features of nondegenerate Hamiltonian systems. Likewise, for degenerate systems, an analogous formalism can be set forth.[9,10] This formulation also stresses the close parallelism between the above-mentioned Shanmugadhasan transformation[7] and the Hamilton–Jacobi method when the canonical Hamiltonian vanishes.[10]

In dealing with Hamiltonian systems, either degenerate or not, it is of great interest to have a precise characterization of canonical transformations. In the nondegenerate case, this characterization can be presented in a very clear and elegant way by requiring the invariance of the Poincaré–Cartan integral under these transformations.[11] To show this result, it is necessary to use Lee Hwa Chung's theorem,[12] which states the only absolute integral invariants under every Hamiltonian system are the symplectic form and the exterior product of this form with itself any number of times.

However, a fully satisfactory definition of canonical transformations has not been attained in the singular case yet. A previous attempt[13] requires canonical transformation to preserve the elementary Poisson brackets (that is, the symplectic form Ω) on the constraint submanifold S. This condition is too strong because the initial phase space M and its symplectic structure Ω must not be considered as the physically relevant objects of the theory, but merely as the starting point to build up the submanifold S and the induced presymplectic structure ω, in terms of which the physical system is represented.

Our standpoint is that, in the degenerate Hamiltonian formalism, canonical transformations must be characterized as those preserving the physically significant submanifold S and its presymplectic structure ω. Thus, a result generalizing Lee Hwa Chung's theorem will a helpful tool in dealing with canonical transformations in the degenerate case. This is the purpose of the present paper.

The language used throughout this work is geometrical[14–18] where concepts, which classically lead to more or less involved equations in terms of coordinates and momenta, can be formulated and handled in a more clear and concise way.

In Sec. 2 we generalize to presymplectic manifolds some well-established results in the symplectic case, such as the concepts of the Hamiltonian vector field, Hamiltonian functions, and the Poincaré–Cartan integral. In Sec. 3, we generalize the above mentioned Lee Hwa Chung's theorem to the presymplectic case. In Sec. 4, previous results are applied to the particular case that the presymplectic manifold has been obtained by imposing a set of constraints on a bigger symplectic space, as is common in the physical applications of the degenerate Hamiltonian formalism. Section 5 is devoted to study the canonical transformations of a constrained Hamiltonian system.

2. HAMILTONIAN FIELDS AND THE POINCARÉ INTEGRAL IN A PRESYMPLECTIC MANIFOLD

Let us consider a manifold S and a presymplectic form on S (i.e., ω is a closed differential 2-form of constant class on S), so that the couple (S, ω) is called a presymplectic manifold.

The 2-form ω defines a differentiable linear map from the tangent vector fields $D(S)$ onto the differential 1-forms $\Lambda^1(S)$, given by

$$D(S) \to \Lambda^1(S),$$
$$\mathbf{X} \rightsquigarrow i_{\mathbf{X}}\omega, \tag{2.1}$$

where $i_{\mathbf{X}}$ denotes the inner product by $\mathbf{X}$. This mapping can-

 0022-2488/84/051348-08$02.50

not be inverted, in general. That is, there are some 1-forms in $\Lambda^1(S)$ which are not in correspondence with any vector field in $D(S)$.

Definition 2.1: A function $h \in \Lambda^0(S)$ is called a *Hamiltonian function* relative to the presymplectic structure defined by ω iff there exists a vector field $\mathbf{X} \in D(S)$ such that

$$i_{\mathbf{X}}\omega = dh. \tag{2.2}$$

In other words, h is a Hamiltonian function iff the linear system (2.2) admits a nonempty family of solutions $\mathbf{X} \in D(S)$. Notice that if $\mathbf{X}_h$ is a particular solution of (2.2), then the general solution $K_h \subset D(S)$ is given by

$$K_h = \mathbf{X}_h + \mathscr{A}(\omega), \tag{2.3a}$$

where $\mathscr{A}(\omega)$ is the submodulus of $D(S)$ defined by

$$\mathscr{A}(\omega) = \{\mathbf{X} \in D(S) / i_{\mathbf{X}}\omega = 0\}. \tag{2.3b}$$

Definition 2.2: A vector field $\mathbf{X} \in D(S)$ is called a *Hamiltonian vector field* relatively to ω iff there exists a function $h \in \Lambda^0(S)$ such that

$$i_{\mathbf{X}}\omega = dh. \tag{2.4}$$

Any Hamiltonian field $\mathbf{X}$ satisfies

$$di_{\mathbf{X}}\omega = 0. \tag{2.5}$$

The converse is not true, since no closed differential form is exact. However, the Poincaré lemma[19] ensures that if Eq. (2.5) is fulfilled, then, for any $p \in S$, there exists an open neighborhood of p, $U \subset S$, and a function $f \in \Lambda^0(U)$ such that

$$i_{\mathbf{X}}\omega - df = 0 \quad \text{on } U. \tag{2.6}$$

This is the reason we introduce the concept of a local Hamiltonian vector field.

Definition 2.3: A vector field $\mathbf{X} \in D(S)$ is called *locally Hamiltonian* relatively to ω iff $i_{\mathbf{X}}\omega$ is closed [i.e., it is a solution of Eq. (2.5)].

Proposition 2.4: $\mathbf{X} \in D(S)$ is locally Hamiltonian relatively to ω iff $L_{\mathbf{X}}\omega = 0$, where L denotes the Lie derivatives.

The proof is immediate after recalling that $L_{\mathbf{X}} = i_{\mathbf{X}}d + di_{\mathbf{X}}$ and that ω is a closed 2-form.

Expressing the class of locally Hamiltonian vector fields relative to ω by

$$D_\omega(S) \equiv \{\mathbf{X} \in D(S) / L_{\mathbf{X}}\omega = 0\}, \tag{2.7}$$

we have the following interesting results:

Proposition 2.5: $D_\omega(S)$ is closed under the Lie bracket.

The proof is immediate since

$$L_{[\mathbf{X},\mathbf{Y}]} = L_{\mathbf{X}}L_{\mathbf{Y}} - L_{\mathbf{Y}}L_{\mathbf{X}}$$

(see, for example, Ref. 19).

Proposition 2.6: Given any point $p \in S$ and any tangent vector at p, $\mathbf{V} \in T_p(S)$, there is at least one locally Hamiltonian vector field $\mathbf{X} \in D_\omega(S)$ such that $\mathbf{X}_p = \mathbf{V}$.

Proof: Using a well-known result in differential geometry,[19] a local chart $(y^1, \ldots, y^s)$ in a neighborhood of $p \in S$ exists such that the presymplectic form ω can be written as

$$\omega_3 = dy^1 \wedge dy^{R+1} + \cdots + dy^R \wedge dy^{2R}, \quad 2R \leqslant s.$$

It is obvious that the vector fields $\partial/\partial y^1, \ldots, \partial/\partial y^s$ are locally Hamiltonian.

For a given $\mathbf{V} \in T_p(S)$, we expand it in the coordinate basis $\{(\partial/\partial y^i)_p\}_{p=1,\ldots,s}$, thus obtaining

$$\mathbf{V} = \sum_{i=1}^{s} V^i \left(\frac{\partial}{\partial y^i}\right)_p, \quad V^i \in \mathbb{R}.$$

Then, a solution for our problem is

$$\mathbf{X} = \sum_{i=1}^{s} V^i \frac{\partial}{\partial y^i}. \qquad \boxtimes$$

At this point it must be noted that any function $f \in \Lambda^0(S)$ depending on any of the variables y^i, $i > 2R$ (i.e., $\partial f/\partial y^i \neq 0$), cannot be a Hamiltonian function.

We are now going to generalize the Poincaré integral theorem for presymplectic manifolds. With this purpose we introduce the concept of a local one-parameter group of diffeomorphisms which consists of a differentiable mapping

$$\Phi: W \to S,$$

$$(t, p) \rightsquigarrow \varphi_t(p),$$

where $W \subset R \times S$ is an open neighborhood of $\{0\} \times S$, having the following properties:

(i) $\forall p \in S$, $(R \times \{p\}) \cap S$ is connected;

(ii) $\forall p \in S$, $\varphi_0(p) = p$;

(iii) If (t', p), $(t + t', p)$, and $(t, \varphi_{t'}(p))$ belong to W, then $\varphi_{t+t'}(p) = \varphi_t(\varphi_{t'}(p))$.

It is also known[19] that the integral orbits of a given differentiable vector field $\mathbf{X} \in D(S)$ permit to define the so-called local one-parameter group generated by $\mathbf{X}$.

Let c be a singular 2-cube on S.[20] Since $|c|$, the support of c, is a compact set in S, there exists a real positive number $\mathscr{E}_c$ such that $[-\mathscr{E}_c, \mathscr{E}_c] \times |c| \subset W$. Therefore, for any $t \in [-\mathscr{E}_c, \mathscr{E}_c]$, the map $\varphi_t: |c| \to S$ defines a diffeomorphism between $|c|$ and $\varphi_t|c|$, and $\varphi_t \circ c$ also is a singular 2-cube on S.

Definition 2.7: Let c be a singular 2-cube on S, $\mathbf{X} \in D(S)$, a differentiable vector field and Φ the local one-parameter group generated by $\mathbf{X}$. For any $t \in [-\mathscr{E}_c, \mathscr{E}_c]$, the Poincaré integral is defined as

$$I(t;c,\mathbf{X}) \equiv \int_{\varphi_t \circ c} \omega.$$

Theorem 2.8: The following conditions are equivalent:

(i) $\mathbf{X} \in D_\omega(S)$ (i.e., it is a locally Hamiltonian field);

(ii) $[dI(t;c,\mathbf{X})/dt]_{t=0} = 0$ for any singular 2-cube on S.

Proof: We have that

$$I(t;c,\mathbf{X}) = \int_{\varphi_t \circ c} \omega = \int_c \varphi_t^* \omega,$$

where φ_t^* is the pullback map associated with φ_t. Deriving then both sides with respect to t and taking $t = 0$, we have

$$\left[\frac{dI(t;c,X)}{dT}\right]_{t=0} = \int_c \lim_{t \to 0}\left(\frac{\varphi_t^* \omega - \omega}{t}\right),$$

but the limit on the right-hand side is nothing but $L_{\mathbf{X}}\omega$. Hence we can write

$$\left[\frac{dI(t;c,X)}{dt}\right]_{t=0} = \int_c L_{\mathbf{X}}\omega,$$

and the theorem follows immediately. $\boxtimes$

3. LEE HWA CHUNG'S THEOREM FOR PRESYMPLECTIC MANIFOLDS

In the case of symplectic manifolds, Lee Hwa Chung's theorem[12] fixes the class of differential forms which are in-

variant under any locally Hamiltonian vector fields. In this section, we arrive at a similar result for presymplectic manifolds.

Theorem 3.1: Let S be a s-manifold, $\omega\in\Lambda^2(S)$ a presymplectic form with constant class $2R$ on S, and α a differential p-form on S, $\alpha\in\Lambda^p(S)$. If $L_{\mathbf{X}}\alpha=0$, for every locally Hamiltonian vector field $\mathbf{X}\in D_\omega(S)$, then

(a) $\alpha=0$, if $p>2R$;

(b) $\alpha=0$, if $p=2l+1$, $l>R$;

(c) $\alpha=f\cdot\omega\wedge\overset{l}{\cdots}\wedge\omega$, if $p=2l$, $l\leqslant R$, where $f\in\Lambda^0(S)$, and is constant on any connected components of S.

Proof: Recalling Eq. (2.3b) and that $L_{\mathbf{X}}=di_{\mathbf{X}}+i_{\mathbf{X}}d$, we have that $\mathscr{A}(\omega)\subset D_\omega(S)$. The hypothesis of the theorem therefore implies that

$$L_{\mathbf{X}}\alpha=0,\quad \forall\mathbf{X}\in\mathscr{A}(\omega). \tag{3.1}$$

Moreover, since $\mathscr{A}(\omega)$ is closed under product by any $f\in\Lambda^0(S)$, we also have that

$$L_{f\mathbf{X}}\alpha=0,\quad \forall\mathbf{X}\in\mathscr{A}(\omega)\ \text{and}\ \forall f\in\Lambda^0(S). \tag{3.2}$$

Comparing both Eqs. (3.1) and (3.2), we arrive at

$$df\wedge i_{\mathbf{X}}\alpha=0,\quad \forall f\in\Lambda^0(S)\ \text{and}\ \forall\mathbf{X}\in\mathscr{A}(\omega). \tag{3.3}$$

For the sake of simplicity, we shall handle Eq. (3.3) at each point $z\in S$ [recall that $T_z(S)$ is a finite-dimensional real vector space, whereas $D(S)$ is a modulus on $\Lambda^0(S)$]. Applying Corollary A.4 of the Appendix, we have

$$i_{\mathbf{X}_z}\alpha_z=0,\quad \forall\mathbf{X}_z\in\mathscr{A}(\omega_z),\ \forall z\in S, \tag{3.4}$$

where $\mathscr{A}(\omega_z)=\{\mathbf{X}_z\in T_z(S)/i_{\mathbf{X}_z}\omega_z=0\}$.

For $p>2R$, using Corollary A.2, we obtain

$$\alpha_z=0,\quad \forall z\in S, \tag{3.5}$$

and, equivalently, $\alpha=0$, so that statement (a) is proved.

Let us now take two locally Hamiltonian vector fields $\mathbf{X},\mathbf{Y}\in D_\omega(S)$. The Poincaré lemma guarantees that, $\forall z\in S$, there exists an open neighborhood of z, $U\subset S$, and two functions, $f, g\in\Lambda^0(U)$, such that

$$i_{\mathbf{X}_U}\omega_U=df\quad\text{and}\quad i_{\mathbf{Y}_U}\omega_U=dg, \tag{3.6}$$

where $\mathbf{X}_U\equiv\mathbf{X}\circ j_U\in D(U)$, $\mathbf{Y}_U\equiv Y\circ j_U\in D(U)$, $\omega_U\equiv\omega\circ j_U\in\Lambda^2(U)$, and j_U: $U\to S$ is the natural injection.

Note that $\mathbf{P}\equiv g\mathbf{X}_U+f\mathbf{Y}_U$ is a Hamiltonian vector field on U relative to ω_U and its corresponding Hamiltonian function is $fg\in\Lambda^0(U)$. Hence, by hypothesis, the differential form $\alpha_U=\alpha\circ j_U\in\Lambda^p(U)$ is invariant under $\mathbf{X}_U$, $\mathbf{Y}_U$, and $\mathbf{P}$, i.e.,

$$L_{\mathbf{X}_U}\alpha_U=L_{\mathbf{Y}_U}\alpha_U=L_{\mathbf{P}}\alpha_U=0, \tag{3.7}$$

which implies

$$df\wedge i_{\mathbf{Y}_U}\alpha_U+dg\wedge i_{\mathbf{X}_U}\alpha_U=0. \tag{3.8}$$

Using (3.6), specializing (3.8) at the point $z\in U$, and taking Proposition 2.6 into account, we can finally write

$$i_{\mathbf{X}_z}\omega_z\wedge i_{\mathbf{Y}_z}\alpha_z+i_{\mathbf{Y}_z}\omega_z\wedge i_{\mathbf{X}_z}\alpha_z=0,\quad \forall z\in S\ \text{and}\ \forall\mathbf{X}_z,\ \mathbf{Y}_z\in T_z(S). \tag{3.9}$$

At this point, the following result is needed:

Lemma 3.2: Let α_z be a p-form on $T_z(S)$ [i.e., $\alpha_z\in\Lambda^p_z(S)$], $p\leqslant 2R$, such that

(i) $\forall\mathbf{X}_z\in\mathscr{A}(\omega_z)$, $i_{\mathbf{X}_z}\alpha_z=0$, (3.4)

(ii) $\forall\mathbf{X}_z,\mathbf{Y}_z\in T_z(S)$,

$$i_{\mathbf{X}_z}\omega_z\wedge i_{\mathbf{Y}_z}\alpha_z+i_{\mathbf{Y}_z}\omega_z\wedge i_{\mathbf{X}_z}\alpha_z=0. \tag{3.9}$$

Then,

—if $p=1$, then $\alpha_z=0$.

—if $p>1$, then there exists $\alpha^{(1)}_z\in\Lambda^{p-2}_z(S)$, fulfilling conditions (3.4) and (3.9) and also

$$i_{\mathbf{X}_z}\alpha_z=i_{\mathbf{X}_z}\omega_z\wedge\alpha^{(1)}_z,\quad \forall\mathbf{X}_z\in T_z(S). \tag{3.10}$$

(The proof is given in the Appendix.)

This lemma provides a recursive algorithm to prove statements (b) and (c) of Theorem 3.1.

(c) Indeed, let us consider the case $p=2l$, $l\leqslant R$.

Starting from $\alpha^{(p)}_z\equiv\alpha_z\in\Lambda^p_z(S)$ and by iterative application of Lemma 3.2, we obtain a sequence of alternated forms:

$$\alpha^{(i)}_z\in\Lambda^{2(l-i)}_z(S),\quad i=0,\ldots,l,$$

such that

$$i_{\mathbf{X}_z}\alpha^{(i)}_z=i_{\mathbf{X}_z}\omega_z\wedge\alpha^{(i+1)}_z,\quad \forall\mathbf{X}_z\in T_z(S)$$

whence, by induction, it follows immediately that

$$\alpha^{(i)}_z=[1/(l+i)!]\alpha^{(l)}_z\omega_z\wedge\overset{l-i}{\cdots}\wedge\omega_z, \tag{3.11}$$

where $\alpha^{(l)}_z\in\Lambda^0_z(S)$ is a real number.

Therefore, the differential p-form $\alpha\in\Lambda^p(S)$ can be written as

$$\alpha=f\cdot\omega\wedge\overset{l}{\cdots}\wedge\omega, \tag{3.12}$$

where $f\in\Lambda^0(S)$.

Finally, since α and ω are invariant under any locally Hamiltonian vector field, we have

$$\mathbf{X}f=0,\quad \forall\mathbf{X}\in D_\omega(S),$$

which, by Proposition 2.6, implies that

$$\mathbf{X}_z f=0,\quad \forall\mathbf{X}_z\in T_z(S)\ \text{and}\ \forall z\in S.$$

Hence, f is constant on any connected component of S, and the proof of the statement (c) is over.

(b) For the case $p=2l+1$, $l<R$, starting from $\alpha^{(0)}_z\equiv\alpha_z$, Lemma 3.2 yields again a sequence of alternated forms:

$$\alpha^{(i)}_z\in\Lambda^{2(l-i)+1}_z(S),\quad i=0,\ldots,l,$$

such that

$$i_{\mathbf{X}_z}\alpha^{(i)}_z=i_{\mathbf{X}_z}\omega_z\wedge\alpha^{(i+1)}_z,\quad \forall\mathbf{X}_z\in T_z(S). \tag{3.13}$$

Since each one of these $\alpha^{(i)}$ fulfills the hypothesis of the Lemma 3.2, and $\alpha^{(l)}_z\in\Lambda^1_Z(S)$, it follows that

$$\alpha^{(l)}_z=0,\quad \forall z\in S,$$

which, by virtue of (3.13), implies $\alpha^{(i)}_z=0$, $i=1,\ldots,l$ and particularly, $\alpha_z=0$, $\forall z\in S$.

Thus, the proof is concluded.

4. CANONICAL SYSTEMS

A canonical system is characterized by a triplet (M, S, Ω), (M,Ω) being a symplectic $2n$-manifold and S an s-submanifold of M. We shall denote the natural injection by j_S: $S\to M$ and the corresponding pullback mapping by j_S^*: $\Lambda(M)\to\Lambda(S)$. We shall assume the closed differential 2-form

$$\omega\equiv j_S^*\Omega\in\Lambda^2(S)$$

has constant rank: rank $\omega = 2R$.

The relationships between $2n$, s, and $2R$ determine the class of the submanifold: namely,

(i) S is said to be a first class submanifold when $2n - s \neq 2R$,

(ii) S is second class when $s = 2R$,

(iii) otherwise S is a mixed submanifold.

This classification is equivalent to the usual one, which is formulated in terms of the Poisson brackets among the constraints defining S.

In this section we are going to particularize the results obtained in the preceeding sections to the case of canonical systems.

Definition 4.1: Let α be a differential p-form on M, $\alpha \in \Lambda^p(M)$. We shall refer to the p-form $j_\alpha^* \in \Lambda^p(S)$ as the *specialization* of α onto S.

Definition 4.2: Two p-forms $\alpha, \beta \in \Lambda^p(M)$ are said to be *weakly equal* on S iff $\alpha \circ j_S = \beta \circ j_S$. Weak equality shall be denoted hereafter by $\alpha \underset{S}{=} \beta$ or $\alpha \approx \beta$.

Definition 4.3: Two p-forms $\alpha, \beta \in \Lambda^p(M)$ are said to be *strongly equal* on S iff $\alpha \underset{S}{=} \beta$ and $d\alpha \underset{S}{=} d\beta$. Throughout this paper we shall write $\alpha \equiv \beta$ for strong equality.

Proposition 4.4: Let ζ^ν, $\nu = 1,...,n-s$, be a set of independent constraint functions defining S, and let $\alpha \in \Lambda^p(M)$. Then, the specialization of α onto S vanishes ($j_S^* \alpha = 0$) if, and only if, there exist $n - s$ differential $(p-1)$-forms $\eta_0 \in \Lambda^{p-1}(M)$, $\nu = 1,...,n-s$, such that

$$\alpha \underset{S}{=} \sum_{\nu=1}^{n-s} \eta_\nu \wedge d\zeta^\nu.$$

Proof: See, for example, Ref. 15.

Definition 4.5: Let $\mathbf{X} \in D(M)$ be a vector field on M. $\mathbf{X}$ is said to be *tangent* to the submanifold S iff

$$\mathbf{X}\zeta \underset{S}{=} 0 \tag{4.1}$$

for any given $\zeta \in \Lambda^0(M)$ such that $\zeta \underset{S}{=} 0$.

We shall denote by $\overline{D(S)}$ the set of those vector fields which are tangent to S.

Condition (4.1) is equivalent to

$$\mathbf{X} \circ j_S \in j_S^T(D(S)).$$

Hence, given any $X \in \overline{D(S)}$, there is a vector field on $D(S)$, which we shall denote by $\mathbf{X}^S$, such that

$$\mathbf{X} \circ j_S = j_{S*}(\mathbf{X}^S), \tag{4.2}$$

i.e., $\forall f \in \Lambda^0(M)$, $\mathbf{X}^S(j_S^* f) = j_S^*(\mathbf{X}f)$, where j_{S*} denotes the Jacobian map.

Proposition 4.6: Let $\mathbf{X} \in \overline{D(S)}$, $\alpha \in \Lambda^p(M)$, then

(i) $j_S^* i_{\mathbf{X}} \alpha = i_{\mathbf{X}^S}(j_S^* \alpha)$,

(ii) $j_S^* L_{\mathbf{X}} \alpha = L_{\mathbf{X}^S}(j_S^* \alpha)$.

The proof follows immediately using very well-known results of differential geometry (see, for example, Ref. 19).

Let $\mathbf{H} \in \overline{D(S)}$ and $\mathbf{H}^S \in D(S)$ be the vector field associated with $\mathbf{H}$ by Eq. (4.2). According to Definition 2.2, $\mathbf{H}^S$ will be Hamiltonian relative to $j_S^* \Omega$ iff a function $f \in \Lambda^0(S)$ exists such that

$$i_{\mathbf{H}^S}(j_S^* \Omega) = df. \tag{4.3}$$

Let us now pick up a function $h_c \in \Lambda^0(M)$ such that $f = j_S^* h_C$ (there is a large class of them) and rewrite (4.3) as

$$j_S^*(i_{\mathbf{H}} \Omega - dh_c) = 0. \tag{4.4}$$

Definition 4.7: A vector field $\mathbf{H} \in D(M)$ is said to be *weakly Hamiltonian* relatively to the canonical structure (M, S, Ω) iff

(i) $\mathbf{H} \in \overline{D(S)}$,

(ii) there exists $h_c \in \Lambda^0(M)$ such that Eq. (4.4) holds.

The function h_c is commonly called a *canonical Hamiltonian* corresponding to $\mathbf{H}$.

We are now going to express Eq. (4.4) in terms of a given set of $2n - s$ constraints defining the submanifold S. These constraints can be always arranged in such a way that:

(i) Some of them $\{\xi^\rho, \rho = 1,...,l\}$, which are said to be *first class*, have weakly vanishing Poisson brackets with any other constraint, and

(ii) the remaining $2n - s - l$ $\{\chi^\alpha, \alpha = 1,...,2n-s-l\}$, which are called *second class* and must be in even number, satisfy the inequality

$$\det(\{\chi^\alpha, \chi^\beta\})_{\alpha,\beta=1,...,2n-s-l} \neq 0. \tag{4.5}$$

Using Proposition 4.4, one can immediate see that condition (4.4) is equivalent to require that $2n - s$ functions a_ρ, $b_\alpha \in \Lambda^0(S)$ exist ($\rho = 1,...,l$; $\alpha = 1,...,2n-s-l$), such that

$$i_{\mathbf{H}} \Omega \approx dh_c + a_\rho \, d\xi^\rho + b_\alpha \, d\chi^\alpha, \tag{4.6}$$

where the sume convention has been used. Realize that the role played by the a_ρ's and b_α's is quite similar to the Lagrange multipliers in many geometrical problems involving the specialization of a differential form onto a submanifold.

Furthermore, condition (i) in definition 4.7 must be also be taken into account. In terms of the constraints, this condition reads:

$$\mathbf{h}\xi^\rho \approx 0 \quad \text{and} \quad \mathbf{H}\chi^\alpha \approx 0,$$

which, according to (4.6), can also be written as

$$\{h_c, \xi^\rho\} \approx 0, \quad \rho = 1,...,l, \tag{4.7}$$

$$\{h_c, \chi^\xi\} + b_\beta \{\chi^\beta, \chi^\alpha\} \approx 0, \quad \alpha, \beta = 1,...,2n-s-l. \tag{4.8}$$

Hence, Eq. (4.7) delimitates the domain of canonical Hamiltonians, and Cramer's linear system (4.8)—remember Eq. (4.5)—determines the Lagrange multipliers b_α, $\alpha = 1,...,2n-s-l$, which are associated with the second-class constraints

$$b_\alpha \approx -c_{\beta\alpha}\{\chi^\alpha, h_c\}, \tag{4.9}$$

where $c_{\beta\alpha}$ is the inverse matrix of $\{\chi^\alpha, \chi^\gamma\}$, i.e.,

$$c_{\beta\alpha}\{\chi^\alpha, \chi^\gamma\} = \delta_\beta{}^\gamma.$$

Substituting (4.9) into (4.6), we arrive at

$$i_{\mathbf{H}} \Omega \approx dh_c + a_\rho \, d\xi^\rho - \{h_c, \chi^\alpha\} c_{\alpha\beta} \, d\chi^\beta, \tag{4.10}$$

where the indeterminacy associated with the first-class constraints appear manifestly.

Summarizing, given a function $h_c \in \Lambda^0(M)$:

(i) The linear system (4.4) will have a nonempty solution if, and only if, condition (4.7) is fulfilled.

(ii) The indeterminacy of the solution $\mathbf{H}$ is related to the arbitrary Lagrange multipliers associated with the first-class constraints.

An explicit expression for $\mathbf{H}$ is

$$\mathbf{H} \approx \{\cdot, h_c^*\} + \{\cdot, a_\rho \xi^\rho\}, \tag{4.11}$$

where the "star function" h_c^* means[21]

$$h_c^* = h_c - \{h_c, \chi^\alpha\} c_{\alpha\beta} \chi^\beta.$$

Definition 4.8: A vector field $\mathbf{H} \in D(M)$ is said to be *locally weakly Hamiltonian* relative to (M, S, Ω) iff

(i) $\mathbf{H} \in \overline{D(S)}$,

(ii) $j_S^*(i_{\mathbf{H}}\Omega)$ is a closed differential form on S.

The latter condition is equivalent to

$$L_{\mathbf{H}^S}(j_S^*\Omega) = 0. \tag{4.12}$$

The proof is similar to the one of Proposition 2.1.

A result analogous to the Poincaré integral theorem holds as well for locally weakly Hamiltonian systems:

Theorem 4.9: Let φ_T be the local one-parameter group generated by a given $\mathbf{H} \in \overline{D(S)}$. Thus, $\mathbf{H}$ is locally weakly Hamiltonian if, and only if,

$$\left[\frac{d}{dt}\int_{\varphi_t \circ c} j_S^*\Omega\right]_{t=0} = 0 \tag{4.13}$$

for any singular 2-cube on $S \subset M$.

We finally have that the generalization of Lee Hwa Chung's theorem proved in the Sec. 3 leads to the following:

Theorem 4.10: Let (M, S, Ω) be a canonical system and $\alpha \in \Lambda^p(M)$ such that

$$L_{\mathbf{H}^S}(j_S^*\alpha) = 0$$

for every locally weakly Hamiltonian field $\mathbf{H}$. Thus we have that:

(i) If either $p > 2R$ or $p = 2l + 1$, $l < R$, then $j_S^*\alpha = 0$.

(ii) If $p = 2l$, $l < R$, then a function $f \in \Lambda^0(M)$ exists such that

$$j_S^*(\alpha - f\Omega \wedge \overset{l}{\cdots} \wedge \Omega) = 0$$

and that $f \circ j$ is a constant on any connected component component of S.

5. CANONICAL TRANSFORMATIONS

Throughout this section, (M, S, Ω) and (M', S', Ω') will denote two given canonical systems and $j_S: S \hookrightarrow M$ and $j_{S'}: S' \hookrightarrow M'$, the natural injections. We shall also assume that $j_S^*\Omega$ is a form of constant rank on S.

Definition 5.1: The map $\Phi: M \to M'$ is said to be a *canonical transformation* of (M, S, Ω) into (M', S', Ω') iff:

(i) Φ is a diffeomorphism;

(ii) $\Phi(S) = S'$;

(iii) the Jacobian map $\Phi_*: T(M) \to T(M')$ maps every locally weakly Hamiltonian field relative to (M, S, Ω) into a locally weakly Hamiltonian field relative to (M', S', Ω').

Thanks to condition (ii), the map Φ induces a diffeomorphism $\varphi: S \to S'$ by merely taking

$$\forall x \in S, \quad \varphi(x) = \Phi(x) \in S'.$$

(We denote this mapping by other symbol than Φ because its domain is not M but S.) Since φ has been purposely defined, we have that the diagram

$$\begin{array}{ccc} M & \xrightarrow{\Phi} & M' \\ {\scriptstyle j_S}\uparrow & & \uparrow{\scriptstyle j_{S'}} \\ S & \xrightarrow{\varphi} & S' \end{array} \tag{5.1}$$

is commutative.

The aim of this section is to give several characterizations of canonical transformations between constrained systems.

Lemma 5.2: If $\mathbf{H} \in \overline{D(S)}$, then $\Phi_* \mathbf{H} \in \overline{D(S')}$.

Proof: $\forall \xi' \in \Lambda^0(M')$, such that $j_{S'}^* \xi' = 0$, we have, taking the commutativity of diagram (5.1) into account, that

$$j_S^*(\Phi^* \xi') = (j_S^* \circ \Phi^*)(\xi') = \varphi^*(j_{S'}^* \xi') = 0. \tag{5.2}$$

If we now make $\Phi_* \mathbf{H}$ act on any function $\xi' \in \Lambda^0(M')$, such that $j_{S'}^* \xi' = 0$, we obtain

$$\Phi_* \mathbf{H}(\xi') = \mathbf{H}(\Phi^* \xi') = 0$$

because $\mathbf{H} \in \overline{D(S)}$ and $\Phi^* \xi'$ vanishes on S. Hence $\Phi_* \mathbf{H} \in \overline{D(S')}$. ⊠

Theorem 5.3: If conditions (i) and (ii) are satisfied by a given map $\Phi: M \to M'$, then Φ is a canonical transformation if, and only if,

$$\Phi^*(j_{S'}^* \Omega') = c\, j_S^* \Omega, \tag{5.3}$$

where $c \in \Lambda^0(S)$ is a locally constant function.

Proof: $\Phi^*((j_{S'}^* \Omega'))$ is a differential 2-form in the submanifold S which, at its turn, is endowed with the presymplectic structure given by $j_S^* \Omega$.

Let $\mathbf{H}$ be locally weakly Hamiltonian relatively to (M, S, Ω). According to Lemma 5.2, since $\mathbf{H} \in \overline{D(S)}$, then $\Phi_* \mathbf{H}$ will belong to $\overline{D(S')}$ as well. Furthermore, using Eq. (5.1), it can be seen in an obvious way that the vector field $(\Phi_* \mathbf{H})^{S'}$ associated with $\Phi_* \mathbf{H}$ by Eq. (4.2) is none but $\varphi_*(\mathbf{H}^S)$. Thus, for every $\mathbf{H} \in \overline{D(S)}$, we can write

$$\begin{aligned} L_{\mathbf{H}^S}(\varphi^*(j_{S'}^* \Omega')) &= \varphi^* \{L_{\varphi_*(\mathbf{H}^S)}(j_{S'}^* \Omega')\} \\ &= \varphi^* \{L_{(\Phi_* \mathbf{H})^{S'}}(j_{S'}^* \Omega')\}. \end{aligned} \tag{5.4}$$

(a) If Φ is a canonical transformation, then $\Phi_* \mathbf{H}$ is a locally weakly Hamiltonian field relatively to (M', S', Ω'). Therefore, we have, by Eq. (4.12), that the right-hand side of (5.4) vanishes. Whence the hypothesis of Theorem 4.10 is fulfilled and, consequently, Eq. (5.3) follows.

(b) Conversely, if Eq. (5.3) holds, we then have that

$$L_{\mathbf{H}^S}(\Phi^*(j_{S'}^* \Omega')) = c L_{\mathbf{H}^S}(j_S^* \Omega) = 0. \tag{5.5}$$

That is, the left-hand side of Eq. (5.4) vanishes. Therefore, since φ is a diffeomorphism, we arrive at

$$L_{(\Phi_* \mathbf{H})^{S'}}(j_{S'}^* \Omega') = 0,$$

which means that $\Phi_* \mathbf{H}$ is locally weakly Hamiltonian relatively to (M', S', Ω').

In most cases occurring in analytical dynamics, the symplectic forms Ω and Ω' are not only closed forms but exact as well and can be derived from the respective Liouville forms, $\theta \in \Lambda^1(M)$ and $\theta' \in \Lambda^1(M')$, i.e., $\Omega = d\theta$ and $\Omega' = d\theta'$. Then there follows immediately:

Corollary 5.4: A map $\Phi: M \to M'$, fulfilling conditions (i) and (ii) of Definition 5.1, is a canonical transformation if, and only if, for every $z \in S \subset M$, there exists an open neighborhood $U \subset M$, a function $F \in \Lambda^0(U)$, and a constant c such that

$$j_S^*(\Phi^* \theta' - c\theta - dF) = 0. \tag{5.6}$$

Lemma 5.5: Let $\mathbf{H} \in D(M)$ be a weakly Hamiltonian field and $h \in \Lambda^0(M)$ a Hamiltonian function for $\mathbf{H}$. If $\Phi: M \to M'$ is a canonical transformation, then $h' \equiv (1/c)\Phi^{*-1} h$ is a Hamiltonian function for $\mathbf{H}' \equiv \Phi_* \mathbf{H} \in D(M')$.

Proof: According to the hypothesis and Eq. (4.4), we have that

$$j_{S'}^{*}(i_{\mathbf{H}}\Omega - dh) = 0. \tag{5.7}$$

Since Φ is canonical, Theorem 5.3 can be applied to express (5.7) as

$$\frac{i}{c} i_{\varphi_*^{-1}(\mathbf{H}'^{S'})}\{\varphi^{*}(j_{S'}^{*}\Omega') - (\varphi^{*}\circ j_{S'}^{*})d(\Phi^{*-1}h)\} = 0$$

or

$$(\varphi^{*}\circ j_{S'}^{*})[i_{(1/c)\mathbf{H}'}\Omega' - d(\Phi^{*-1}h)] = 0.$$

Finally, as φ^{*} is a diffeomorphism, we arrive at

$$j_{S'}^{*}[i_{(1/c)\mathbf{H}'}\Omega' - d(\phi^{*-1}h)] = 0, \tag{5.8}$$

which proves the lemma.

Using the Poincaré lemma, a similar result can be proved to hold for locally weakly Hamiltonian fields. Taking Eq. (4.11) into account, we can then use Lemma 5.5 to write $\Phi_{*}\mathbf{H}$ as

$$\Phi_{*}\mathbf{H} \underset{S}{=} \{\cdot, c\tilde{h}\}'^{*} + \{\cdot, \tilde{a}_{\alpha}\tilde{\xi}^{\alpha}\}', \tag{5.9}$$

here $\{\ ,\ \}'$ and $\{\ ,\ \}'^{*}$ are, respectively, the Poisson and Dirac brackets corresponding to (M', S', Ω'), and

$$\tilde{h} \equiv \Phi^{*-1}h, \quad \tilde{a}_{\alpha} = \Phi^{*-1}a_{\alpha}, \quad \tilde{\xi}^{\alpha} = \Phi^{*-1}\xi^{\alpha}.$$

If we now apply $\mathbf{H}$ to any Hamiltonian function g, we obtain, taking (5.9) and (4.7) into account, that

$$(\mathbf{H}g) \underset{S}{=} \{g, h\}^{*} \tag{5.10}$$

and, for $\Phi_{*}\mathbf{H}$ acting on $\tilde{g}$, we also have

$$(\Phi_{*}\mathbf{H})\tilde{g} \underset{S}{=} c\{\tilde{g}, \tilde{h}\}'^{*}, \tag{5.11}$$

where $\tilde{g} \equiv \Phi^{*-1}g$.

Comparison of both equations, (5.10) and (5.11), suggest to us the following:

Theorem 5.6: A map $\Phi: M \to M'$ fulfilling conditions (i) and (ii) definition 5.1 is a canonical transformation if, and only if,

(a) For every couple of functions g and h, which are Hamiltonian relative to (M, S, Ω),

$$\{\tilde{g}, \tilde{h}\}'^{*}\circ\varphi \approx (1/c)\{g, h\}^{*}. \tag{5.12}$$

(b) If $g \in \Lambda^{0}(M)$ is Hamiltonian relative to (M, S, Ω), then $\tilde{g} \equiv \Phi^{*-1}g$ is Hamiltonian relative to (M', S', Ω') as well.

Proof: Provided that Φ is a canonical transformation and taking Eqs. (5.10), (5.11) and the commutativity of (5.1) into account, it follows that

$$\begin{aligned}\{g, h\}^{*} &= j_{S}^{*}(\mathbf{H}g) = (\varphi^{*}\circ j_{S'}^{*}\circ\Phi^{*-1})(\mathbf{H}g)\\ &= \varphi^{*}\circ j_{S'}^{*}[(\Phi_{*}\mathbf{H})(\tilde{g})]\\ &= c\{\tilde{g}, \tilde{h}\}'^{*},\end{aligned}$$

and (a) is proved.

Statement (b) follows immediately applying Lemma 5.5.

Let us now conversely assume (a) and (b) and try to prove Φ is a canonical transformation. According to Eq. (5.12), for every weakly Hamiltonian vector field $\mathbf{H} \in D(M)$ with the associated Hamiltonian function $h \in \Lambda^{0}(M)$, and every Hamiltonian function $g \in \Lambda^{0}(M)$, we have that

$$[(\Phi_{*}\mathbf{H} - c\mathbf{X}_{\tilde{h}})\tilde{g}]\circ j_{S'}^{*} = 0, \tag{5.13}$$

where $\mathbf{X}_{\tilde{h}} \in \overline{D(S')}$ is a solution of

$$j_{S'}^{*}(i_{\mathbf{X}}\Omega' - d\tilde{h}) = 0. \tag{4.4'}$$

The assumption (b) guarantees the existence of such a solution $\mathbf{X}_{\tilde{h}}$. It is also this assumption which ensures that $\tilde{g} \in \Lambda^{0}(M)$ will be Hamiltonian relative to (M', S', Ω'). From Eq. (5.13) we can therefore state that

$$[(\Phi_{*}\mathbf{H} - c\mathbf{X}_{\tilde{h}})f]\circ j_{S'} = 0. \tag{5.14}$$

Finally, taking Proposition 2.6 into account, Eq. (5.14) implies immediately that

$$(\Phi_{*}\mathbf{H} - c\mathbf{X}_{\tilde{h}})\circ j_{S'} \in \mathscr{A}(j_{S'}^{*}\Omega')$$

or

$$\Phi_{*}\mathbf{H} \underset{S'}{=} c\mathbf{X}_{\tilde{h}} + \mathscr{A}(j_{S}^{*}\Omega')$$

which, by inner product with $j_{S'}^{*}\Omega'$, yields

$$j_{S'}^{*}(i_{\Phi_{*}\mathbf{H}}\Omega') = c\, j_{S'}^{*}(i_{\mathbf{X}_{\tilde{h}}}\Omega') = c\, j_{S'}^{*}\, d\tilde{h} = j_{S'}^{*}\, d(c\tilde{h}),$$

whence $\Phi_{*}\mathbf{H} \in D(M')$ is a weakly Hamiltonian field relatively to (M', S', Ω') and $c\tilde{h}$ is an associated Hamiltonian function. ⊠

Theorem 5.6 will be a useful tool to characterize canonical transformations of constrained Hamiltonian systems in terms of the Dirac brackets. Indeed, let us consider a Shanmugadhasan set of coordinates and momenta $q_a, p_B; Q_f^1, P_g^1; Q_k^2, P_h^2$ $(a, b = 1,...,r; f, g = 1,...,l; k, h = 1,...,t)$ defined on some open domain $U \subset M$. In terms of this coordinates the submanifold S is defined by

$$\begin{aligned}&\text{(first class)} \quad P_g^1 \approx 0, \qquad g = 1,...,l,\\ &\text{(second class)} \quad Q_k^2 \approx P_h^2 \approx 0, \quad k, h = 1,...,t\end{aligned}$$

(r and t are related to the dimension $2n$ and s by $2r = s - l$ and $2t = 2n - s - l$), the differential 2-forms Ω and $j_S^{*}\Omega$ can be written as

$$\Omega = \sum_{a=1}^{r} dq_a \wedge dp_a + \sum_{f=1}^{l} dQ_f^1 \wedge dP_f^1 + \sum_{k=1}^{t} dQ_k^2 \wedge dP_k^2$$

and

$$j_S^{*}\Omega = \sum_{a=1}^{r} dq_a \wedge dp_a$$

and, according to (4.7), Hamiltonian functions are characterized by

$$\left(\frac{\partial F}{\partial Q_g^1}\right)_{(q, p; Q^1, 0, 0, 0)} = 0, \quad g = 1,...,l,$$

that is, the elementary Hamiltonian functions are q_a, p_B, P_f^1 $(a, b = 1,...,r; f = 1,...,l)$. (Realize that, although the second-class constraints Q_h^2, P_k^2 also satisfy the latter relations, they are not relevant since their Dirac brackets with any other function vanish.)

In the most cases of interest in mathematical physics, a canonical transformation Φ acts from (M, S, Ω) into itself. In terms of the above set of Shanmugadhasan variables, Φ is expressed by $2n$ functions $\tilde{z}_{\alpha}(z_{\beta})$ $(\alpha, \beta = 1,...,2n)$, where z_{β} denotes generically the coordinates and momenta q, p; Q^1, P^1; Q^2, P^2.

Theorem 5.6 then states that Φ is a canonical transformation of (M, S, Ω) into itself if, and only if,

(i) $\tilde{Q}^2_k(q,p;Q^1,0;0,0) = 0,$

$$\tilde{P}^2_k(q,p;Q^1,0;0,0) = 0, \quad k,h = 1,...,t, \tag{5.15}$$

$$\tilde{P}^1_g(q,p;Q^1,0;0,0) = 0, \quad g = 1,...,l$$

that is, $\Phi(S) = S$.

(ii) Φ transforms Hamiltonian functions into Hamiltonian functions, that is,

$$\{\tilde{q}_a, P^1_g\} \approx 0, \quad \{\tilde{p}_b, P^1_g\} \approx 0, \quad \{\tilde{P}^1_f, P^1_g\} \approx 0. \tag{5.16}$$

(iii) There exists a constant $c \in R$ such that

$$\{\tilde{q}_a, \tilde{p}_b\}^* \approx c\delta_{ab} \tag{5.17}$$

and

$$\{\tilde{q}_a, \tilde{q}_b\}^* \approx \{\tilde{p}_a, \tilde{p}_b\}^* \approx \tilde{P}^1_f, \tilde{P}^1_g\}^* \approx \{\tilde{q}_a, \tilde{P}^1_f\}^* \approx \{\tilde{p}_b, \tilde{P}^1_f\}^* \approx 0. \tag{5.18}$$

6. CONCLUSION AND OUTLOOK

In this paper we have studied in depth constrained Hamiltonian systems. In particular, we have introduced the concepts of Hamiltonian function and Hamiltonian vector field relative to a given constrained Hamiltonian system.

The Hamiltonian study of a dynamical system leads to the concept of canonical transformation in a natural way (i.e., that which preserves the canonical formalism). In the nondegenerate formalism, the invariant integral of Poincaré[8] and the theorem of Lee Hwa Chung[12] permit us to characterize canonical transformations as those which preserve the symplectic structure apart from a multiplicative constant ($\Phi^* \Omega' = c\Omega$, $c \in \mathbb{R}$) or, equivalently, as those preserving the Poisson brackets (except for a constant factor).

In a similar way, we have here derived a generalization of Lee Hwa Chung's theorem for constrained Hamiltonian systems which, insofar as the generalization of the Poincaré integral invariant has been already obtained, has allowed us to characterize canonical tranformations of a given constrained system into itself as those preserving: (i) the submanifold S and (ii) the induced presymplectic structure on S [i.e., $\Phi^*(j_S^* \Omega) = c\, j_S^* \Omega$; $c \in \mathbb{R}$].

We have also derived a formulation for the latter condition in terms of the elementary Dirac brackets.

Previous attempts to generalize the concept of canonical transformations for a constrained system (M, S, Ω) were based on the condition of preserving the Poisson bracket on the constraint submanifold S. In the present work we have intended to clarify that this condition is too restrictive. This is owing to the fact that the physical system is completely represented in the presymplectic manifold $(S, j_S^* \Omega)$ and the degrees of freedom beyond S are physically irrelevant. On the contrary, in order to avoid unphysical restrictions, our study of canonical tranformations starts from the most primary characteristic, that is, to transform weakly Hamiltonian vector fields into weakly Hamiltonian vector fields. Then, the results obtained in Secs. 2, 3, and 4 permit us to give more elaborated characterizations of canonical transformations for constrained systems, either in terms of the presymplectic form $j_S^* \Omega$ or in terms of the Dirac brackets and weakly Hamiltonian functions.

It is finally interesting to remark that a good definition for canonical transformations will be of great help to study the symmetries of a given constrained Hamiltonian system, and a forthcoming paper will be devoted to this problem.

ACKNOWLEDGMENTS

This work has been partially supported by the Ayuda Comisión Asesora de Investigación y Técnica No. 0435.

APPENDIX

Throughout this appendix $\mathscr{E}$ is a finite dimensional linear space ($\dim \mathscr{E} = s$) and $\omega \in \Lambda^2(\mathscr{E})$ is an alternated 2-form.

The 2-form ω defines the linear mapping

$$\mathscr{E} \to \mathscr{E}^* = \Lambda^1(\mathscr{E}), \qquad \mathbf{V} \rightsquigarrow i_{\mathbf{V}} \omega, \tag{A1}$$

the kernel of which is $\mathscr{A}(\omega) = \{\mathbf{V} \in \mathscr{E} / i_{\mathbf{V}} \omega = 0\}$. Also, it can be easily proved that its image is

$$i_{\omega}\, \omega = \mathscr{A}^{\perp}(\omega) \equiv \{\lambda \in \mathscr{E}^* / i_{\mathbf{V}} \lambda = 0, \ \forall \mathbf{V} \in \mathscr{A}(\omega)\} \tag{A2}$$

The rank of ω is defined by

$$\operatorname{rank} \omega = \dim \mathscr{A}^{\perp}(\omega), \tag{A3}$$

which is always an even number

$$\operatorname{rank} \omega = 2R \leqslant s. \tag{A4}$$

Let $\Lambda^p(\mathscr{E}/\mathscr{A}(\omega))$ be the space of alternated p-forms on the quotient space $\mathscr{E}/\mathscr{A}(\omega)$, and let us consider the mapping:

$$\Lambda^p(\mathscr{E}/\mathscr{A}(\omega)) \to \Lambda^p(\mathscr{E}), \qquad \alpha \rightsquigarrow \tilde{\alpha}$$

such that

$$\forall \mathbf{V}_1,...,\mathbf{V}_p \in \mathscr{E}, \quad \tilde{\alpha}(\mathbf{V}_1,..,\mathbf{V}_p) \equiv \alpha([\mathbf{V}_1],..,[\mathbf{V}_p]),$$

where $[\mathbf{V}_i] \in \mathscr{E}/\mathscr{A}(\omega)$ is the class of $\mathbf{V}_i \in \mathscr{E}$.

Proposition A.1: The mapping $\alpha \to \tilde{\alpha}$ defines an isomorphism between $\Lambda^p(\mathscr{E}/\mathscr{A}(\omega))$ and

$$\Lambda^{p\perp}(\omega) \equiv \{\gamma \in \Lambda^p(\xi) / i_{\mathbf{V}} \gamma = 0, \ \forall \mathbf{V} \in \mathscr{A}(\omega)\} \subset \Lambda^p(\mathscr{E}). \tag{A5}$$

The latter result is an immediate consequence of proposition (I.1.12) of Ref. 19.

Since

$$\dim(\mathscr{E}/\mathscr{A}(\omega)) = \dim \mathscr{A}^{\perp}(\omega) = 2R, \tag{A6}$$

we have that $p > 2R \Rightarrow \Lambda^p(\mathscr{E}/\mathscr{A}(\omega)) = \{0\}$, which, by Proposition A.1, implies the following:

Corollary A.2: If $\alpha \in \Lambda^p(\mathscr{E})$, $p > 2R$ and $i_{\mathbf{V}} \alpha = 0$, for any $\mathbf{V} \in \mathscr{A}(\omega)$, then $\alpha = 0$.

Now, let $\mathscr{F}$ be a linear subspace of $\mathscr{E}$ ($\dim \mathscr{F} = m$).

Proposition A.3: If $\gamma \in \Lambda^p(\mathscr{F})$, $p < s - m$, and $\lambda \wedge \gamma = 0$, for any $\lambda \in \mathscr{F}^{\perp}$, then $\gamma = 0$.

Proof: Let us choose a basis $\epsilon^1,...,\epsilon^{s-m}$ of $\mathscr{F}^{\perp}$ (i.e., the subspace of 1-forms annihilating $\mathscr{F}$). There will exist $s - m$ vectors

$$\mathbf{X}_1,...,\mathbf{X}_{s-m} \in \mathscr{E}$$

such that

$$i_{\mathbf{X}_j}\epsilon^l = \delta^l_j, \quad l,j = 1,...,s-m.$$

By the hypothesis $\epsilon^j \wedge \gamma = 0, j = 1,...,s-m$, and taking into account that

$$i_{\mathbf{X}_j}(\epsilon^j \wedge \gamma) = \gamma - \epsilon^j \wedge i_{\mathbf{X}_j}\gamma, \quad j = 1,...,s-m,$$

we conclude

$$\gamma = \epsilon^j \wedge i_{\mathbf{X}_j}\gamma, \quad j = 1,...,s-m,$$

whence it immediately follows that

$$\gamma = \epsilon^p \wedge \epsilon^j \wedge (i_{\mathbf{X}_j} i_{\mathbf{X}_p}\gamma), \quad j,l = 1,...,s-m.$$

A recursive application of this method leads to

$$\gamma = 0 \quad \text{if} \quad p < s-m$$

or

$$\gamma = \epsilon_1 \wedge \cdots \epsilon^{s-m} \wedge (i_{\mathbf{X}_{s-m}} \cdots i_{\mathbf{X}_1}\gamma) \quad \text{if} \quad p \geqslant s-m.$$

Particularizing the latter proposition for the cases $\mathscr{F} = 0$ and $\mathscr{F} = \mathscr{A}(\omega)$ we have, respectively:

Corollary A.4: If $\gamma \in \Lambda^p(\mathscr{E})$, $p < s$, and $\gamma \wedge \gamma = 0$, for any $\lambda \in \mathscr{E}^*$, then $\gamma = 0$.

Corollary A.5: If $\gamma \in \Lambda^p(\mathscr{E})$, $p < 2R = \text{rank}\,\omega$, and $\lambda \wedge \gamma = 0$ for any $\lambda \in \mathscr{A}^\perp(\omega)$, then $\gamma = 0$.

Finally, we are going to prove the main result of this appendix which is needed in the proof of Theorem 3.1.

Proposition A.6: If $\mathscr{A} \in \Lambda^p(\mathscr{E})$, $p \leqslant 2R$, satisfies

$$\text{(i)}\ \forall \mathbf{V} \in \mathscr{A}(\omega), \quad i_{\mathbf{V}}\alpha = 0, \tag{A7}$$

$$\text{(ii)}\ \forall \mathbf{X}, \mathbf{Y} \in \mathscr{E}, \quad i_{\mathbf{X}}\omega \wedge i_{\mathbf{Y}}\alpha + i_{\mathbf{Y}}\omega \wedge i_{\mathbf{X}}\alpha = 0, \tag{A8}$$

then, either $p = 1$, which implies $\alpha = 0$, or $p > 1$, which implies that there exists $\alpha^{(1)} \in \Lambda^{p-2}(\mathscr{E})$, fulfilling (A7) and (A8), such that

$$\forall \mathbf{X} \in \mathscr{E}, \quad i_{\mathbf{X}}\alpha = i_{\mathbf{X}}\omega \wedge \alpha^{(1)}.$$

Proof: In accordance with Eq. (A2), we have that, given any $\lambda \in \mathscr{A}^\perp(\omega)$, there exists $\mathbf{V} \in \mathscr{E}$ such that $\lambda = i_{\mathbf{V}}\omega$.

Condition (A6) and Corollary A.4 ensure that, given any two $\mathbf{V}_1, \mathbf{V}_2, \in \mathscr{E}$ such that $i_{\mathbf{V}_1}\omega = i_{\mathbf{V}_2}\omega = \lambda$; then

$$i_{\mathbf{V}_1}\alpha = i_{\mathbf{V}_2}\alpha. \tag{A9}$$

Therefore, there is no ambiguity in defining the mapping:

$$\Phi: \mathscr{A}^\perp(\omega) \to \Lambda^{p-1}(\mathscr{E}),$$
$$\lambda \rightsquigarrow \Phi(\lambda) = i_{\mathbf{V}}\alpha, \tag{A10}$$

$\mathbf{V}$ being any vector such that $i_{\mathbf{V}}\omega = \lambda$. Then, taking $\mathbf{X} = \mathbf{Y}$ in Eq. (A8), we obtain

$$\forall \mathbf{X} \in \mathscr{E}, \quad i_{\mathbf{X}}\omega \wedge i_{\mathbf{X}}\alpha = 0,$$

which, due to Eq. (A2), is equivalent to

$$\lambda \wedge \Phi(\lambda) = 0, \quad \forall \lambda \in \mathscr{A}^\perp(\omega).$$

This implies that either (i) $\Phi(\lambda) = 0$ if $p = 1$ or (ii) there exists $\alpha^{(1)} \in \Lambda^{p-2}(\mathscr{E})$, such that

$$\Phi(\lambda) = \lambda \wedge \alpha^{(1)} \quad \text{if} \quad p > 1.$$

Recalling now (A10), we see that this is equivalent to either (i) $\alpha = 0$, if $p = 1$, or (ii) there exists $\alpha^{(1)} \in \Lambda^{p-2}(\mathscr{E})$ such that

$$i_{\mathbf{X}}\alpha = i_{\mathbf{X}}\omega \wedge \alpha^{(1)}, \quad \forall \mathbf{X} \in \mathscr{E} \quad \text{if} \quad p > 1. \tag{A11}$$

To complete the proof, we must still check whether $\alpha^{(1)}$ fulfills conditions (A7) and (A8).

By inner product of both sides of Eq. (A11) with any $\mathbf{V} \in \mathscr{E}$, and, since $i_{\mathbf{X}}$ and $i_{\mathbf{V}}$ anticommute, we obtain that

$$i_{\mathbf{V}}\omega \wedge i_{\mathbf{X}}\alpha^{(1)} + i_{\mathbf{X}}\omega \wedge i_{\mathbf{V}}\alpha^{(1)} = 0, \quad \forall \mathbf{V}, \mathbf{X} \in \mathscr{E}. \tag{A12}$$

That is, condition (A8) holds good for $\alpha^{(1)}$ also.

Taking $\mathbf{V} \in \mathscr{A}(\omega)$ in Eq. (A12), it yields

$$i_{\mathbf{X}}\omega \wedge i_{\mathbf{V}}\alpha^{(1)} = 0, \quad \forall \mathbf{X} \in \mathscr{E}, \quad \forall \mathbf{V} \in \mathscr{A}(\omega),$$

which, by virtue of Eq. (A2), is equivalent to

$$\lambda \wedge i_{\mathbf{V}}\alpha^{(1)} = 0, \quad \forall \lambda \in \mathscr{A}^\perp(\omega), \quad \forall \mathbf{V} \in \mathscr{A}(\omega);$$

applying corollary A.5, it implies that

$$i_{\mathbf{V}}\alpha^{(1)} = 0, \quad \forall \mathbf{V} \in \mathscr{A}(\omega),$$

and the proof is completed.

[1] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[2] P. A. M. Dirac, Lectures in quantum mechanics, Belfer Graduate School of Science, Yeshiva University, New York, 1964.
[3] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[4] A. J. Hanson, T. Regge, and C. Teitelboim, *Constraint Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976).
[5] M. J. Gotay, "Presymplectic manifolds, Geometric constraint theory...," Preprint 80-063, University of Maryland, 1979.
[6] See, for example, *The Proceedings of Action at a Distance Theory, Classical and Quantum Aspects, Barcelona 1981*, Lecture Notes in Physics, Vol. 162 (Springer-Verlag, Berlin, 1982).
[7] S. Shanmugadhasan, J. Math. Phys. **14**, 677 (1973).
[8] E. Cartan, *Leçons sur les Invariants Integraux* (Hermann, Paris, 1922).
[9] F. Benavent and J. Gomis, Ann. Phys. (N.Y.) **118**, 476 (1979).
[10] D. Dominici and J. Gomis, J. Math. Phys. **21**, 2124 (1980); **23**, 256 (1982).
[11] F. Gantmacher, *Lectures in Analytical Mechanics* (Mir, Moscow, 1970).
[12] L. H. Chung, Proc. R. Soc. London, 237 (1945).
[13] S. Kaptanoglu, Phys. Lett. B **98**, 77 (1981).
[14] J. Sniatycki, Ann. Inst. H. Poincaré, **20**, 305 (1974).
[15] A. Lichnerowicz, C. R. Acad. Sci. Paris A **280**, 523 (1975).
[16] M. R. Menzio and W. Tulczyjew, Ann. Inst. H. Poincaré, A **28**, 23 (1978).
[17] M. Gotay, J. Nester, and G. Hinds, J. Math. Phys. **19**, 2388 (1978).
[18] G. Marmo, N. Mukunda, and J. Samuel, Riv. Nuovo Cimento **6**, 1 (1983).
[19] C. Godbillon, *Geometric Differentielle et Mécanique Analitique* (Hermann, Paris, 1969).
[20] M. Spivak, *Cálculo en variedades* (Reverté, Barcelona, 1972).
[21] P. G. Bergmann and A. Komar, Phys. Rev. Lett. **4**, 432 (1960).

On the derivation of hydrodynamics from molecular dynamics

Michael G. Mürmann
Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 6900 Heidelberg, Federal Republic of Germany

(Received 31 May 1983; accepted for publication 29 July 1983)

We study the transition from molecular dynamics to hydrodynamics. Convergence of subsequences and absolute continuity of the limit are proven for the local distributions of mass and momentum. The corresponding limit dynamics is discussed with regard to the foundations of hydrodynamics. A Chapman–Enskog-like procedure leads formally to local equilibrium with the parameter functions obeying the system of the Euler equations.

PACS numbers: 03.40.Gc, 05.20.Dd

INTRODUCTION

The rigorous derivation of the foundations of hydrodynamics from molecular dynamics is the most challenging, but also most difficult problem in nonequilibrium statistical mechanics. In fact, there is general agreement that at present there is no hope for the solution of this problem. In order to get a deeper insight into the underlying dynamical structures and to develop methods for a solution, one can proceed in two different ways. One can study simplified models, which are exactly solvable, but in spite of their idealized character show the desired physical features. A typical example is the one-dimensional systems of hard rods[1] (for further examples see, e.g., Refs. 2–5). An alternative possibility, which we shall stress, is the study of the original model as to partial rigorous results and formal expansions which explain the transition to hydrodynamics and might be made rigorous one day.

We premise a section on the general form of the approximation of molecular dynamics by continuum dynamics in order to fit our procedure into the general frame. Our approach follows Lanford's work on the Boltzmann equation.[6] According to the respective situation we shall distinguish three degrees of approximation.

In the second section we prove the convergence of subsequences and the absolute continuity of the limit of the local distributions of mass and momentum for quite general deterministic initial configurations. We discuss the corresponding limit dynamics in the third section and illustrate in particular how the hydrodynamical pressure arises. Finally we treat the case of random initial configurations. We formulate the evolution equations of the correlation functions as a singular perturbation problem. A formal approximation, whose derivation is reminiscent of the Chapman–Enskog method, leads to the correlation functions of local equilibrium. Their parameters, inserted into the conservation equations, satisfy the system of the Euler equations.

The present work originated in the further development of the preliminary results of Ref. 7. Our procedure is roughly comparable to an old paper of Morrey,[8] as we realized during its accomplishment. He treats for the most part the random case. Though his paper is technically too involved and there are some gaps in it, it is a remarkable paper for that time and already indicates the way one should attack the problem.

1. APPROXIMATION OF MOLECULAR DYNAMICS BY HYDRODYNAMICS

A hydrodynamical behavior of molecular systems is expected to hold approximately if the microscopic quantities are small compared to their corresponding macroscopic ones. We choose as the basic parameter to express their relation the value

$$\alpha = \frac{\text{microscopic unit length}}{\text{macroscopic unit length}},$$

and formulate such an approximation mathematically by an idealized limit theorem with $\alpha \to 0$ or by an expansion in α. According to the actual physical situation, further quantities like time, mass, etc., have to be suitably scaled in dependence on α.

In this section we shall demonstrate by means of typical examples in what sense these approximations hold. First we have to make precise how microscopic states are approximated by macroscopic ones, before we treat the approximation of the dynamics. In this paper we are dealing with classical systems in the d-dimensional Euclidean space $\mathbb{R}^d (d \geqslant 1)$. A microscopic state is given by an element

$$\mathbf{x}_N = (x_1, \ldots, x_N) \in \mathbb{R}^{2Nd}$$

with

$$x_i = (q_i, v_i) \in \mathbb{R}^d \times \mathbb{R}^d \quad (i = 1, 2, \ldots, N),$$

where N is the number of particles (molecules) and q_i, resp., v_i denotes the position, respectively, velocity of the ith particle.

We describe a macroscopic state by a density f, defined on $\mathbb{R}^{2d}$, with the meaning

$$\int_A \int_B f(q,v)\,dq\,dv = \text{mass of fluid elements located in } A \text{ with velocities in } B.$$

In order to approximate microscopic states by macroscopic ones we conceive both as measures. We replace $\mathbf{x}_N$ by $\Sigma_i \delta_{x_i}$. This neglects the order of the indices and thus refers to indistinguishable particles. It has the further advantage of treating systems with variable particle numbers N as objects in one space. The density f will be identified with the measure having this density with respect to the Lebesgue measure.

 0022-2488/84/051356-08$02.50

The approximation of microscopic states by macroscopic ones is realized with respect to the weak topology of measures. For this purpose the microscopic state $\Sigma_i \delta_{x_i}$ has to be properly normalized to $m(\alpha)\Sigma_i \delta_{x_i}$, where $m(\alpha)$ can be interpreted as the mass of a single particle in a macroscopic scale.

Thus, roughly speaking, a state $m(\alpha)\Sigma_i \delta_{x_i}$ is approximated by f, if for microscopically large, but macroscopically small sets $A, B \subset \mathbb{R}^d$,

$$m(\alpha)|\{i: q_i \in A,\ v_i \in B\}| \approx \int_A \int_B f(q,v)\,dq\,dv.$$

Rigorous results involve the limit $\alpha \to 0$ and the dependence of the x_i on α. We suppress the quotation of α to make the formulas more lucid, whereas the normalization $m(\alpha)$ will be explicitly specified.

Now supposed that there is a microscopic time evolution, which is given on the macroscopic scale by

$$T_t^\alpha : \mathbf{x}_N(0) \mapsto \mathbf{x}_N(t) \quad (t \geqslant 0,\ N \geqslant 0).$$

A macroscopic time evolution

$$T_t : f_0 \to f_t \quad (t \geqslant 0),$$

defined for a suitable class of smooth densities is an approximation to the microscopic dynamics, if the following holds.

If for each $\alpha > 0$ there is given an initial configuration $\mathbf{x}_N(0)$ such that $m(\alpha)\Sigma_i \delta_{x_i(0)} \to f_0$ weakly as $\alpha \downarrow 0$ then for each $t \geqslant 0$, or at least in some interval $[0,T]$, $m(\alpha)\Sigma_i \delta_{x_i(t)} \to f_t$ weakly as $\alpha \downarrow 0$.

Such a limiting behavior means a drastic reduction of the degrees of freedom, since the approximated density only depends on the occupation numbers of microscopically large sets. It can be interpreted as deterministic macroscopic behavior of matter.

There is one example where such a result is true, namely the Vlasov or weak-coupling limit.[9,10] In this limit $m(\alpha) = \alpha^d$ (dense gas) with the Newtonian dynamics with respect to a weak, long-range potential realized by the scaling $\phi_\alpha(q) = \alpha^d \phi(q)$.

The limit dynamics is given by the Vlasov equation. This form of a purely deterministic behavior is an exception. In most cases, where a limit dynamics exists, there are initial configurations, which are not approximated by it. But they are in a minority. The overwhelming majority behaves as in the deterministic case. At this stage probabilities have to be introduced to attach small weight to the exceptional configurations. From a probabilistic point of view the corresponding limit theorems are weak laws of large numbers. Remark that the notion of probability has a different meaning than in the ensemble conception. For a detailed discussion see Refs. 6 and 11.

The most important example is the Boltzmann–Grad limit[6] with $m(\alpha) = \alpha^{d-1}$ (rarefied gas) and elastic collisions of hard spheres of radius α. This limit dynamics is governed by the Boltzmann equation. In the model, which we shall treat in this paper, a further reduction of the kind of approximation is necessary. We briefly introduce the model in order to formulate this type of behavior.

We deal with the case of dense gases: $m(\alpha) = \alpha^d$, with the dynamics given by a smooth, even potential ϕ, which we rescale to $\phi_\alpha(q) = \phi(q/\alpha)$. This keeps the total cross section fixed. The corresponding force is

$$F_\alpha(q) = -\nabla \phi_\alpha(q) = \alpha^{-1} F(q/\alpha) \quad \text{with } F = -\nabla \phi.$$

The equations of motion are

$$\dot{q}_i = v_i \quad \dot{v}_i = \sum_{j \neq i} F_\alpha(q_i - q_j). \tag{1.1}$$

Since the approximation of states is realized with respect to the weak topology, we express the time evolution of $\alpha^d \Sigma_i \delta_{x_i(t)}$ by its application to smooth test functions φ

$$\begin{aligned}\frac{d}{dt}\alpha^d \sum_i \varphi(q_i(t), v_i(t)) &= \alpha^d \sum_i \nabla_q \varphi(q_i(t), v_i(t)) v_i(t) \\ &\quad + \alpha^{d-1} \sum_i \nabla_v \varphi(q_i(t), v_i(t)) \sum_{j \neq i} F\left(\frac{q_i(t) - q_j(t)}{\alpha}\right).\end{aligned} \tag{1.2}$$

The last term behaves singularly as $\alpha \downarrow 0$ because of the factor α^{-1}, since $\Sigma_{j \neq i} F((q_i(t) - q_j(t))/\alpha)$ is of order 1. This is the crucial difference to the Vlasov and Boltzmann–Grad limit, where the coresponding term is of finite order.[12] No limit theorems of the former kinds can be expected to hold. But realize that the singularity cancels for certain functions φ, e.g., those which do not depend on the velocity. This corresponds to the application of the reduced measure $\alpha^d \Sigma \delta_{q_i(t)}$. More generally, we shall see in the next section, that the singularity, at least formally, cancels for the local distributions of the conserved quantities, i.e., for the partly signed and vector-valued measures

$$\rho_t^\alpha = \alpha^d \sum_i \delta_{q_i(t)}, \tag{1.3a}$$

$$v_t^\alpha = \alpha^d \sum_i v_i(t) \delta_{q_i(t)}, \tag{1.3b}$$

$$e_t^\alpha = \alpha^d \sum_i \left\{ \frac{1}{2} v_i(t)^2 + \frac{1}{2} \sum_{j \neq i} \phi\left(\frac{q_i(t) - q_j(t)}{\alpha}\right) \right\} \delta_{q_i(t)}. \tag{1.3c}$$

Thus the third kind of approximation only concerns the local distributions of the conserved quantities.

Rigorously we prove under suitable, quite general conditions on the potential and the initial configurations the following results.

(1) For each sequence $\alpha_n \downarrow 0$ there exist a subsequence $\alpha_{n(k)} \downarrow 0$ such that $\rho_t^{\alpha_{n(k)}}$ and $v_t^{\alpha_{n(k)}}$ converge uniformly on each compact time interval.

(2) Each limit of ρ_t^α and v_t^α as $\alpha \downarrow 0$ has a density with respect to the Lebesgue measure.

As we shall explain at the end of the next section, we conjecture that under the mentioned general conditions, only these results are possible. Stronger results require special initial configurations, e.g., local equilibrium. (See Sec. 4).

The three different kinds of results are presumably related to the behavior of a test particle. In the purely deterministic case its dynamics can be approximated by a deter-

ministic one.[9,10] In the stochastic case, its time evolution converges to a nonlinear Markov process.[13] The behavior of the whole system can be conceived in both cases as resulting from the composition of all particles by some law of large numbers, in the first case a deterministic one. Finally, in our model the singularity of the motion of the single particles makes a similar convergence result rather unlikely. A limit behavior can only be expected to hold for special collective quantities, which locally vary slowly, namely the conserved quantities.

2. DETERMINISTIC CONFIGURATIONS: RIGOROUS RESULTS

2.1. General results

We now set up the model in detail. Let

$$\phi:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$$

be a two-body potential, which is

(i) real valued and C^1 on $\mathbb{R}^d\setminus\{0\}$,
(ii) even, $\phi(q)=\phi(-q)$,
(iii) superstable.

We use the definition of superstability in the following form[14]:

Let $\mathbb{R}^d$ be decomposed into cubes: $\mathbb{R}^d=\bigcup_{k\in\mathbb{Z}^d} W_k$ with $W_{(k^1,\dots,k^d)}=\{q=(q^1,\dots,q^d):k^l\leqslant q^l<k^l+1,\ 1\leqslant l\leqslant d\}$. Then there exists $A>0$, $B\geqslant 0$ such that for $q_1,\dots,q_N\in\mathbb{R}^d$, $N\geqslant 2$:

$$\sum_{1\leqslant i<j\leqslant N}\phi(q_i-q_j)\geqslant\sum_{k\in\mathbb{Z}^d}[A\,N(\mathbf{q}_N,W_k)^2-BN(\mathbf{q}_N,W_k)]$$

with $\mathbf{q}_N=(q_1,\dots,q_N)$, $N(\mathbf{q}_N,C)=|\{i:q_i\in C\}|$ for $C\subset\mathbb{R}^d$. (2.1)

Concerning the initial configurations we make the following assumptions. For each $\alpha>0$ there is a configuration $\mathbf{x}^\alpha_{N(\alpha)}=\mathbf{x}^\alpha_{N(\alpha)}(0)$ such that the following expressions are uniformly bounded in $\alpha>0$. As before we omit the superscript α.

$$\text{(total mass)}\qquad \alpha^d N(\alpha)=\alpha^d\sum_i 1\leqslant M,$$

$$\text{(total energy)}\qquad \alpha^d\sum_i\left\{\frac{1}{2}v_i^2+\frac{1}{2}\sum_{j\neq i}\phi\left(\frac{q_i-q_j}{\alpha}\right)\right\}\leqslant E,$$

$$\text{(spread)}\qquad \alpha^d\sum_i q_i^2\leqslant D.$$

In the following these general conditions are always assumed without further mention. The time evolution is given by (1.1), resp. (1.2). The following boundedness properties are easy consequences of the general assumptions.

Lemma 2.1: (i) The potential energy and the kinetic energy

$$\frac{1}{2}\alpha^d{\sum_{i,j}}'\phi\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)\qquad\left[{\sum_{i,j}}'=\sum_{\substack{i,j\\ i\neq j}}\right],$$

$$\alpha^d\sum_i\frac{1}{2}v_i(t)^2$$

are separately bounded, uniformly in $\alpha>0$, $t\geqslant 0$.

(ii) For each $T>0$ there exists $D(T)>0$ such that $\alpha^d\sum_i q_i(t)^2\leqslant D(T)$ for $\alpha>0$, $0\leqslant t\leqslant T$.

Proof: (i) By the conservation of total mass and energy their bounds are preserved in time. Since $v_i(t)^2\geqslant 0$, the assertion concerning the potential energy is clear. For the kinetic energy we use the stability of the potential (replace A by its lower bound 0 in the superstability estimate)

$$\frac{1}{2}\alpha^d\sum_i v_i(t)^2\leqslant E-\frac{1}{2}\alpha^d{\sum_{i,j}}'\phi\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)$$

$$\leqslant E+B\alpha^d\sum_i 1\leqslant E+BM.$$

(ii) This follows from the estimate

$$\alpha^d\sum_i|q_i(t)-q_j(0)|^2=\alpha^d\sum_i\left|\int_0^t v_i(s)ds\right|^2$$

$$\leqslant\alpha^d\sum_i t\int_0^t v_i(s)^2\,ds$$

and part (i).

Now denote $\mu^\alpha_t:=\alpha^d\sum_i\delta_{x_i(t)}$.

Proposition 2.2: For each $T>0$ the set $\{\mu^t_\alpha:\alpha>0,\ 0\leqslant t\leqslant T\}$ is tight.

Proof: For $\alpha>0$, $0\leqslant t\leqslant T$ there follows from Lemma 2.1

$$\int|x|^2 d\mu^\alpha_t(x)=\alpha^d\sum_i(|q_i(t)|^2+|v_i(t)|^2)\leqslant C(T)$$

for some constant $C(T)>0$. Hence

$$\mu^\alpha_t(\{x:|x|\geqslant R\})\leqslant C(T)/R^2$$

which implies the result.

From this proposition we can conclude the following: Let $\{t_1,t_2,\dots\}$ be a dense subset of $\mathbb{R}^d$. Then for each sequence $\alpha_n\downarrow 0$ there exists a subsequence $\alpha_n(k)\downarrow 0$ such that $\mu^{\alpha_{n(k)}}_{t_j}$ converges weakly as $k\to\infty$ for each t_j.

This property is not very helpful. To get interesting results we need convergence for each $t\geqslant 0$. This needs suitable estimates of the differences $\mu^\alpha_t-\mu^\alpha_s$, which are generally not valid for the reasons mentioned in the last section. We can get sufficient estimates, however, for the local distributions of certain conserved quantities.

2.2. Convergence

If we apply the time derivatives (1.2) to the local distributions (1.3) of the conserved quantities, we get

$$\frac{d}{dt}\int\varphi(q)d\rho^\alpha_t(q)=\frac{d}{dt}\alpha^d\sum_i\varphi(q_i(t))$$

$$=\alpha^d\sum_i\nabla\varphi(q_i(t))v_i(t),\qquad(2.2a)$$

$$\frac{d}{dt}\int\varphi(q)dv^\alpha_t(q)=\frac{d}{dt}\alpha^d\sum_i\varphi(q_i(t))v_i(t)$$

$$=\alpha^d\sum_i(\nabla\varphi(q_i(t))v_i(t))v_i(t)$$

$$+\alpha^{d-1}\sum_i \varphi(q_i(t))\sum_{j\neq i}' \times F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)$$

$$=\alpha^d\sum_i(\nabla\varphi(q_i(t))v_i(t))v_i(t)+\frac{1}{2}\alpha^d\sum_{i,j}'\frac{\varphi(q_i(t))-\varphi(q_j(t))}{\alpha}\times F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right). \tag{2.2b}$$

In the last transformation we use the symmetry of ϕ. The singular term now remains bounded as $\alpha\downarrow 0$, if F decays sufficiently strongly at infinity. Rigorous estimates will follow below.

We omit the derivative of the energy distribution (1.3c), since we shall not need it. One can similarly verify that the singularity α^{-1} cancels at least formally in the same way.

For the mass and momentum distributions we also have tightness.

Proposition 2.3: For each $T>0$ the sets $\{\rho_t^\alpha:\alpha>0,\ 0\leqslant t\leqslant T\}$ and $\{\nu_t^\alpha:\alpha>0,\ 0\leqslant t\leqslant T\}$ are tight.

Proof: The tightness of the mass distribution is a direct consequence of the proof of Proposition 2.2.

The tightness of the momentum distribution follows from the estimate

$$\alpha^d\sum_i|v_i(t)|\mathbf{1}_{\{|q_i(t)|>R\}}\leqslant\left(\alpha^d\sum_i\mathbf{1}_{\{|q_i(t)|>R\}}\right)^{1/2}(\alpha^d\sum_i v_i(t)^2)^{1/2}.$$

(We denote the indicator function of a set C by $\mathbf{1}_C$.)

Now we fix a denumerable dense subset $\{t_1,t_2,\ldots\}$ of $\mathbb{R}^+$. For each sequence $\alpha_n\downarrow 0$ there exists a subsequence $\alpha_{n(k)}\downarrow 0$ such that $\rho_{t_j}^{\alpha_{n(k)}}$ and $\nu_{t_j}^{\alpha_{n(k)}}$ converge weakly as $k\to\infty$ for each t_j. We want to show that under suitable assumptions on ϕ the convergence holds for each $t\geqslant 0$. This is true, if for each $T>0$ and each continuously differentiable bounded φ with bounded derivatives

$$\frac{d}{dt}\int\varphi(q)d\rho_t^\alpha(q)\text{and}\frac{d}{dt}\int\varphi(q)d\nu_t^\alpha(q)$$

is uniformly bounded in $\alpha>0$ and $0\leqslant t\leqslant T$.

From (2.2b) it is easily seen, that this holds under the condition

$$\alpha^d\sum_{i,j}'\left|\frac{q_i(t)-q_j(t)}{\alpha}\right|\left|F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)\right|\text{ is uniformly bounded in }\alpha>0,\quad 0\leqslant t\leqslant T\text{ for each }T>0. \tag{2.3}$$

[For (2.2a) it holds without further assumptions.] In the dual of the bounded Lipschitz norm, which generates the weak topology,[15]

$$\|\mu\|_{\mathrm{BL}}^*=\sup\left\{\left|\int\varphi\,d\mu\right|:\|\varphi\|_{\mathrm{BL}}\leqslant 1\right\}$$

with

$$\|\varphi\|_{\mathrm{BL}}=\sup\{|\varphi(q)|:q\in\mathbb{R}^d\}+\sup\{|\varphi(q_1)-\varphi(q_2)|/|q_1-q_2|:q_1,q_2\in\mathbb{R}^d,\ q_1\neq q_2\},$$

the convergence of $\rho_t^{\alpha_{n(k)}}$ and $\nu_t^{\alpha_{n(k)}}$ is then even stronger, namely, uniform on $0\leqslant t\leqslant T$ for each $T>0$, and the limit distributions are weakly continuous in t. So we have to show the validity of (2.3) under suitable assumptions. We shall prove it for two different cases. Both reduce the boundedness condition (2.3) to that of the potential energy and require the common assumption:

there exists $\alpha>0$, $\delta>0$ such that

$$|q|\ |\nabla\phi(q)|\leqslant\alpha\phi(q)\quad\text{for }0<|q|\leqslant\delta. \tag{2.4}$$

This condition is not too restrictive. It prevents a singularity of infinite order at 0 and is fulfilled, if $\phi(q)$ behaves like $|q|^{-\alpha}$ near 0.[16] Furthermore we either assume that ϕ has a finite range or is of the following type which was studied by Dobrushin, Fisher, and Ruelle:

there exists $0<\delta_1<\delta_2$ and monotone decreasing functions

$$\varphi_1:(0,\delta_1]\to\mathbb{R}^+,\qquad \varphi^2:[\delta_2,\infty)\to\mathbb{R}^+ \tag{2.5}$$

such that (nonintegrable singularity at 0)

$$\phi(q)\geqslant\varphi_1(|q|)\quad\text{for }0<|q|\leqslant\delta_1,\qquad \int_0^{\delta_1}\varphi(t)t^{d-1}\,dt=+\infty, \tag{2.5a}$$

(integrable decay of the negative part at ∞)

$$\phi(q)\geqslant-\varphi_2(|q|)\quad|q|\geqslant\delta_2,\qquad \int_{\delta_2}^{\infty}\varphi_2(t)t^{d-1}\,dt<\infty. \tag{2.5b}$$

These potentials are superstable. We shall call them DFR potentials. In this case we have to assume additionally that $-|q|\ |\nabla\phi(q)|$ satisfies condition (2.5b) also.

We first treat the case of a finite range potential. The crucial step is the following lemma, which even holds under the general assumptions.

Lemma 2.4: $\alpha^d\sum_{i,j}'\mathbf{1}_{\{|q_i(t)-q_j(t)|<\alpha\}}$ is uniformly bounded in $\alpha>0$, $t\geqslant 0$.

Proof: Let $\mathbb{R}^d=\cup_k W_k$ be the decomposition of $\mathbb{R}^d$ into cubes as it is given in the definition of superstability (2.1). Fix $\alpha>0$, $t\geqslant 0$ and let $k,k'\in\mathbb{Z}^d$ be such that there exists $i\neq j$ with $|q_i(t)-q_j(t)|\leqslant\alpha$; $q_i(t)/\alpha\in W_k$; $q_j(t)/\alpha\in W_{k'}$. Then $|k^l-k'^l|\leqslant 1$ holds for the coordinates of k, k' $(1\leqslant l\leqslant d)$. For each k there are 3^d different k' with this condition. For these k,k', the number of pairs (i,j) with $q_i(t)/\alpha\in W_k$; $q_j(t)/\alpha\in W_{k'}$ can be estimated by

$$N(\mathbf{q}_N(t)/\alpha,W_k)\,N(\mathbf{q}_N(t)/\alpha,W_k')\leqslant\begin{cases}\frac{1}{2}(N(\mathbf{q}_N(t)/\alpha,W_k)^2+N(\mathbf{q}_N(t)/\alpha W_{k'})^2) & \text{for } k\neq k'\\ N(\mathbf{q}_N(t)/\alpha,W_k)^2 & \text{for } k=k'.\end{cases}$$

There follows from superstability (2.1)

$$\alpha^d\sum_{i,j}'\mathbf{1}_{\{|q_i(t)-q_j(t)|<\alpha\}}\leqslant 3^d\alpha^d\sum_k N\left(\frac{\mathbf{q}_N(t)}{\alpha},W_k\right)^2$$

$$\leqslant 3^d a^d A^{-1}\left\{\frac{1}{2}\sum_{i,j}{}' \phi\left(\frac{q_i(t)-q_j(t)}{a}\right)+BN\left(\frac{\mathbf{q}_N(t)}{a},\mathbb{R}^d\right)\right\},$$

which is uniformly bounded in $t>0$, $a\geqslant 0$ by Lemma 2.1. Now let ϕ be a potential of finite range. Without restriction we assume

$$\phi(q)=0 \quad \text{for } |q|\geqslant 1.$$

Condition (2.4) and the finite range property imply that there exists $\beta>0$ such that

$$|q|\ |F(q)|\leqslant \alpha\phi(q)+\beta \mathbb{1}_{\{|q|\leqslant 1\}} \quad \text{for } q\in\mathbb{R}^d,$$

and (2.3) follows from Lemmas 2.1 and 2.4.

The treatment of a DFR potential (2.5) is based on an ideal of Morrey,[8] which he used in his proof of Theorem 5.1. We formulate a more general statement in a separate lemma, which we shall need later on again.

Lemma 2.5: Let ϕ satisfy (2.5), and let ϕ_1 be a two-body potential with the properties

(1) there exists $\alpha,\delta>0$ such that

$$\phi_1(q)\leqslant\alpha\phi(q) \quad \textit{for } |q|\leqslant\delta,$$

(2) $-\phi_1$ satisifes (2.5b).

Then the potential $\Psi(q):=\phi(q)+(\alpha\phi(q)-\phi_1(q))$ is superstable and $a^d\Sigma_{i,j\phi_1}{}'([q_i(t)-q_j(t)]/a)$ is uniformly bounded in $a>0$, $t\geqslant 0$.

Proof: One easily verifies that Ψ is a DFR potential and hence superstable. The stability estimate of Ψ implies the second assertion.

We apply this Lemma under the mentioned assumptions to $\phi_1(q)=|q|\ |\nabla\phi(q)|$ and get (2.3). So we finally proved the following theorem.

Theorem 2.6: Assume that ϕ satisfies (2.4) and either has a finite range or is a DFR potential with the additional property that $-|q|\ |\nabla\phi(q)|$ satisfies (2.5b). Then for each sequence $a_n\downarrow 0$ there exists a subsequence $a_{n(k)}\downarrow 0$ such that $\rho_t^{a_{n(k)}}$ and $\nu_t^{a_{n(k)}}$ converge uniformly on each compact time interval with respect to the weak topology, and the limit distributions are weakly continuous in t.

2.3. Absolute continuity

The proof of the absolute continuity of the accumulation points of ρ_t^a and ν_t^a as $a\downarrow 0$ holds even for limits at fixed time and needs weaker conditions than the convergence of subsequences.

Theorem 2.7: Let ϕ satisfy at least one of the following conditions:

(1) $\phi\geqslant 0$,
(2) ϕ has finite range,
(3) ϕ is a DFR potential,

and let $t\geqslant 0$ and $a_n\downarrow 0$ be such that $\rho_t^{a_n}\to\rho_t$ and $\nu_t^{a_n}\to\nu_t$ weakly. Then ρ_t is absolutely continuous with respect to the Lebesgue measure, and ν_t is absolutely continuous with respect to ρ_t.

Proof: The first part of the proof only needs the superstability and coincides for the different cases.

We again use the decomposition of $\mathbb{R}^d$ into the cubes W_k, $k\in\mathbb{Z}^d$. Let $C\subset\mathbb{R}^d$ be a bounded set, and denote by $K(a)$ the number of cubes W_k with $(a^{-1}C)\cap W_k\neq 0$. Then the superstability (2.1) implies

$$\begin{aligned}\frac{1}{2}a^d\sum_{i,j}{}' &\mathbb{1}_C(q_i(t))\mathbb{1}_C(q_j(t))\ \phi\left(\frac{q_i(t)-q_j(t)}{a}\right)\\ &\geqslant a^d\left[A\sum_k N\left(\frac{\mathbf{q}_N(t)}{a},(a^{-1}C)\cap W_k\right)^2-BN(\mathbf{q}_N(t),C)\right]\\ &\geqslant A\left(\left[a^d\sum_k N\left(\frac{\mathbf{q}_N(t)}{a},(a^{-1}C)\cap W_k\right)\right]^2\Big/a^dK(a)\right)\\ &\quad -BM=A([a^dN(\mathbf{q}_N(t),C)]^2/a^dK(a))-BM.\end{aligned}$$

Now assume that the potential energy is bounded subsets $C\subset\mathbb{R}^d$, i.e., the first expression of the above estimates, is uniformly bounded in $a>0$ and C. The proof of this fact is different for the different assumptions and will be given below.

Then by the above estimate there exists a constant $L>0$ such that

$$a^dN(\mathbf{q}_N(t),C)\leqslant L(a^dK(a))^{1/2} \quad \text{for each } a>0.$$

If C is a finite union of bounded rectangles, then $a^dK(a)\to|C|$, the Lebesgue measure of C, as $a\downarrow 0$, and hence

$$\limsup_{a\downarrow 0} a^dN(\mathbf{q}_N(t),C)\leqslant L\ |C|^{1/2}.$$

The absolute continuity of ρ_t then follows from the following Lemma:

Lemma 2.8: Let $(\Omega,\mathcal{A})$ be a measurable space, and let $\mathcal{C}$ be a ring which generates $\mathcal{A}$. Let μ,ν be σ-finite measures on $\mathcal{A}$ with the property

for each $\epsilon>0$ there exists $\delta>0$ such that $\mu(A)<\epsilon$ for every $A\in\mathcal{C}$ with $\nu(A)<\delta$.

Then μ is absolutely continuous with respect to ν.

The easy proof of the Lemma uses outer measures to carry over the assumed property to sets A in $\mathcal{A}$ with "$<\epsilon$" replaced by "$\leqslant\epsilon$."

It remains to show the uniform boundedness of the potential energy in bounded subsets. It is trivial for the nonnegative case (1).

Let ϕ have a finite range. We again assume without restriction $\phi(q)=0$ for $|q|\geqslant 1$. Then

$$\begin{aligned}\frac{1}{2}a^d\sum_{i,j}{}' &\mathbb{1}_C(q_i(t))\mathbb{1}_C(\mathbf{q}_j(t))\phi\left(\frac{q_i(t)-q_j(t)}{a}\right)\\ &=\frac{1}{2}a^d\left[\sum_{i,j}{}'\phi\left(\frac{q_i(t)-q_j(t)}{a}\right)\right.\\ &\quad\left.-\sum_{i,j}{}'\mathbb{1}_{\{q_i(t)\notin C \text{ or } q_j(t)\notin C\}}\phi\left(\frac{q_i(t)-q_j(t)}{a}\right)\right].\end{aligned}$$

Since ϕ is bounded below by superstability, the second term is bounded above by Lemma 2.4.

In the case of a DFR potential the stronger results holds that $a^d\Sigma'_{i,j}\ |\phi((q_i(t)-q_j(t))/a)|$ is uniformly bounded in $a>0$. The proof uses the same idea as the proof of Lemma 2.5, though it directly does not apply. The potential $\Psi(q)=\phi(q)+(\phi(q)-|\phi(q)|)$ is a DFR potential. The property (2.5a) directly carries over from ϕ to Ψ, and property (2.5b) holds with φ_2 replaced by $3\varphi_2$. From this the uniform boundedness of $a^d\Sigma'_{i,j}\ |\phi([q_i(t)-q_j(t)]/a)|$ follows as in Lemma 2.5.

Finally, the absolute continuity of ν_t with respect to ρ_t follows from the estimate

$$\alpha^d \sum_i \mathbb{1}_C(q_i(t))|v_i(t)| \leqslant \left(\alpha^d \sum_i \mathbb{1}_C(q_i(t))\right)^{1/2} \left(\alpha^d \sum_i v_i(t)^2\right)^{1/2}.$$

2.4. Remarks

We conjecture that under the general assumptions concerning the initial configurations, the form of the results cannot be improved for the following reasons.

There is great evidence—see the intuitive, though not rigorous example in Ref. 16—that a small number of particles may collect a nonvanishing part of the energy in finite time and transport it to infinity or concentrate it in a small region. So neither tightness nor absolute continuity can be expected to hold for the energy distribution.

Furthermore, since the time evolution of ρ_t and ν_t is not closed and hence presumably not unique, ρ_t^α and ν_t^α may be near different limit evolutions for different small α. Then no convergence for $\alpha \downarrow 0$ can hold, but only convergence of subsequences.

3. ON THE LIMIT DYNAMICS

In the last chapter we proved a convergence and a regularity theorem for the local distributions of the mass and momentum. Now we study the time evolution of these limit distributions with regard to conclusions for the limit dynamics.

So let $\alpha_n \downarrow 0$ such that $\rho_t^{\alpha_n}$ and $\nu_t^{\alpha_n}$ converge weakly on some time interval $0 \leqslant t \leqslant T$. We denote the density of $\lim_{n\to\infty} \rho_t^{\alpha_n}$ with respect to the Lebesgue measure by ρ_t and the density of $\lim_{n\to\infty} \nu_t^{\alpha_n}$ with respect to ρ_t by u_t $(0 \leqslant t \leqslant T)$.

It is easy to derive the continuity equation in a weak form. We integrate Eq. (2.2a) with respect to time, take the limit $\alpha_n \downarrow 0$ and differentiate. Then there follows

$$\frac{d}{dt}\int \varphi(q)\rho_t(q)dq = \int \nabla\varphi(q)\, u_t(q)\, \rho_t(q)dq$$

for continously differentiable, bounded functions φ with bounded derivatives. One cannot proceed in the same way with Eq. (2.2b) since the derivative consists of terms whose convergence may not hold for similar reasons as mentioned in Sec. 2.4. But we can study this equation in the integrated version and examine in particular the behavior of the force term in order to explain how the hydrodynamical pressure arises in the limit. We first prove a convergence theorem for this term and then discuss its limit behavior in a formal way.

We represent the force as a vector-valued measure. So denote the force at time t exerted on particles in a Borel set $A \subset \mathbb{R}^d$ by

$$\overline{F}_t^\alpha(A) = \alpha^{d-1}{\sum_{ij}}' \mathbb{1}_A(q_i(t))F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)$$

$$= \alpha^{d-1}\sum_{i,j} \mathbb{1}_A(q_i(t))\mathbb{1}_{CA}(q_j(t))F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right).$$

This equality holds, since F is odd, and is analogous to the transformation of (2.2b).

Thus $\overline{F}_t^\alpha$ is the vector-valued measure

$$\overline{F}_t^\alpha = \alpha^{d-1}\sum_i \left(\sum_{j\neq i} F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)\right)\delta_{q_i(t)}.$$

If we integrate a test function φ with respect to $\overline{F}_t^\alpha$, then we just get the force term of (2.2b). Integration with respect to time yields

$$\int_0^t \left(\int \varphi(q)\, d\overline{F}_s^\alpha(q)\right)ds$$

$$= \int_0^t \left(\frac{1}{2}\alpha^d {\sum_{ij}}' \frac{\varphi(q_i(s))-\varphi(q_j(s))}{\alpha} F\left(\frac{q_i(s)-q_j(s)}{\alpha}\right)\right)ds.$$

This can be approximated for small α and smooth φ by

$$\int_0^t \left(\frac{1}{2}\alpha^d {\sum_{ij}}' \left(\nabla\varphi(q_i(s))\frac{q_i(s)-q_j(s)}{\alpha}\right) F\left(\frac{q_i(s)-q_j(s)}{\alpha}\right)\right)ds. \tag{3.1}$$

The difference of both expressions tends to 0 as $\alpha \downarrow 0$ if

$$\alpha^d {\sum_{ij}}' \left|\frac{q_i(t)-q_j(t)}{\alpha}\right|^2 \left|F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)\right|$$

is uniformly bounded in $\alpha > 0$, $0 \leqslant t \leqslant T$. (3.2)

Conditions, under which (3.2) holds, are derived in the same way as for (2.3). In the finite range case (2.3) implies (3.2), and in the case of a DFR potential, the decay of $|F(q)|$ at infinity has to be one power stronger.

Expression (3.1) can be represented as

$$\int \nabla\varphi(q)\, d\Pi_t^\alpha(q)$$

with the tensor valued measure Π_t^α given by

$$(\Pi_t^\alpha)^{k,l} = \int_0^t \frac{1}{2}\alpha^d \left[\sum_i \left(\sum_{j\neq i} \frac{q_i^k(s)-q_j^k(s)}{\alpha} \times F^l\left(\frac{q_i(s)-q_j(s)}{\alpha}\right)\right)\delta_{q_i(s)}\right]ds,$$

$$(1 \leqslant k,l \leqslant d).$$

We want to derive for Π_t^α a convergence theorem similar to Theorem 2.6. But again for reasons as mentioned in Sec. 2.4 there is no tightness. We can avoid this difficulty by replacing the weak topology by the vague one. The set $\{\Pi_t^\alpha : \alpha > 0,\ 0 \leqslant t \leqslant T\}$ is relatively compact with respect to the vague topology, if

$$\alpha^d {\sum_{ij}}' \left|\frac{q_i(t)-q_j(t)}{\alpha}\right| \left|F\left(\frac{q_i(t)-q_j(t)}{\alpha}\right)\right|$$

is uniformly bounded in $\alpha > 0$, $0 \geqslant t \geqslant T$ (compare 2.3).

The same condition is needed to estimate the differences $\Pi_t^\alpha - \Pi_s^\alpha$. Then we can proceed as in Sec. 2.2 and prove a convergence theorem for Π_t^α. If we combine this with the approximation of $\int_0^t \overline{F}_s^\alpha\, ds$ by $-\nabla\Pi_t^\alpha$ in the sense of distributions we get the following.

Theorem 3.1: Assume that ϕ satisfies (2.4) and either has a finite range or is a DFR potential with the additional property that $-|q|^2|F(q)|$ satisfies (2.5b). Then for each sequence $\alpha_n \downarrow 0$ there exists a subsequence $\alpha_{n(k)} \downarrow 0$ such that $\Pi_t^{\alpha_{n(k)}}$ con-

verges vaguely to some Π_t and $\int_0^t \overline{F}_s^{a_n(k)} ds$ converges to $-\nabla \Pi_t$ as distributions $(0 \leqslant t \leqslant T)$. The dependence of Π_t on t is vaguely continuous.

Thus the limit behavior of the force leads us to the investigation of Π_t. We do this in a formal way, i.e., we calculate as if Π_t had a sufficiently smooth density. Let P_t denote the density of $(d/dt)\,\Pi_t$ with respect to the Lebesgue measure. Then $\int \varphi(q) d\overline{F}_t^{a_n(k)}(q)$ converges to

$$\int \nabla \varphi(q) P_t(q) dq = -\int \varphi(q) \nabla P_t(q) dq.$$

We now may set $\varphi = \mathbb{1}_A$ for sets $A \subset \mathbb{R}^d$ with sufficiently smooth boundary and get

$$\overline{F}_t^{a_n(k)}(A) \to -\int_A \nabla P_t(q) dq = -\int_{\partial A} n(q) P_t(q) d\sigma(q)$$

with the outer normal n and the surface measure σ on ∂A. This is the usual hydrodynamical form of the force with pressure tensor P_t.

Obviously, the same reasoning that led to Theorem 3.1 holds for the local distributions of the potential energy and the kinetic energy and for the $\mu_t^a (a > 0,\ t \geqslant 0)$, integrated with respect to time. We omit an explicit formulation of a corresponding theorem, since we shall draw no physical conclusions from it.

4. RANDOM CONFIGURATIONS: A FORMAL EXPANSION

As conjectured in Sec. 2.4 the kind of results of Sec. 2 cannot be improved under the general assumptions concerning the initial configurations. On the other hand, they are obviously not satisfactory. For stronger results we shall need special initial configurations.

In this section we shall approach this problem by treating the case of random configurations. The evolution equations for the correlation functions, the BBGKY equations, lead to a singular perturbation problem. Approximate outer solutions are the correlation functions of local equilibrium. So they are the natural candidates for special random initial configurations. Because of the great difficulties the procedure in this chapter is only formal.

Let $(\rho_{n,t}^a)_{n>1}$ denote the sequence of correlation functions of random configurations, which evolve according to the dynamics (1.1). They satisfy the BBGKY equations (see, e.g., Ref. 6):

$$\begin{aligned}\frac{\partial \rho_{n,t}^a(\mathbf{x}_n)}{\partial t} = &-\sum_{i=1}^n v_i \nabla_{q_i} \rho_{n,t}^a(\mathbf{x}_n) \\ &- a^{-1} \sum_{i=1}^n \left(\sum_{j \neq i} F\left(\frac{q_i - q_j}{a}\right)\right) \nabla_{v_i} \rho_{n,t}^a(\mathbf{x}_n) \\ &- a^{-1} \sum_{i=1}^n \int F\left(\frac{q_i - q_{n+1}}{a}\right) \\ &\times \nabla_{v_i} \rho_{n+1,t}^a(\mathbf{x}_n, x_{n+1}) dx_{n+1} \quad (n \geqslant 1). \end{aligned} \tag{4.1}$$

We introduce the local correlation functions by means of the following transformation of the space variables:

$$\begin{aligned} q_1 &= q, \\ q_2 &= q + a\xi_2, \\ &\vdots \\ q_n &= q + a\xi_n \end{aligned}$$

The corresponding correlation functions describe the distribution of particles in a microscopic neighborhood of the macroscopic point q. The appropriate normalization is

$$\begin{aligned} &r_{n,t}^a(q, v_1, \xi_2, v_2, \ldots, \xi_n, v_n) \\ &\quad := a^d \rho_{n,t}^a(q, v_1, q + a\xi_2, v_2, \ldots, q + a\xi_n, v_n). \end{aligned}$$

Equations (4.1) are transformed into equations of the form

$$\frac{\partial r_{n,t}^a}{\partial t} - v_1 \nabla_q r_{n,t}^a = a^{-1}\left[L_n r_{n,t}^a + C_n r_{n+1,t}^a\right] \quad (n \geqslant 1). \tag{4.2}$$

The operators L_n and C_n do not depend on a and t and do not act on the macroscopic variable q.

We write it shortly in the form

$$D_t \mathbf{r}_t^a = a^{-1} A \mathbf{r}_t^a \tag{4.2'}$$

for the sequence of functions $\mathbf{r}_t^a = (r_{n,t}^a)_{n>1}$. This is a familar kind of singular perturbation problem. It is well known in physics from the Champman–Enskog solution of the Boltzmann equation.

The first-order outer problem is the equation $A\mathbf{r}_t = 0$. In order to solve this equation we transform the BBGKY equations to a purely microscopic level by setting additionally

$$\begin{aligned} &q = q_0 + a\xi \text{ with a fixed } q_0, \\ &\tau = ta^{-1}. \end{aligned}$$

The corresponding equations for the microscopic correlation functions $f_{n,\tau}(\xi, v_1, \xi_2, v_2, \ldots, \xi_n, v_n) := a^{-d} r_{n,a\tau}(q_0 + a\xi, v_1, \xi_2, \ldots, v_n)$ do not depend on a. They are

$$\frac{\partial f_{n,\tau}}{\partial \tau} - v_1 \nabla_\xi f_{n,\tau} = L_n f_{n,\tau} + C_n f_{n+1,\tau} \quad (n \geqslant 1),$$

with the same operators L_n and C_n as above. Remember that they do not act on the transformed variables.

The equilibrium correlation functions in these variables,

$$f_{n,\mathrm{eq}}(\xi, v_1, \xi_2, \ldots, v_n; \rho, u, e) \quad (n \geqslant 1),$$

are time-independent solutions. They are parameterized by the mean density ρ, velocity u, and energy e. Since they are translational-invariant, they do not depend on ξ and hence satisfy $\nabla_\xi f_{n,\mathrm{eq}} = 0$. There follows

$$L_n f_{n,\mathrm{eq}} + C_n f_{n,\mathrm{eq}} = 0 \quad (n \geqslant 1).$$

Therefore the correlation functions of local equilibrium given by

$$\begin{aligned} &r_{n,t}(q, v_1, \xi_2, \ldots, v_n) \\ &\quad = a^d f_{n,\mathrm{eq}}(\xi, v_1, \xi_2, \ldots, v_n; \rho(q,t), u(q,t), e(q,t)), \quad (n \geqslant 1), \end{aligned}$$

with arbitrary functions $\rho(q,t)$, $u(q,t)$, $e(q,t)$, are solutions of the outer equation $A\mathbf{r}_t = 0$.

If we insert these solutions into the conservation equa-

tions (2.2a) and (2.2b) completed by the conservation equation for the energy, and modified for the random case by taking the expectation, we get the system of the Euler equations

$$\frac{\partial}{\partial t}\rho + \nabla_q(\rho u) = 0,$$

$$\frac{\partial}{\partial t}(\rho u) + \nabla_q(\rho u\, u + p) = 0,$$

$$\frac{\partial}{\partial t}(\rho e) + \nabla_q\{\rho u(e+p)\} = 0.$$

The pressure p can be explicitly represented as a function of ρ, u, e.[8] This equation of state is the same dependence as the one derived from the virial theorem,[17] which is not surprising after our calculations in Sec. 3.

The procedure that we carried out in this section is a modification of the Chapman–Enskog method. It seems to be extremely difficult to make it rigorous. An even more difficult problem would be to study the transition from more general initial configurations in an initial time layer of the order α to local equilibrium distributions.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 123). The final version of the results was derived during my inspiring stay at the Institute for Mathematics and its Applications, University of Minnesota. I want to thank their directors, Professor H. Weinberger and Professor G. R. Sell, for their kind hospitality.

[1]C. Boldrighini, R. L. Dobrushin, Yu. A. Sukhov, "The one-dimensional hard rod caricature on hydrodynamics," preprint 1982.
[2]R. L. Dobrushin and R. Siegmund-Schulze, "The hydrodynamical limit for systems of particles with independent evolution," Math. Nachr. **105**, 199–224 (1982).
[3]A. Galves, C. Kipnis, C. Marchioro, and E. Presutti, "Nonequilibrium measures which exhibit a temperature gradient: study of a model," Commun. Math. Phys. **81**, 127–147 (1981).
[4]H. Rost, "Diffusion de lignes dures dans la droite réelle:comportement macroscopique et équilibre local," preprint 1983.
[5]H. Spohn, "Kinetic equations from Hamiltonian systems: Markovian limits," Rev. Mod. Phys. **53**, 569–615 (1980).
[6]O. E. Lanford, "Time evolution of large classical systems," in *Dynamical Systems: Theory and Applications*, edited by J. Moser, *Springer Lecture Notes in Physics*, Vol. 38, pp. 1-11 (Springer, Berlin Heielberg New York, 1975).
[7]M. G. Mürmann, "Limit dynamics of dense gases," in *Colloquia Mathematica Societatis János Bolyai*, Vol. 27, pp. 787–797 (North-Holland, Amsterdam, 1982).
[8]C. B. Morrey, "On the derivation of the equations of hydrodynamics from statistical mechanics," Comm. Pure Appl. Math. **8**, 279–326 (1955).
[9]W. Braun and K. Hepp, "The Vlasov dynamics in the $1/N$ limit of interacting classical particles," Commun. Math. Phys. **56**, 101–113 (1977).
[10]H. Neunzert, "Mathematical investigations on particle-in-cell methods," Fluid Dyn. Trans. **9**, 229–254 (1978).
[11]O. E. Lanford, "On a derivation of the Boltzman equation," Soc. Math. France Astérisque **40**, 117–137 (1976).
[12]In the Boltzmann–Grad limit this term also becomes singular for an interacting particle, but this is compensated by the fact that the number of interacting particles at a fixed time becomes relatively small.
[13]See Ref. 5, Theorem 3.2 for the exact formulation in which sense this holds.
[14]D. Ruelle, "Superstable interactions in classical statistical mechanics," Commun. Math. Phys. **18**, 127–159 (1970).
[15]R. M. Dudley, "Convergence of Baire measures," Studia Math. **27**, 251–268 (1966).
[16]J. Fritz and R. L. Dobrushin, "Non-equilibrium dynamics of two-dimensional infinite particles systems with a singular interaction," Commun. Math. Phys. **57**, 67–81 (1977).
[17]J. de Boer,, "Molecular distribution and equation of state of gases," Rep. Prog. Phys. **12**, 305–375 (1949).

A KdV soliton propagating with varying velocity

C. Au and P. C. W. Fung
Physics Department, University of Hong Kong, Hong Kong

(Received 10 February 1983; accepted for publication 22 July 1983)

Using one solution to the KdV equation found previously by Au and Fung [Phys. Rev. B **25**, 6460 (1982)] we apply the Bäcklund transformation again to obtain a new set of solutions which are divergent at certain points in the special case where the vacuum parameter is zero. While one set of solutions is static, the other set is an asymptotic one-soliton solution propagating with a varying velocity in a "transient" domain of space and time. To demonstrate the main features of our discovery, we have carried out a detailed numerical analysis of our analytical solutions.

PACS numbers: 03.40.Kf

I. INTRODUCTION

Recently,[1] we have obtained analytical solutions to the KdV equation via a Bäcklund transformation[2,3] using a differential geometrical approach. The appearance of a vacuum parameter b (which represents the disturbance at infinite distance) in our solutions leads us to the result that a KdV soliton can propagate to both directions and that the amplitude width and velocity of the soliton are functions of this vacuum parameter b. Our result suggests that different vacuum states have different effects on the observable physical state.

In this paper, we follow up our previous work by obtaining another set of new analytical solutions to the KdV equation using one solution obtained earlier via a repeated use of our Bäcklund transformation in the special case $b = 0$. Our new set of solutions are divergent at certain points, and the soliton propagates with varying velocity values as it propagates through a spatial region bounded by the stated singularities. As the propagation properties are rather complicated, we carry out a numerical analysis of our analytical solutions, in order to demonstrate the main features of our new soliton.

II. NEW ANALYTICAL SOLUTIONS TO THE KdV EQUATION ($b = 0$)

Previously, we have obtained solutions to the KdV equation

$$u_t + u_{xxx} + 12uu_x = 0 \tag{1}$$

via the Bäcklund transformation

$$u^* = b, \tag{2}$$

$$u^* = u(x, t), \tag{3}$$

$$u^* = -u(x, t) - y^2 + \lambda, \tag{4}$$

where $y(x, t)$ must satisfy

$$y_x = -2u(x, t) - y^2 + \lambda, \tag{5}$$

$$y_t = -4[u(x, t) + \lambda]y_x + 2u_{xx} - 4u_x y. \tag{6}$$

These solutions are

$$u^* = b, \tag{7}$$

$$u^* = b - \frac{1}{(x - 12bt - x_0)^2}, \quad \lambda = 2b, \tag{8}$$

$$u^* = \lambda - b - (\lambda - 2b)\frac{(Ce^{\sqrt{\lambda - 2b}r} - e^{-\sqrt{\lambda - 2b}r})^2}{(Ce^{\sqrt{\lambda - 2b}r} + e^{-\sqrt{\lambda - 2b}r})^2}, \tag{9}$$

where $r = x - 4(b + \lambda)t$. Solution (9) has been analyzed.[1] In this paper we shall take our solution (8) under the simple situation $b = 0$ (and $x_0 = 0$), namely,

$$u = -1/x^2 \tag{10}$$

and attempt to obtain explicit expressions for the function y. Obviously, since our "seed" in (10) is a divergent function at $x = 0$, our solutions to the KdV equation found later will also have divergent properties. Using this set of y solutions, we can obtain a new set of u^* solutions by using again our Bäcklund transformation (4). In our procedure, we first write (6) in the following form, using (10):

$$4(\lambda - 1/x^2)y_x + y_t = -4[3/x^4 + (2/x^3)y]. \tag{11}$$

The characteristic equations related to the above first-order inhomogeneous linear partial differential equation are

$$\frac{dx}{4(\lambda - 1/x^2)} = dt = \frac{dy}{-4[3/x^4 + (2/x^3)y]}. \tag{12}$$

Note that for $\lambda > 0$, $\lambda < 0$, and $\lambda = 0$, the solutions (5) and (12) belong to three different types, giving the following solutions after some elementary manipulation:

$$\left.\begin{aligned} y &= \frac{1}{x(\lambda x^2 - 1)} + \frac{\lambda x^2}{|\lambda x^2 - 1|}(\pm\sqrt{\lambda}) \\ y &= \frac{1}{x(\lambda x^2 - 1)} + \frac{\lambda x^2}{(\lambda x^2 - 1)}(\pm\sqrt{\lambda}) \end{aligned}\right\} \quad \text{for } \lambda > 0, \tag{13}$$

$$y = \frac{1}{x(\lambda x^2 - 1)} + \frac{\lambda^{3/2}x^2}{(\lambda x^2 - 1)}\left(\frac{Ce^{\sqrt{\lambda}\xi} - e^{-\sqrt{\lambda}\xi}}{Ce^{\sqrt{\lambda}\xi} + e^{-\sqrt{\lambda}\xi}}\right) \quad \text{for } \lambda \neq 0, \tag{14}$$

$$y = -\frac{1}{x} + \frac{x^2}{\frac{1}{3}x^3 + 4t + D} \quad \text{for } \lambda = 0. \tag{15}$$

Here D is a constant, and we are not considering the case of complex y in this paper; also, the time-dependent argument ξ is given by

$$\xi = x + \frac{1}{2\sqrt{\lambda}}\ln\left\{\left|\frac{\sqrt{\lambda}x - 1}{\sqrt{\lambda}x + 1}\right|\right\} - 4\lambda t \tag{16}$$

for $\lambda > 0$ and

$$\xi = x - (1/k)\arctan(kx) - 4\lambda t \tag{17}$$

for $\lambda > 0$.

 0022-2488/84/051364-06$02.50

We would like to remark that superficially C can take on any value in (14); however, there are only two different classes of solution specified by $C = \pm 1$, as mentioned in Ref. 1.

Substituting relations (13)–(15) into (4), keeping in mind solution (10), we obtain

$$\left.\begin{aligned} u^* &= -\lambda\frac{(\lambda x^2+1)}{(\lambda x^2-1)^2} \mp \frac{2\lambda^{3/2}x}{(\lambda x^2-1)|\lambda x^2-1|} \\ u^* &= -\lambda\frac{(\lambda x^2+1)}{(\lambda x^2-1)^2} \mp \frac{2\lambda^{3/2}x}{(\lambda x^2-1)^2} \end{aligned}\right\} \quad \text{for } \lambda > 0, \tag{18}$$

$$u^* = \frac{\lambda x^2+1}{x^2} - \left\{\frac{1}{x^2(\lambda x^2-1)^2} + \frac{2\lambda^{3/2}x}{(\lambda x^2-1)^2}\left(\frac{Ce^{\sqrt{\lambda}\xi}-e^{-\sqrt{\lambda}\xi}}{Ce^{\sqrt{\lambda}\xi}+e^{-\sqrt{\lambda}\xi}}\right) + \frac{\lambda^3x^4}{(\lambda x^2-1)^2}\frac{[Ce^{\sqrt{\lambda}\xi}-e^{-\sqrt{\lambda}\xi}]^2}{[Ce^{\sqrt{\lambda}\xi}+e^{-\sqrt{\lambda}\xi}]^2}\right\} \quad \text{for } \lambda \neq 0 \tag{19}$$

(Loh has obtained this solution[4] for $\lambda > 0$), and

$$u^* = \frac{2x}{\frac{1}{3}x^3+4t+D} - \frac{x^4}{(\frac{1}{3}x^3+4t+D)^2} \quad \text{for } \lambda = 0. \tag{20}$$

These are our new divergent solutions to the KdV equation for $\lambda \neq 0$. It is not difficult to check by direct substitution that they satisfy the KdV equation. By inspection, we see that there is a static contribution to solution (19). The amplitude here is a function of distance. There is a term, namely,

$$\frac{2\lambda^{3/2}}{(\lambda x^2-1)} \times \left(\frac{Ce^{\sqrt{\lambda}\xi}-e^{-\sqrt{\lambda}\xi}}{Ce^{\sqrt{\lambda}\xi}+e^{-\sqrt{\lambda}\xi}}\right),$$

which is a function of time and distance. For small x, this term influences the propagation properties of u^* around the region $-\sqrt{\lambda} < x < \sqrt{\lambda}$. At large x, this terms tends to zero, and the whole solution (19) then becomes

$$u^* = \lambda \operatorname{sech}^2\sqrt{\lambda}(x - 4\lambda t) \tag{21}$$

for $C = 1$. We thus call solution (19) the asymptotic one-soliton KdV solution.

III. ANALYSIS

In order to demonstrate the interesting propagating properties of soliton (19), we shall compute u as a function of x as time evolves, taking $C = 1.0$ and $\lambda = 1.0$ (we shall drop the superscript * in u^* from now on, and consider only the situation $\lambda > 0$).

Note that ξ is the only quantity depending on time in solution (19). In order to study some basic dynamic characteristics of solution (19), we consider the variation of x with t according to (16) for constant ξ. We choose $\xi = 0$ in particular and show such an x–t plot in Fig. 1.

As we follow the condition $\xi = 0$ when time evolves from $-\infty$ to ∞, we see that the x space is separated into three regions (see Fig. 1):

(I) from $x = -\infty$ to $x = -1/\sqrt{\lambda}$;

(II) from $x = -1/\sqrt{\lambda}$ to $x = 1/\sqrt{\lambda}$;

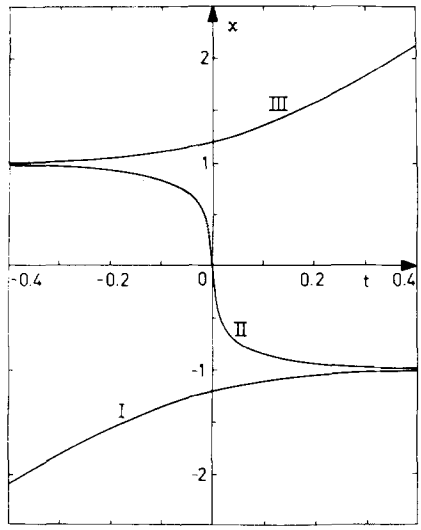


FIG. 1. x vs t following $\xi = 0$ according to Eq. (16).

(III) from $x = 1/\sqrt{\lambda}$ to $x = \infty$.

At every instant, there are three values of x for which the condition $\xi = 0$ is satisfied.

To show some basic specific propagation characteristics of the soliton, we plot in Fig. 2 the variation of the disturbance u with x at various specified times according to Eq. (19). At $t = -2.0$, Fig. 2(a) shows that a one-soliton appears on the left side. At $t = 0$ [Fig. 2(b)], the soliton has entered the middle branch II and we see that as if it has been cut off at $x = -1/\sqrt{\lambda} = -1$ and reappears at $x = 1/\sqrt{\lambda} = 1$. We shall study the interesting features of propagation through branch II shortly. In the meantime we show the soliton at $t = 1.0$ and $t = 2.0$ in Figs. 2(c) and 2(d). We observe in our analysis of other graphs that the amplitude of the soliton at large x is unity and remains the same most of the time [see the asymptotic one-soliton solution as represented by (20) for large x]. Close to branch II, however, the amplitude has values different from $\lambda = 1$.

It is worth noting that previously the velocity of a KdV soliton is taken to keep a constant value of 4λ. Even with the inclusion of our vacuum parameter, the velocity remains constant as it propagates. The solution to our soliton found in this investigation, however, gives a varying velocity as the soliton propagates. The dependence of velocity on distance is indicated in Fig. 3. From this study, we observe another new feature of nonlinearity. We would point out that the KdV equation has a well-known one-soliton solution with uniform velocity, and at the same time the equation also allows a one-soliton solution corresponding to propagation with varying velocity, as reported here. In linear physics, the change of velocity is caused by an external influence. Here in a nonlinear situation, the solution itself affects its own velocity, like a dog bites its own tail. In our case, from Eq. (6), we see that the solution $u(x, t) = -1/x^2$ decides the velocity variation since ξ is deduced from (12). Based on Eq. (4), $u(x, t) = -1/x^2$ constitutes part of the new solution $u^*(x, t)$. It is exactly due to the singularity of $u(x)$ at $x = 0$ that the velocity tends to $-\infty$ at $x = 0$. In fact, such a nonlinear nature is

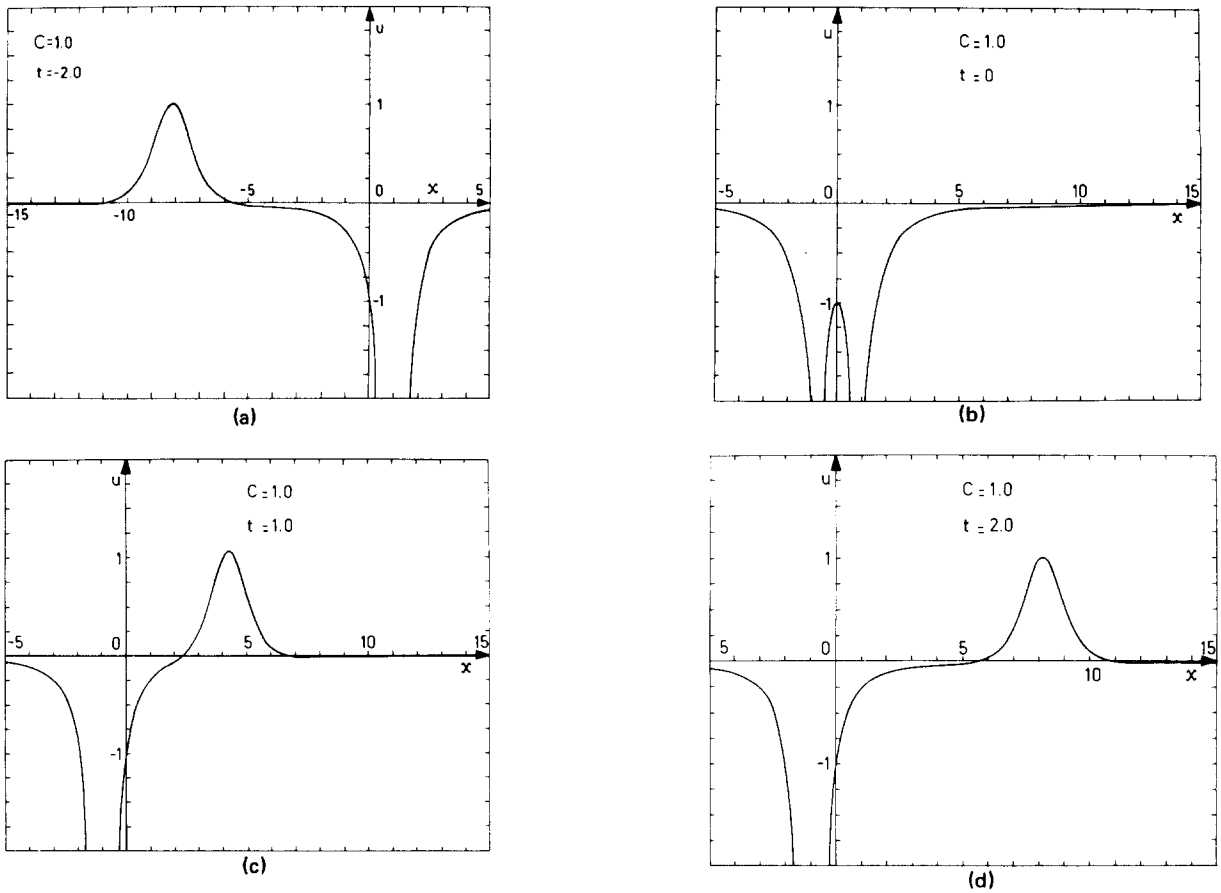


FIG. 2. A KdV soliton propagates to the right according to Eq. (19), taking $C = 1.0$ and $\lambda = 1.0$. The u–x plot is calculated at (a) $t = -2.0$, (b) $t = 0.0$, (c) $t = 1.0$, (d) $t = 2.0$.

governed by the nonlinear term uu_x in the KdV equation.

Now we shall leave the asymptotic behavior and turn to study the transient behavior of the soliton as it propagates through branch II. As clearly seen in Eq. (19), the u solution is composed of a static part and a dynamic part. In Figs. 4(a)–(m) we show the u–x plot during different instants of time: -0.25, -0.20, -0.15, -0.10, -0.05, 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.50, and 0.87. In each diagram, the dotted line indicates the static contribution and the dash–dot line indicates the dynamic contribution of Eq. (19). The solid line shows the overall u solution. During the transient period specified roughly in time scale by $-1.0 < t < 1.0$, we do not observe the "normal" propagation of the soliton through branch II. In fact, the value of u vibrates up and down in all the three branches, I, II, and III, while propagating slowly to the right. We would note that there are two singular points specified by $x = \pm 1/\sqrt{\lambda}$ in Eq. (19). When the soliton propagates, the dynamic contribution influences the static contribution, causing rather complicated behavior around branch II. The points $x = \pm 1/\sqrt{\lambda}$ may or may not be singular points at different times, depending on whether u approaches the vertical lines $x = \pm 1/\sqrt{\lambda}$ from the left or from the right. For example, in Fig. 4j, at $t = 0.20$, $x = 1$ is not a singular point. Across these two lines, there are discontinuities of the u solution during the transient domain. Slightly later than $t = 0.87$ in this case ($C = 1.0$, $\lambda = 1.0$), branch II joins branch III and the soliton propagates out in the manner indicated in Fig. 2.

IV. CONCLUSIONS

(1) In this investigation, we have used the divergent KdV solution $u = -1/x^2$ obtained earlier[1] and obtained a new set of solutions to the KdV equation via our Bäcklund transformation (4)–(6) again. This set of solutions for a nonzero value of the λ parameter is given in (18)–(19). The solution given in (18) is static and more obvious. The dynamic

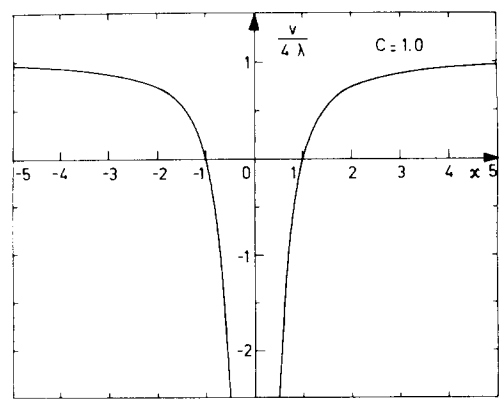
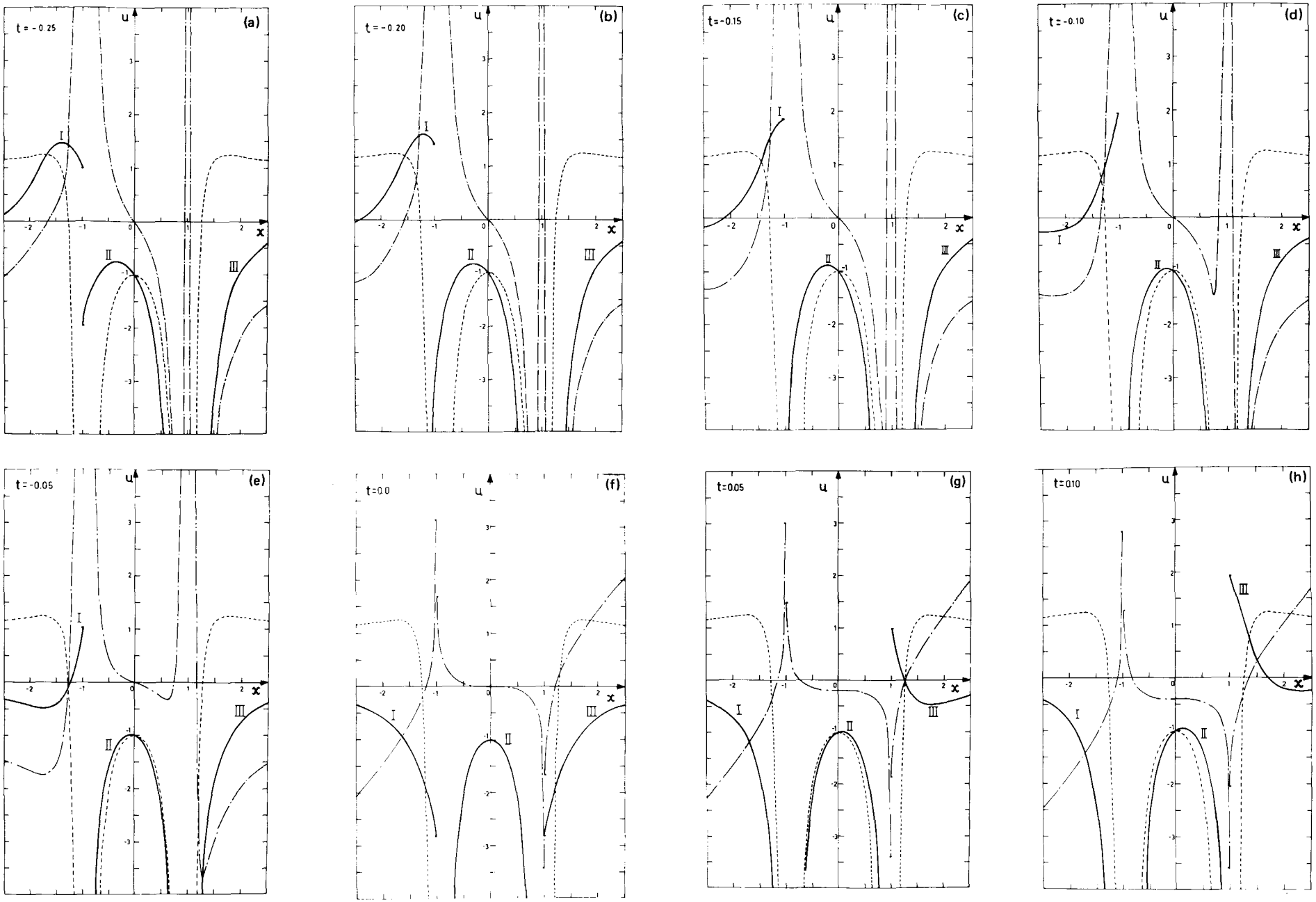
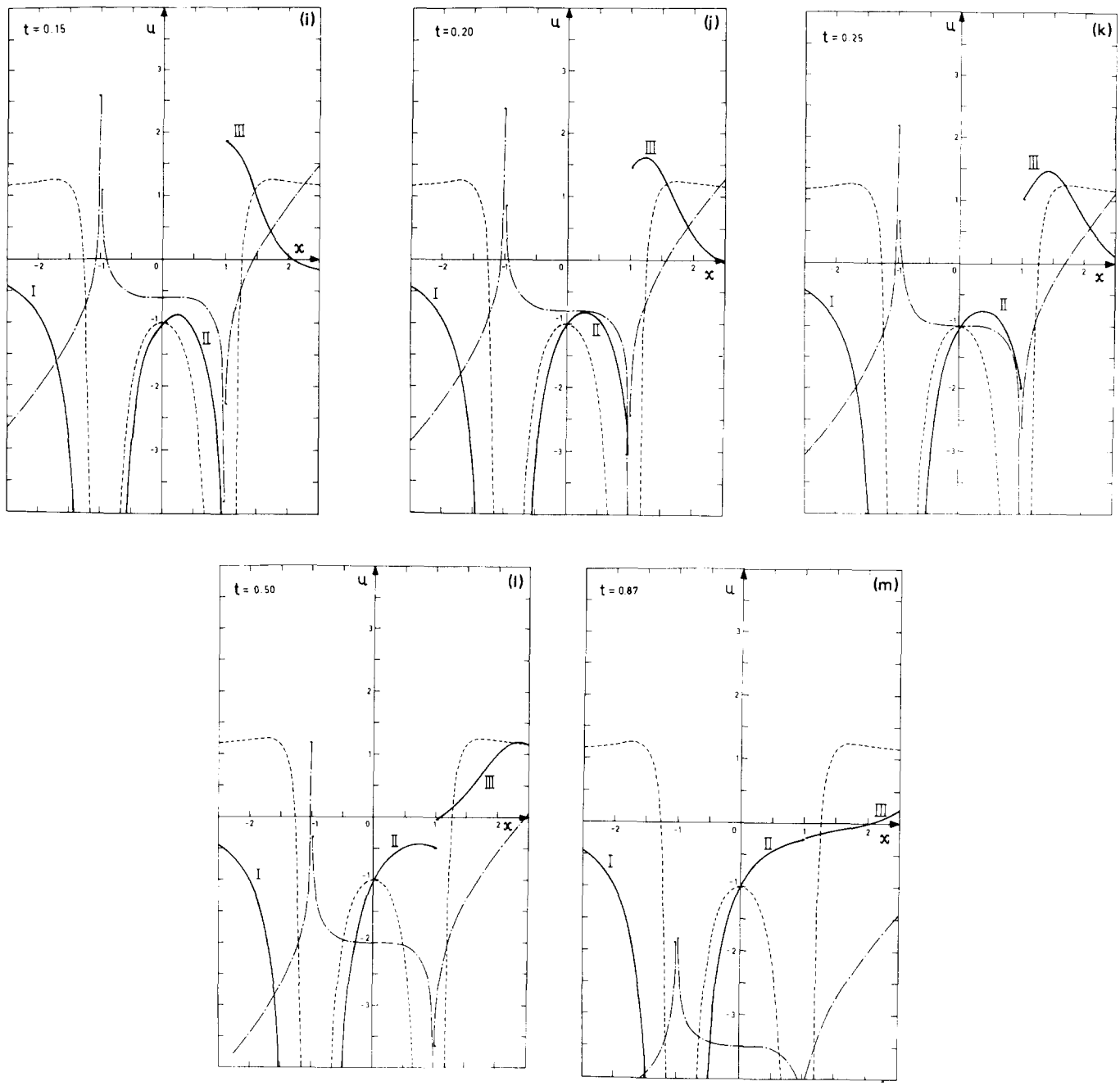


FIG. 3. Variation of the normalized soliton velocity $v/(4\lambda)$ with position x.



FIGS. 4 (a)–(h). Transient behavior of the one-soliton KdV solution according to Eq. (19) for $C = 1.0$, $\lambda = 1.0$ at different time values: (a) $t = -0.25$; (b) $t = -0.20$; (c) $t = -0.15$; (d) $t = -0.10$; (e) $t = -0.05$; (f) $t = 0$; (g) $t = 0.05$; (h) $t = 0.10$.



FIGS. 4(i)–(m). (i) $t = 0.15$; (j) $t = 0.20$; (k) $t = 0.25$; (l) $t = 0.50$; (m) $t = 0.87$.

solution (19) is an asymptotic one-soliton solution. These solutions can be divergent at $x = 0, \pm 1/\sqrt{\lambda}$.

(2) Solution (19) has a transient domain in space and time. The existence of singular points at $x = \pm 1/\sqrt{\lambda}$ separates the space into three regions, I, II, and III. At time $t = 0$, no soliton satisfying the normal soliton definition appears [Fig. 2(b)]. As time evolves, disturbance u of each branch oscillates while propagating relatively slowly to the right. When branch II has joined up to branch III, the soliton appears and transmits to the right. The soliton, however, does not travel with a constant velocity ($= 4\lambda$) as assumed in the past. The variation of velocity with distance is shown in Fig. 3. In linear physics, we need an external force to change the velocity of an object. In the nonlinear regime, the value of the velocity can be "self-adjusted" to vary.

(3) Since we have chosen a coordinate system such that the singularities $-1/\sqrt{x}$, $1/\sqrt{x}$ are symmetrical with respect to the origin, there appears to be an "absolute time" instant ($t = 0$) in our solution (19). The solution to u at $t = -t_1$ say [see Fig. 4(a), $t_1 = 0.25$] is identical to u at $t = 0.25$ if x is replaced by $-x$. This image property in time is very interesting and we believe that it is significant physically.

(4) If we have physical situations represented by the

KdV equation, the features of the propagating soliton (19) reported here could be detected. Further analysis needs to be carried out before we can understand fully the physical implications of a KdV soliton propagating with varying velocity. In particular, we call for new experimental work like a transmission line system to study the transient behavior of a KdV system.

(5) In our analysis reported in this paper, we have only treated the situation where the vacuum parameter b is zero, corresponding to $u = 0$ at $x \to \infty$. Our soliton studied here is unidirectional under this special situation. For nonzero b, we anticipate that the soliton can propagate to both directions. Before we understand the $b \neq 0$ case, we must study carefully the simpler case ($b = 0$), which has brought up a new concept in the KdV soliton—a soliton can propagate with varying velocity.

[1] C. Au and P. C. W. Fung, "The Vacuum States of the Korteweg-deVries Equation," Phys. Rev. B **25**, 6460 (1982).

[2] A. D. Wahlquist and F. B. Estabrook, "Prolongation Structure of Nonlinear Evaluation Equations," J. Math. Phys. **16**, 1–7 (1975).

[3] Lu Wen (Loo Win), Yao Qiyuan, Ou Zhi (Au Chi), He Shaohui, Tao Puzhen, and He Shigiang, "A Differential Geometrical Approach for the Solution of KdV Equation," Acta Sci. Naturalium Univ. Sunyatseni **3**, 29–43 (1979) (in Chinese). English translation in American Institute of Physics Document No. PAPS PRBMD-25-6460-22.

[4] Loh Ching-Yuen, Department of Applied Mathematics, University of Western Ontario, Canada, private communication, 1982.

A series of new analytical solutions to the nonlinear equation $y_t + y_{xxx} - 6y^2y_x + 6\lambda y_x = 0$. I

P. C. W. Fung and C. Au
Physics Department, University of Hong Kong, Hong Kong

(Received 30 September 1983; accepted for publication 16 December 1983)

Following the Bäcklund transformation and using the theorems proved by the same authors, previously we obtained sets of new solutions to the KdV equation and the nonlinear equation $y_t + y_{xxx} - 6y^2y_x + 6\lambda y_x = 0$ which transforms into the modified KdV equation when $\lambda = 0$. In this paper we present another new series of solutions to the above nonlinear equation. One of our analytical solutions under a certain special condition (vacuum parameter $b = 0$ and $\lambda = 0$) is found to be identical to the one-soliton solution obtained via the inverse scattering method.

PACS numbers: 03.40.Kf

I. INTRODUCTION

The same authors have employed the Bäcklund transformation[1,2] approach to study the mathematical and physical properties of a number of nonlinear equations. Using our methodology, via Bäcklund transformation we put in a "seed" (a KdV solution) and obtained previously a set of solutions to the KdV equation. In particular, using $u = b$ as our "seed," we have found[2] certain new soliton solutions to the KdV equation

$$u_t + u_{xxx} + 12uu_x = 0, \tag{1}$$

and have discovered that the solutions contain the vacuum parameter b (in the KdV case, b is the asymptotic value of the solution as $x \to \pm\infty$) which has important physical significance; the soliton velocity, amplitude, and width are all functions of b. In the second paper of our series,[3] we have used rather powerful theorems to relate the solutions of the KdV equation and nonlinear equation

$$y_t + y_{xxx} - 6y^2y_x + 6\lambda y_x = 0 \tag{2}$$

which transforms into the modified KdV equation if $\lambda = 0$. The Miura[4] transformation is a special case of one of our theorems. One set of our solutions to Eq. (2) is a kink–antikink solution,[3] which tends to $\pm\sqrt{\lambda - 2b}$ as $x \to \pm\infty$, instead of simply b in the case of KdV equation. It is interesting to note that the solutions so far found for the KdV equation are nontopological,[3,5] while the solutions[3] to Eq. (2) are topological. It would be meaningful to find out whether other solutions to (2) can be nontopological via our Bäcklund transformation; in this paper we shall provide an answer to this equation based on all the solutions obtained so far.

In the third and fourth of our papers,[5,6] using the KdV solution $u = -1/x^2$ as the seed, we have derived other new solitons to both the KdV equation and Eq. (2) using our theorems[3] which bridge these solutions. While these new soliton solutions to the KdV equation are close to the nature of the conventional one-soliton solutions (nontopological), the solutions to Eq. (2) are kink–antikink solutions (topological). We have discovered that the new soliton solutions to (1) and (2) change their velocities, amplitudes, and widths as they travel from remote distances. Such properties lead us to discover that soliton solutions can show "annihilation and creation" phenomena as they propagate.[7] It is worth noting that these two equations [(1) and (2)] are classical wave equations.

In this and the following papers, we shall use another seed, namely,

$$u = (\mu - b) - (\mu - 2b)\left(\frac{Ce^{(\mu-2b)^{1/2}r} - e^{-(\mu-2b)^{1/2}r}}{Ce^{(\mu-2b)^{1/2}r} + e^{-(\mu-2b)^{1/2}r}}\right)^2 \tag{3}$$

[where $r = x - 4(b + \mu)t$ and C, μ are constants and b is the vacuum parameter [2]] to find other new analytical solutions to the KdV equation (1) and Eq. (2).

Since both the KdV and the modified KdV equations have been applied to many physical situations, new solutions to these equations are of both physical and mathematical significance.

II. SEVERAL SETS OF NEW SOLUTIONS TO THE NONLINEAR EQUATION $y_t + y_{xxx} - 6y^2y_x + 6\lambda y_x = 0$, USING BÄCKLUND TRANSFORMATION

According to theorems stated in Ref. 3, if u is a solution to the KdV equation (1) and y satisfies

$$y_x = \lambda - 2u(x,t) - y^2, \tag{4}$$

$$y_t = -4(\lambda + u)y_x + 2u_{xx} - 4u_xy, \tag{5}$$

then y is a solution to (2), and

$$u^* \equiv \lambda - u - y^2 \tag{6}$$

is also a solution to (1).

We have already obtained a one-soliton KdV solution (3) with nonzero vacuum parameter. Now we use this solution (3) as our seed, we find that Eq. (5) can be expressed as

$$y_t + 4(\lambda + b + \mu' - L^2)y_x = -4\mu'^2 + 16\mu'L^2 - 12L^4 + 8(\mu'L - L^3)y, \tag{7}$$

where the real number

$$L \equiv \sqrt{\mu'}\left(\frac{C\exp\{\mu'^{1/2}[x - 4(b+\mu)t]\} - \exp\{-\mu'^{1/2}[x - 4(b+\mu)t]\}}{C\exp\{\mu'^{1/2}[x - 4(b+\mu)t]\} + \exp\{-\mu'^{1/2}[x - 4(b+\mu)t]\}}\right), \tag{8}$$

and

$$\mu' \equiv \mu - 2b.$$

 0022-2488/84/051370-02$02.50

The parameter μ' cannot be equal to zero, otherwise seed (3) will become $u(x,t) = b$, leading to solutions which have already been found.[2]

The characteristic equations of (7) are

$$dt = \frac{dx}{4(\lambda + b + \mu' - L^2)} = \frac{dy}{-4\mu'^2 + 16\mu' L^2 - 12L^4 + 8(\mu' L - L^3)y}. \tag{9}$$

We now solve Eqs. (9) and (4), and obtain the following real solutions:

$$y = -L \quad (\text{for } \lambda' = \mu' > 0), \tag{10}$$

$$y = -L + \frac{\lambda'}{\lambda' - L^2} \frac{1}{L/(2(\lambda' - L)) + (1/(4\lambda'^{1/2}))\ln(|\sqrt{\lambda'} + L|/|\sqrt{\lambda'} - L|) - 4\lambda' t + D} \quad (\text{for } \lambda' = \mu' > 0), \tag{11}$$

$$y = (\mu' - L^2)/L \quad (\text{for } \lambda' = 0, \mu' \neq 0), \tag{12}$$

$$y = \frac{\mu' - L^2}{L} - \frac{\mu'}{L^2} \frac{1}{[1/L - (1/(2\sqrt{\mu'}))\ln(|L + \sqrt{\mu'}|/|L - \sqrt{\mu'}|)] - 4\mu' t + D} \quad (\text{for } \lambda' = 0, \mu' > 0), \tag{13}$$

$$y = \frac{L(L^2 - \mu')}{\lambda' - L^2} \pm \frac{\sqrt{\lambda'}(\lambda' - \mu')}{\lambda' - L^2} \quad (\text{for } \lambda' > 0, \mu' \neq 0), \tag{14}$$

$$y = \frac{L(L^2 - \mu')}{\lambda' - L^2} + \frac{\sqrt{\lambda'}(\lambda' - \mu')}{\lambda' - L^2} \frac{Ce^{\lambda'^{1/2}\xi} - e^{-\lambda'^{1/2}\xi}}{Ce^{\lambda'^{1/2}\xi} + e^{-\lambda'^{1/2}\xi}} \quad (\text{for } \lambda' > 0, \mu' > 0, \text{ and } \lambda' \neq \mu'), \tag{15}$$

where

$$\xi = \frac{1}{2}\left(\frac{1}{\sqrt{\lambda'}} \ln\frac{|\sqrt{\lambda'} - L|}{|\sqrt{\lambda'} + L|} + \frac{1}{\sqrt{\mu'}} \ln\frac{|\sqrt{\mu'} + L|}{|\sqrt{\mu'} - L|}\right) - 4(\lambda' - \mu')t. \tag{16}$$

We have substituted solutions (10) – (15) directly into Eq. (2) and found that Eq. (2) is satisfied.

As a demonstration of the complexity of our new solutions corresponding to different parameters, we list the following two forms of the rather simple solution (12):for $\mu' > 0$,

$$y = \pm 2\sqrt{\mu'}/\sinh\{2\sqrt{\mu'}[x - 4(b + \mu)t]\} \quad (\text{for } C = \pm 1); \tag{17}$$

and for $\mu' < 0$,

$$y = \pm 2\sqrt{|\mu'|}/\sin\{2\sqrt{|\mu'|}[x - 4(b + \mu)t]\} \quad (\text{for } C = \pm 1). \tag{18}$$

We would remark that solution (12) is derived under the condition $\lambda' = \lambda - 2b = 0$. If we set $b = 0$, automatically $\lambda = 0$; then Eq. (2) becomes the modified KdV equation. In this case, for the sake of comparison, solution (18) takes the form

$$y = \pm 2\eta/\sin(2\eta x + 8\eta x^3 t), \tag{19}$$

where $\eta \equiv (|\mu'|)^{1/2}$.

In this article, we have concentrated mainly on new solutions pertaining to the case where the parameter $\mu' > 0$. The other new solutions corresponding to $\mu' < 0$ will be published in the following article.

III. DISCUSSIONS

(1) We believe that if a nonlinear equation can be taken to represent a physical situation, then its solutions represent certain physical phenomena. The new solutions reported here have definite physical significance. We call for further experimental investigations pertaining to the situations represented by Eq. (1) and (2). Along the line of theoretical research, it would be interesting to analyze various properties of our new solutions (10)–(15). We shall follow up this work and report the result in subsequent paper(s).

(2) We would remark that the y solutions (10)–(15) satisfy Eq. (2), and the system (4) and (5). However, if y is given by (10) – (15) then $-y$ also satisfies Eq. (2) (Ref. 6) but not Eq. (4) and (5). Moreover, as we have pointed out,[6] if $y(x,t)$ is a solution to (2), then $y(x + 6\lambda t, t)$ is a solution to the modified KdV equation. Also, since (10)–(15) are solutions to the modified KdV equation if $\lambda = 0$, we denote it as $y_{\lambda=0}$. Then $y_{\lambda'=0}(x - 6\lambda t, t)$ evaluated at $\lambda = 0$ is a solution to Eq. (2). Based on such results, we can obtain rather wide types of solutions to Eq. (2).

(3) The modified KdV solution (19) is identical to the one-soliton solution obtained using the inverse scattering method.[8,9] It is easy to see the stated identity if one transforms Eq. (2) into their format.

(4) At first inspection, one observes that solution (10) is a topological solution, while (16) is nontopological. Solution (17), however, has a periodic boundary condition.

[1] H. D. Walhquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[2] C. Au and P. C. W. Fung, Phys. Rev. B **25**, 6460 (1982).
[3] P. C. W. Fung and C. Au, Phys. Rev. B **26**, 4035 (1982).
[4] R. M. Miura, J. Math. Phys. **9**, 1902 (1968).
[5] C. Au and P. C. W. Fung, "A KdV soliton propagating with varying velocity," J. Math. Phys. **25**, 1364 (1984).
[6] P. C. W. Fung and C. Au, J. Phys. A **16**, 1611 (1983).
[7] C. Au and P. C. W. Fung, "Annihilation and Creation of a KdV Soliton" (submitted to Phys. Rev., 1983).
[8] M. Wadati, J.Phys. Soc. Jpn. **34**, 1289 (1973).
[9] In their work, they have expressed the modified KdV equation as $y_t + y_{xxx} + 6y^2 y_x = 0$. Their solution is $y = \pm 2\eta\,\mathrm{sech}(8\eta^3 t - 2\eta x + \delta)$.

Exact solution of some nonlinear evolution equations

Yukiko Tagami
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 28 October 1983; accepted for publication 2 December 1983)

A method is developed and subsequently used to study periodic solutions of (1) the one-dimensional time-dependent Ginzburg–Landau equation, (2) an analog of the sine–Gordon equation, and (3) a special system of nonlinear equations associated with a problem of interface.

PACS numbers: 03.40.Kf, 02.30.Hq

1. INTRODUCTION

Progress in recent years in the study of nonlinear evolution equations has made significant contributions to our understanding of many physical systems. In this paper we develop a systematic method of solving typical one-dimensional nonlinear evolution equations of broad physical interest. Specifically, we investigate (1) the one-dimensional time-dependent Ginzburg–Landau (TDGL) equation, (2) a dissipative analog of the sine–Gordon equation (SGE), and (3) one-dimensional case of a set of coupled nonlinear diffusion equations derived by Helfand and the author[1] (the HT system) for the analysis of interfacial properties in polymers. Our method is based on the use of a set (triad) of periodic as well as nonperiodic functions associated with the stationary solution of SGE. Extensive use of elliptic functions[2,3] are made throughout the paper.

Section 2 along with the Appendix establishes a solution triad associated with the stationary solution of SGE. Using the solution triad we then calculate solutions for the one-dimensional TDGL equation and an analog of SGE in Secs. 3 and 4, respectively. Investigated in Sec. 5 is the one-dimensional HT system as a practical application of our solution triad. Recapitulation follows in Sec. 6.

2. SOLUTION TRIAD OF THE SINE–GORDON SYSTEM

In this paper we consider a class of one-dimensional nonlinear evolution equations that pertain to the multitudinous ramification of the following equation:

$$\frac{\partial}{\partial t}\phi(y,t)=\frac{\partial^2\phi}{\partial y^2}-\sin\phi, \tag{2.1}$$

where ϕ ($0\leqslant\phi\leqslant\pi/2$) refers to an angle, and y and t stand, respectively, for a spatial coordinate and time. Equation (2.1) is an analog of SGE, and will be discussed in more detail in Sec. 4.

Let us point out here that the solution of the stationary equation (2.1), i.e.,

$$\frac{d^2}{dy^2}\phi^0(y)-\sin\phi^0=0, \tag{2.2}$$

can be written in the form

$$\phi^0(y)=4\arctan\left(\frac{1-h}{1+h}\right)^{1/2}=2\arccos(h), \tag{2.3}$$

where h ($|h|\leqslant 1$) constitutes the following triad, which we shall hereafter refer to as the *h-triad*

$$h(y)=\begin{cases}\mathrm{sn}(ky,1/k), & k>1, \quad (2.4a)\\ \tanh y, & (k=1), \quad (2.4b)\\ k\,\mathrm{sn}(y,k), & 1>k>0. \quad (2.4c)\end{cases}$$

Here $\mathrm{sn}(y,k)$ is the Jacobian elliptic function with modulus k. In Eqs. (2.4a) and (2.4c) the argument y should read y^* which has been shifted from y of Eq. (2.2) by a quarter period, viz.,

$$y^*=y+K(k), \quad k\neq 1. \tag{2.5}$$

The derivation of Eqs. (2.3) and (2.4a)–(2.4c) is found in the Appendix. Since the form (2.4b) may be regarded as the limit, when k goes to 1, of either of the form (2.4a) or form (2.4c), we shall refer to the former case as the one with $k=1$ for convenience, although the modulus k there has become meaningless. Setting the h-triad in the form of Eqs. (2.4a)–(2.4c) is tantamount to fixing the location of the kink at $y=0$, being $h(0)=0$, as well as the kink width which is determined by $k=h'(0)$ (see Sec. 3).

It will be worth noting that the h-triad (2.4a)–(2.4c), upon adjustment of the modulus k, constitutes the solution of the equation of motion of a pendulum: the cases (2.4a), (2.4c), and (2.4b) correspond, respectively, to the case when the pendulum is undergoing revolution, when it is undergoing oscillation, and when it slows down as it moves up to the upside-down position and stops there. The utility of thus-introduced h-triad will be demonstrated in the remaining sections.

3. SOLUTION OF ONE-DIMENSIONAL TDGL EQUATION

One-dimensional TDGL equation for the order parameter in super conductivity[4] may be written as

$$\frac{\partial}{\partial t}u(x,t)=L\left\{\frac{\partial^2}{\partial x^2}u(x,t)+g\left[u(x,t)-\tfrac{1}{6}u(x,t)^3\right]\right\}, \tag{3.1}$$

where L and g are constants, among which g is usually taken to be large. Equation (3.1) is associated with the potential energy

$$V[u(x)]=-\tfrac{1}{2}u^2+(1/4!)u^4, \tag{3.2}$$

and constitutes an analog of the $\lambda\phi^4$ equation. If we assume the traveling-wave solution for Eq. (3.1)

$$u(x,t)=M(x-vt), \tag{3.3}$$

with v the velocity of the traveling wave, then $M(x)$ satisfies the following equation:

$$\frac{d^2M}{dx^2}+\frac{v}{L}\frac{dM}{dx}+g\left(1-\frac{1}{6}M^2\right)M=0, \tag{3.4}$$

the kink solution of which is known[4] as

$$M_K(x)=\pm\sqrt{6}\tanh(\sqrt{(g/2)}\,x). \tag{3.5}$$

Proper sign is determined by a given boundary condition.

 0022-2488/84/051372-06$02.50

Suppose we seek the solution of Eq. (3.4) of the type of a slowly propagating wave. For this purpose we express the solution in the form of expansion in terms of a small parameter $v/L = \epsilon$ ($\epsilon > 0$) as

$$M(x;\epsilon) = M_0(x) + \epsilon M_1(x) + \epsilon^2 M_2(x) + \cdots \tag{3.6}$$

and substitute $M(x;\epsilon)$ into Eq. (3.4).

The equation for $M_0(x)$ is as follows:

$$\frac{d^2M_0}{dx^2} + g\left(1 - \frac{1}{6}M_0^2\right)M_0 = 0. \tag{3.7}$$

With the knowledge that $M_0(x)$ assumes in a particular case the form (3.5), we easily construct the general solution of Eq. (3.7) in terms of the h-triad defined by Eqs. (2.4a)–(2.4c)

$$M_0(x) = \pm\sqrt{12/(1+k^2)}\,h\left(\sqrt{g/(1+k^2)}\,x\right). \tag{3.8}$$

Expression (3.5) is recovered from Eq. (3.8) by letting the modulus k approach 1. Unlike the single-kink solution of the case $k = 1$, the case $k > 1$ and the case $1 > k > 0$ give periodic solutions corresponding to an alternating kink-antikink structure. The equation satisfied by $M_1(x)$ of Eq. (3.6) is the following:

$$\frac{d^2M_1}{dx^2} + g\left(1 - \frac{1}{2}M_0^2\right)M_1 = -\frac{dM_0}{dx}. \tag{3.9}$$

In light of M_0 of Eq. (3.8), we see that $M_1(x)$ is a function of h^2, with $h = h\left(\sqrt{g/(1+k^2)}\,x\right)$. Putting

$$M_1(x) = M_1(h^2), \tag{3.10}$$

therefore and denoting $dM_1/d(h^2)$ by M_1', we rewrite Eq. (3.9) as

$$\begin{aligned} &2h^2(1-h^2)(k^2-h^2)M_1'' \\ &\quad + [3h^4 - 2(1+k^2)h^2 + k^2]M_1' \\ &\quad + ((1+k^2)/2 - 3h^2)M_1 \\ &= \mp(3/g)^{1/2}[(1-h^2)(k^2-h^2)]^{1/2}. \end{aligned} \tag{3.11}$$

The observation that this equation admits a hypergeometric function of h^2 as its stationary solution leads to the following solution:

$$\begin{aligned} M_1(x) = \pm\frac{1}{\sqrt{3g}}\frac{1}{k^2k'^4}\sqrt{(1-h^2)(k^2-h^2)} \\ \times\left\{\left[k'^2u - (1+k^2)E(u) \right.\right. \\ \left. + h\left(\sqrt{\frac{1-h^2}{k^2-h^2}} + k^2\sqrt{\frac{k^2-h^2}{1-h^2}}\right)\right]^2 \\ \left. - k'^2h^2\left(\frac{1}{k^2-h^2} - \frac{k^4}{1-h^2}\right)\right\}, \end{aligned} \tag{3.12}$$

where $u = F(\sin^{-1}(h/k), k)$ and $E(u) = E(\sin^{-1}(h/k), k)$ are the incomplete elliptic integrals of the first and the second kind, respectively, with modulus k. $k' = \sqrt{1-k^2}$ and $h = h(y)$ with

$$y = \sqrt{g/(1+k^2)}\,x. \tag{3.13}$$

Especially when $k = 1$, Eq. (3.12) reduces to

$$M_1(x) = \mp\frac{1}{\sqrt{3g}}\tanh^2\left(\sqrt{\frac{g}{2}}x\right)\left(2 + \cosh^2\left(\sqrt{\frac{g}{2}}x\right)\right), \tag{3.14}$$

while when $h = k\,\mathrm{sn}(y, k)$, $1 > k > 0$, Eq. (3.12) becomes explicitly

$$\begin{aligned} M_1(x) = \pm\frac{1}{\sqrt{3g}}\frac{1}{kk'^4}\,\mathrm{dn}\,y\,\mathrm{cn}\,y\{[k'^2y - (1+k^2)E(y) \\ + \mathrm{sn}\,y(\mathrm{dc}\,y + k^4\,\mathrm{cd}\,y)]^2 - k'^2(\mathrm{tn}^2\,y - k^6\,\mathrm{sd}^2\,y)\}, \end{aligned} \tag{3.15}$$

where $\mathrm{dn}\,y = \mathrm{dn}(y, k)$, $\mathrm{cn}\,y = \mathrm{cn}(y, k)$, and $\mathrm{sd}\,y = \mathrm{sd}(y, k)$ are the Jacobian elliptic functions. Calculation of the functions $M_2(x)$, $M_3(x)$,... in Eq. (3.6) will similarly be facilitated by making use of the h-triad.

Of the prominent features of the thus-obtained traveling-wave solution $M(x;\epsilon) = M_0(x) + \epsilon M_1(x) + \cdots$ let us look at the kink width $\Delta(k)$, which is defined in the usual fashion

$$\Delta(k) = \left[\left|\frac{dM(x)}{dx}\right|_{x=0}\right]^{-1}, \tag{3.16}$$

$x = 0$ being the location of the kink. Noting that $dM_1(x)/dx$ (at $x = 0$) vanishes, we obtain to $O(\epsilon)$

$$\Delta(k)/\Delta(1) = (1+k^2)/2k, \tag{3.17}$$

with $\Delta(1) = 1/\sqrt{3g}$. That the ratio (3.17) increases from 1 as k deviates from 1 in either direction will imply that in the case of the TDGL equation the kink width tends to broaden once the periodicity sets in.

4. SOLUTION OF SGE ANALOG

Consider the following equation [see Eq. (2.1)]:

$$\frac{\partial}{\partial t}u(x,t) = L\left\{\frac{\partial^2}{\partial x^2}u(x,t) - g\sin u\right\}, \tag{4.1}$$

which is the counterpart of Eq. (3.1) in the sense that both equations are derived from the same equation except Eq. (4.1) is now associated with the potential energy

$$V[u(x)] = 1 - \cos u. \tag{4.2}$$

Equation (4.1) has been used in the investigations of Josephson current-voltage characteristics[5] in a large junction limit, and of kink dynamics.[6] As was the case with Eq. (3.1), we assume the solution of Eq. (4.1) to be of the type of a slowly propagating traveling wave expressible in the forms (3.3) and (3.6).

Equation for $M_0(x)$ is the following:

$$\frac{d^2M_0}{dx^2} - g\sin M_0 = 0, \tag{4.3}$$

which is solved in light of Sec. 3 as

$$M_0(x) = 4\arctan\left(\frac{1-h}{1+h}\right)^{1/2}, \tag{4.4}$$

where $h = h(\sqrt{g}\,x)$ stands for our h-triad of Eqs. (2.4a)–(2.4c). In the next step we write down the equation for $M_1(x) = M_1(h^2)$:

$$\frac{d^2M_1}{dx^2} - gM_1\cos M_0 = -\frac{dM_0}{dx}, \tag{4.5}$$

or

$$\begin{aligned} &4h^2(1-h^2)(k^2-h^2)M_1'' \\ &\quad + 2[3h^4 - 2(1+k^2)h^2 + k^2]M_1' \\ &\quad - (2h^2-1)M_1 = (2/\sqrt{g})(k^2-h^2)^{1/2}. \end{aligned} \tag{4.6}$$

Since Eq. (4.6) again admits a hypergeometric function of h^2 as its stationary solution, its solution derives straightforward:

$$M_1(x) = -\frac{1}{\sqrt{g}k^2k'^2}\sqrt{k^2-h^2}\left\{\left[k'^2u - E(u) + h\sqrt{\frac{1-h^2}{k^2-h^2}}\right]^2 - \frac{k'^2h^2}{k^2-h^2}\right\}. \tag{4.7}$$

where u and $E(u)$ are given in Eq. (3.12) and $h = h(y)$ with

$$y = \sqrt{g}x. \tag{4.8}$$

Especially when $k = 1$, Eq. (4.7) becomes

$$M_1(x) = \frac{1}{\sqrt{g}}\tanh^2(\sqrt{g}x)\cosh(\sqrt{g}x), \tag{4.9}$$

while when $h = k\,\mathrm{sn}(y, k)$ Eq. (4.7) becomes explicitly

$$M_1(x) = -\frac{1}{\sqrt{g}kk'^2}\,\mathrm{cn}\,y\{[k'^2y - E(y) + \mathrm{dn}\,y\,\mathrm{tn}\,y]^2 - k'^2\,\mathrm{tn}^2\,y\}. \tag{4.10}$$

Calculation of the functions $M_2(x)$, $M_3(x)$,... will similarly be performed by taking advantage of the h-triad.

As in Sec. 3, evaluation of the kink width $\Delta(k)$ is in order. Since dM_1/dx (at $x = 0$) vanishes, we obtain to $O(\epsilon)$

$$\Delta(k)/\Delta(1) = 1/k, \tag{4.11}$$

with $\Delta(1) = 1/2\sqrt{g}$. The behavior of $\Delta(k)$ for the case of an SGE analog visualized in Eq. (4.11) is clearly distinguishable from the corresponding behavior for the TDGL equation described by Eq. (3.17).

5. SOLUTION OF THE ONE-DIMENSIONAL HT SYSTEM

As another example of the expedience of the h-triad, let us investigate the one-dimensional version of the system of nonlinear partial differential equations derived by Helfand and the author[1] which has proved to be a useful tool in the analysis of properties of interface between immiscible polymers. Since it is not the purpose of the present work to go into detail about the chemical aspects of the system, only a cursory description of the system is provided here.

Consider a mixture of two mutually immiscible polymers A and B. By a polymer A (or B) we imply a chain of Z (usually taken to be a large number) identical beads of A species (B species). We assume for simplicity that a bead of A species and that of B species are physically indistinguishable. Then the properties of A-B interface exhibited by the mixture is characterized by the densities $\rho_A(y)$ and $\rho_B(y)$ of bead A and bead B, respectively, at location y, defined by

$$\frac{\rho_A(y)}{\rho_0} = \tilde{\rho}_A(y) = \int_0^1 dt\, q_A(y, 1-t)q_A(y, t), \tag{5.1}$$

$$\frac{\rho_B(y)}{\rho_0} = \tilde{\rho}_B(y) = \int_0^1 dt\, q_B(y, 1-t)q_B(y, t), \tag{5.2}$$

where ρ_0 is the density of pure A (or B) bead evaluated at locations far away from where the mixing takes place. $q_A(y, t)$ and $q_B(y, t)$ are the solutions of the following system:

$$\frac{1}{\chi Z}\frac{\partial}{\partial t}q_A(y, t) = \frac{\partial^2 q_A}{\partial y^2} - q_B{}^2q_A - \frac{1}{\epsilon}(q_A{}^2 + q_B{}^2 - 1)q_A, \tag{5.3}$$

$$\frac{1}{\chi Z}\frac{\partial}{\partial t}q_B(y, t) = \frac{\partial^2 q_B}{\partial y^2} - q_A{}^2q_B - \frac{1}{\epsilon}(q_A{}^2 + q_B{}^2 - 1)q_B. \tag{5.4}$$

Here χ is a small number related to the strength of interaction between A and B beads, and ϵ ($\epsilon > 0$) is a small parameter indicative of the effect of finite compressibility. Unlike the case of the two equations of the preceding sections, t ($0 < Zt < Z$) above does not stand for time but the location of a bead on a polymer chain expressed by the distance measured along the chain. As the concept of "velocity" and thus the traveling-wave picture in such a 1 + 1-dimensional coordinate system is meaningless, we shall attempt the usual linear stability analysis, instead, to obtain the time dependent solutions of Eqs. (5.1)–(5.4) in the form of linear perturbation near the corresponding stationary solutions.

The symmetry with respect to A and B in Eqs. (5.3) and (5.4) allows the solutions to be written as

$$q_A(y, t) = R(y, t)\sin\theta(y, t), \tag{5.5}$$

$$q_B(y, t) = R(y, t)\cos\theta(y, t), \tag{5.6}$$

with

$$R(y, t;\epsilon) = R_0(y, t) + \epsilon R_1(y, t) + \epsilon^2 R_2(y, t) + \cdots, \tag{5.7}$$

$$\theta(y, t;\epsilon) = \theta_0(y, t) + \epsilon\theta_1(y, t) + \epsilon^2\theta_2(y, t) + \cdots. \tag{5.8}$$

We easily obtain

$$R_0 = 1. \tag{5.9}$$

The equation for $\theta_0(y, t)$ writes as

$$\theta_0'' - \frac{1}{4}\sin(4\theta_0) = \frac{1}{\chi Z}\frac{\partial\theta_0}{\partial t}, \tag{5.10}$$

where θ_0' stands for $\partial\theta_0/\partial y$. Use of the result of Sec. 2 yields the stable solution of Eq. (5.10) as follows:

$$\theta_0(y, t) = \theta_0{}^0(y) + \delta\theta_0(y)\exp(-\lambda t), \tag{5.11}$$

with

$$\theta_0{}^0(y) = \arctan\left(\frac{1-h}{1+h}\right)^{1/2}, \tag{5.12a}$$

$$\delta\theta_0(y) = Ch(y), \tag{5.12b}$$

$$\lambda = k^2\chi Z \tag{5.12c}$$

(C is a constant).

The function $R_1(y, t)$ satisfies the following equation:

$$\theta_0'^2 + \tfrac{1}{4}(1 - \cos(4\theta_0)) + 2R_1 = 0, \tag{5.13}$$

which is solved as

$$R_1(y) = \tfrac{1}{8}(h^2 - (2 + k^2)/3). \tag{5.14}$$

The equation for $\theta_1(y, t)$ is as follows:

$$2R_1'\theta_0' + R_1\theta_0'' + \theta_1'' - \tfrac{3}{4}R_1\sin(4\theta_0) - \theta_1\cos(4\theta_0) = \frac{1}{\chi Z}\left(R_1\frac{\partial\theta_0}{\partial t} + \frac{\partial\theta_1}{\partial t}\right), \tag{5.15}$$

which is solved as

$$\theta_1(y,t) = \theta_1^0(y) + \delta\theta_1(y)\exp(-\lambda t), \tag{5.16}$$

with

$$\theta_1^0(y) = -\frac{\sqrt{1-h^2}}{32k^2}\left\{(7k^2-2)h - 2\sqrt{\frac{k^2-h^2}{1-h^2}}\,[(2k^2-1)E(u)+k'^2u]\right\}, \tag{5.17a}$$

$$\delta\theta_1(y) = -\frac{C}{16}k^2h\left\{\left[E(u)+\frac{\sqrt{(1-h^2)(k^2-h^2)}}{h}\right]^2 + (1+k^2) - \frac{k^2}{h^2} - u^2\frac{E}{K} - 2\log\frac{\Theta(u)}{\Theta(0)}\right\}. \tag{5.17b}$$

Here $K = K(k)$, $E = E(k)$, and $\Theta(u)$ refer, respectively, to the complete elliptic integrals of the first and the second kind, and the Jacobian theta function, all with modulus k. u is that of Eq. (3.12). In Eqs. (5.16) and (5.17b) λ and C are those of Eqs. (5.12c) and (5.12b). Especially when $k = 1$ (the case of single kink solution) Eqs. (5.17a) and (5.17b) reduce, respectively, to

$$\theta_1^0(y;k=1) = -\tfrac{3}{32}h\sqrt{1-h^2} = -\tfrac{3}{32}\tanh y\,\operatorname{sech} y, \tag{5.18a}$$

$$\delta\theta_1(y;k=1) = -\frac{C}{16}h\log(1-h^2) = \frac{C}{8}\tanh y\log(\cosh y). \tag{5.18b}$$

Altogether the density profile function of bead A is given by

$$\tilde\rho_A(y) = \frac{1}{2}(1-h)\left\{1+\frac{\epsilon}{16k^2}\left((5k^2+2)h^2-(7k^2-2)h-4k^2(2+k^2) + 2\sqrt{\frac{(1+h)(k^2-h^2)}{1-h}}\,[(2k^2-1)E(u)+k'^2u]\right)\right\}\left\{1+\frac{2C}{\lambda}(1-e^{-\lambda})h\sqrt{\frac{1+h}{1-h}}\left(1+\frac{\epsilon}{16k^2}\left((7k^2-2)h - 2\sqrt{\frac{k^2-h^2}{1-h^2}}\,[(2k^2-1)E(u)+k'^2u] - k^4\left(\left[E(u)+\frac{\sqrt{(1-h^2)(k^2-h^2)}}{h}\right]^2 + (1+k^2) - \frac{k^2}{h^2} - u^2\frac{E}{K} - 2\log\frac{\Theta(u)}{\Theta(u)}\right)\right)\right)\right\}, \tag{5.19}$$

and that of bead B by

$$\tilde\rho_B(y) = \tilde\rho_A(-y). \tag{5.20}$$

The implication of the results shown above will be evident in the following sketch of the interface picture derived therefrom. The thickness of the interface, or the kink width, is obtained by replacement of $M(x)$ by $\tilde\rho_A(y)$ in Eq. (3.16). We obtain, for the unperturbed term in Eq. (5.19),

$$\Delta(k)/\Delta(1) = 1/k, \tag{5.21}$$

with $\Delta(1) = 2$. The coincidence of Eq. (5.21) with Eq. (4.11) is predictable on account of the similarity between Eqs. (4.4) and (5.12a). Another quantity often used to characterize the strength of kink-antikink interaction is the interface tension γ, which is defined[1] as

$$\gamma(k) = \text{const}\int_{-y_0}^{y_0} dy\,[\tilde\rho_A(y)\tilde\rho_B(y) - \tilde\rho_A(-y_0)\tilde\rho_B(-y_0)] = \text{const}\int_{-y_0}^{y_0} dy[(1-h^2(y)) - (1-h^2(-y_0))], \tag{5.22}$$

where y_0 represents a quarter period of the h-triad (see Appendix). (For the nonperiodic case $k = 1$ we take y_0 to be ∞.) Again using the unperturbed term in Eq. (5.19) we immediately obtain

$$\gamma(k)/\gamma(1) = E(k) - k'^2K(k). \tag{5.23}$$

It is easily seen that the function $E(k) - k'^2K(k)$ varies continuously from 1 to 0 as k decreases from 1 to 0. For the case $k > 1$, since $E(k) - k'^2K(k) = E(\kappa)/\kappa$, with $\kappa = 1/k$, we see that this function continuously increases from 1 as k increases from 1.

In periodically or nonperiodically striated structures obtained upon the mixing of two mutually immiscible polymers, observations appear to point to the tendency that the thickness of the interface becomes broader and the interfacial tension smaller once the periodicity sets in. From this we infer that in such a system only the case $1 \geqslant k > 0$, but not the case $k > 1$, is realized under usual experimental conditions. Stationary solutions of $\tilde\rho_A(y)$ and $\tilde\rho_B(y)$ for the case $k = 1$ have been analyzed in Ref. 1, while periodic solutions for the HT system have not been presented elsewhere.

6. RECAPITULATION

A class of one-dimensional nonlinear evolution equations that is known to have soliton solutions is investigated, and exact periodic solutions are presented explicitly for such equations of physical interest as (1) the one-dimensional TDGL equation, (2) an SGE analog, and (3) an HT system of equations for polymeric interfaces. Our method consists of introduction of the h-triad in the fashion described in the Appendix and identification of the known solutions of the type of hyperbolic tangent function with a member of the thus-introduced h-triad. The periodic solutions calculated in this paper will be of use in the analysis of many a periodic physical systems.

Furthermore, in view of the two different types of behaviors of the kink width calculated for (1) [viz., Eq. (3.17)] and for (2) as well as for (3) together [viz., Eqs. (4.11) and (5.21)],

and in conjunction with the remark made at the end of Sec. 5, we consider it to be often the case that in a physical system only the alternative of the case $k \geqslant 1$ or the case $1 \geqslant k > 0$, but not both, will become amenable to observation. This should correspond to the pendulum motion mentioned in Sec. 2 in which each of the three cases $k > 1$, $k = 1$, $1 > k > 0$ refers to a completely different physical picture from each other.

ACKNOWLEDGMENT

The author wishes to thank Professor Katsuro Sawada for enlightening discussions.

APPENDIX

Equation (2.3) with the h-triad (2.4a)–(2.4c) is derived. Consider the equation

$$\frac{d^2}{dy^2}\phi^0(y) - \sin\phi^0 = 0, \tag{2.2}$$

and assume its solution to be of the form suggested from the solitary wave solution of SGE[7]

$$\phi^0(y) = 4\arctan f(y), \tag{A1}$$

with $f(y) > 0$. Then the function $f(y)$ satisfies the following equation:

$$(1+f^2)f_{yy} - 2ff_y^2 - f(1-f^2) = 0, \tag{A2}$$

where $f_y = df/dy$. It can be easily seen that the solution $f(y)$ of the following equation:

$$f_y^2 = Af^4 + Bf^2 + C \tag{A3}$$

(A, B, and C are constants) also satisfies Eq. (A2), provided

$$B = 2A + 1, \quad C = A. \tag{A4}$$

Hence the solution of Eq. (A2) is given by

$$\int [Af^4 + (2A+1)f^2 + A]^{-1/2}\, df = y + \text{const} \tag{A5}$$

with $A > -1/4$. Equation (A5) is made explicit as follows:

(i) case when $A > 0$.

If we put

$$a^2 = [2A + 1 + \sqrt{4A+1}]/2A,$$

$$b^2 = [2A + 1 - \sqrt{4A+1}]/2A,$$

and

$$k^2 = (a^2 - b^2)/a^2,$$

then Eq. (A5) is solved as

$$f(y) = \sqrt{k'}\,\mathrm{tn}(y/(1-k'), k). \tag{A6}$$

Application of the Landen's transformation to Eq. (A6) finally gives

$$f(y) = \left(\frac{1 - \mathrm{cd}(ky, 1/k)}{1 + \mathrm{cd}(ky, 1/k)}\right)^{1/2}, \quad k > 1. \tag{A7}$$

In Eqs. (A6) and (A7) $\mathrm{tn}(y, k)$ and $\mathrm{cd}(ky, 1/k) = \mathrm{dc}(y, k)$ are the Jacobian elliptic functions, and $k' = \sqrt{1-k^2}$.

(ii) case when $A = 0$.

Equation (A5) is solved as

$$f(y) = e^{-y} = \left(\frac{1 - \tanh y}{1 + \tanh y}\right)^{1/2}. \tag{A8}$$

e^{+y} gives the other solution.

(iii) case when $0 > A > -1/4$.

If we put

$$a^2 = -[2A + 1 + \sqrt{4A+1}]/2A,$$

$$b^2 = -[2A + 1 - \sqrt{4A+1}]/2A,$$

and

$$k^2 = (a^2 - b^2)/a^2,$$

then Eq. (A5) gives

$$f(y) = \sqrt{k'}\,\mathrm{nd}(y/(1+k'), k), \tag{A9}$$

which becomes, upon performing the Landen's transformation,

$$f(y) = \left(\frac{1 - k\,\mathrm{cd}(y, k)}{1 + k\,\mathrm{cd}(y, k)}\right)^{1/2}, \quad 1 > k > 0. \tag{A10}$$

In Eq. (A9) $\mathrm{nd}(y, k)$ is the Jacobian elliptic function. Altogether we have the solution of Eq. (2.2) in the form

$$\phi^0(y) = 4\arctan\left(\frac{1-\tilde{h}}{1+\tilde{h}}\right)^{1/2} = 2\arccos(\tilde{h}), \tag{A11}$$

with $|\tilde{h}| \leqslant 1$, where $\tilde{h}$ stands for the following triad

$$\tilde{h}(y, k) \equiv \begin{cases} \mathrm{cd}(ky, 1/k), & k > 1 \quad \text{(A12a)} \\ \tanh y, & \text{(A12b)} \\ k\,\mathrm{cd}(y, k), & 1 > k > 0. \quad \text{(A12c)} \end{cases}$$

Equation (A12b) corresponds to a single kink solution, while Eqs. (A12a) and (A12c) refer, respectively, to periodic solutions with period $4(1/k)K(1/k)$ and $4K(k)$, where $K(k)$ is the complete elliptic integral of the first kind with modulus k. Neither form (A12a) nor form (A12c) tends, in the limit when k goes to 1, to the form (A12b). However, we may transform the triad $\tilde{h}(y, k)$ into more familiar forms

$$\tilde{h}(y, k) = \begin{cases} \mathrm{sn}(ky^*, 1/k), & k > 1 \quad \text{(A13a)} \\ \tanh y, & \text{(A13b)} \\ k\,\mathrm{sn}(y^*, k), & 1 > k > 0, \quad \text{(A13c)} \end{cases}$$

by shifting the argument by a quarter period, viz.,

$$y^* = y + K(k), \quad k \neq 1. \tag{A14}$$

We then redefine y^* as y to obtain the h-triad

$$h(y, k) \equiv \begin{cases} \mathrm{sn}(ky, 1/k), & k > 1 \quad \text{(2.4a)} \\ \tanh y, & (k = 1) \quad \text{(2.4b)} \\ k\,\mathrm{sn}(y, k), & 1 > k > 0. \quad \text{(2.4c)} \end{cases}$$

Incidentally, the h-triad is associated with the equation of motion of a pendulum, which is Eq. (2.2) except the sign in front of $\sin\phi^0$ is now opposite. The solution of the pendulum equation written in the form of Eq. (A11) is

$$\phi^0(y) = 4\arctan\left(\frac{1-\sqrt{1-h^2}}{1+\sqrt{1-h^2}}\right)^{1/2} = 2\arcsin(h). \tag{A15}$$

Unlike the $\tilde{h}$-triad, in h-triad both the form for $k > 1$ and that for $1 > k > 0$ tend, in the limit when k goes to 1, continuously to the form $\tanh y$. As seen in the main text the h-triad, by way of Eq. (2.3), will ever appear in our problem in the form h^2; accordingly, the choice of two solutions in the case (ii) will in no way affect the consequence of our calculation.

[1]E. Helfand and Y. Tagami, J. Chem. Phys. **56**, 3592 (1972).
[2]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 3rd ed. (Springer, Berlin, 1966).
[3]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).
[4]K. Kawasaki and T. Ohta, Physica A **116**, 573 (1982).
[5]J. R. Waldram, A. B. Pippard, and J. Clarke, Philos. Trans. R. Soc. London **268**, 265 (1970).
[6]M. Büttiker and R. Landauer, Phys. Rev. A **23**, 1397 (1981).
[7]A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proc. IEEE **61**, 1443 (1973).

Backscattering and localization of high-frequency waves in a one-dimensional random medium

Uriel Frisch and Jean-Luc Gautero
C.N.R.S., Observatoire de Nice, B.P. 252, 06007 Nice Cedex, France

(Received 24 May 1983; accepted for publication 2 December 1983)

Total reflection of waves by a one-dimensional random medium is superficially inconsistent with the decoupling of forward and backward propagating waves in the WKB limit (when no real turning points are present). Actually, complex turning points yield exponentially small reflection terms; in the random case, their cumulative effect over large distances can be evaluated using Fokker–Planck techniques, and total reflection is recovered. We also calculate the localization length as a function of wavenumber and strength of the random fluctuations.

PACS numbers: 03.40.Kf, 02.30.Hq

I. INTRODUCTION

We shall be interested in the one-dimensional Helmholtz equation written in the form

$$\frac{d^2\Psi}{dx^2}(x) + k^2 V(x)\Psi(x) = 0, \quad V(x) = n^2(x); \tag{1.1}$$

k is the wavenumber; $n(x) > 0$ is a (real) refraction index assumed to be analytic for real x and statistically homogeneous (translation-invariant). $V(x)$ will be denoted, somewhat improperly, the "potential." It is known that, in one dimension, randomness prevents propagation in the following sense: (i) a wave incident upon a semiinfinite random medium is totally reflected[1,2]; (ii) in the bilaterally infinite case, those solutions of (1.1) satisfying outgoing wave conditions at $\pm\infty$ are "localized," i.e., have their amplitude decreasing for $|x|\to\infty$ asymptotically like $e^{-|x|/\delta}$, where δ is the localization length.[3,4]

Localization is a very well-known concept in solid state physics.[3] It may also play an important role in connection with macroscopic media, e.g., for propagation in very long optical fibers. It is easily shown that the condition for nearly total reflection by a finite random medium is that it should extend over a distance large compared to the localization length.

Localization and total reflection results are established using ergodic theory[4,5]; therefore, in general, the localization length can only be calculated numerically. A more quantitative theory can be made however when the randomness is small, e.g., when

$$n(x) = n_0 + \epsilon\mu(x), \tag{1.2}$$

where μ has zero mean value and $\epsilon\to 0$: it may then be shown that the (complex) reflection coefficient for a medium of length L/ϵ^2 tends, as $\epsilon\downarrow 0$, to a Markov diffusion process; an explicit equation is obtained for the mean reflection and transmission coefficients, and this localization length varies as ϵ^{-2}.[6,7] A number of applications have been worked out along similar lines.[8–10]

We shall here be interested in a different limit: the fluctuations may or may not be weak but the wavenumber k, proportional to the frequency, becomes very large, so that the wavelength is small compared to the scale of fluctuations. In the limit $k\to+\infty$, the solution of (1.1) can be represented locally in the form of a WKB expansion

$$\Psi(x) = A^+\Psi^+(x) + A^-\Psi^-(x), \tag{1.3}$$

$$\Psi^\pm(x) = V^{-1/4}(x)\exp\left\{\pm\left[ik\int_o^x V^{1/2}(y)\,dy + k^{-1}\alpha_1(x) + k^{-2}\alpha_2(x) + \cdots\right\}\right]. \tag{1.4}$$

The trouble with the (traditional) WKB approximation is that it predicts zero reflection of an incoming wave (say, a pure ψ^+ wave), and this to all orders in k^{-1}. Indeed, Ψ^+ and Ψ^- are separately (asymptotic) solutions of the Helmholtz equation (1.1) provided there are no real turning points [zeros for $V(x)$] or singularities.[11] Therefore, we have the paradoxical situation that ergodic theory predicts total reflection and WKB theory predicts no reflection. As in other high frequency problems, such as the intermittency problem, this difficulty can be resolved by going into the complex domain: the analytic "potential" $V(x)$ will usually have complex turning points (and possibly singularities). Such turning points contribute exponentially small (in k) terms to the reflection coefficient, which are not captured by real WKB theory, but are by its complex versions.[12–17] Since the statistics of $V(x)$ are assumed to be translation-invariant, so will be the statistics of the real parts of the turning points. Therefore, there can be cumulative effects of many exponentially small terms over a large distance, yielding a finite contribution, even in the limit $k\to+\infty$. It is our purpose in this paper to evaluate the localization length in the limit $k\to+\infty$ and also to find its dependence on the strength of the randomness.

The essence of the method and the outline of the paper are as follows: Turning points giving the largest contributions are usually those closest to the real axis. Such points are widely spaced. It is therefore possible to calculate their individual contributions using complex WKB theory with a single pair of complex conjugate singular points (Sec. II). When moving along the real axis and upon crossing a pair of randomly located complex conjugate turning points, the ratio of the amplitude of the WKB solution undergoes a random change which is exponentially small for large k. After an exponential rescaling of distances, a Markov diffusion process is obtained for this ratio. The statistics of the reflection coefficients are then calculated using Fokker–Planck techniques (Sec. III). Finally, in Sec. IV, we calculate the localization length and study the dependence on wavenumber and on the strength of randomness.

 0022-2488/84/051378-04$02.50

II. COMPLEX WKB THEORY FOR THE DETERMINISTIC CASE

Ordinary (real) WKB theory is unable to capture exponentially small terms in the wavenumber k. In the complex WKB theory, it is found that exponentially small terms are contributed by the complex turning points and singularities of the "potential" $V(x)$. We shall restrict our attention mostly to the former. It is possible to carry out complex WKB theory with many turning points. This, however, will not be necessary. It will turn out that in the limit $k \to +\infty$ the dominant contributions come from those turning points which are closest to the real axis. Since, eventually, we shall be working in a probabilistic context where there is a distribution of turning points, the most relevant turning points will be widely spaced; it is then sufficient to patch connection formulae relating to individual pairs of complex conjugate turning points. The logics of the bootstrapping procedure is quite the same as in the intermittency calculation for the nonlinear Langevin equation discussed in Ref. 18.

To investigate the deterministic problem, we therefore assume that the "potential" $V(x)$ has a single pair of turning points located at $a \pm ib$ $(b>0)$. With the elementary WKB solutions $\Psi^+(x)$ and $\Psi^-(x)$ defined by (1.4), we assume that the wave function Ψ for large $|x|$ can be represented in the following form:

$$\Psi(x) = A_<^+ \Psi^+(x) + A_<^- \Psi^-(x), \quad x \to -\infty, \tag{2.1a}$$

$$\Psi(x) = A_>^+ \Psi^+(x) + A_>^- \Psi^-(x), \quad x \to +\infty. \tag{2.1b}$$

Complex WKB theory gives connection formulae relating the $A_<$'s to the $A_>$'s, namely, to leading order

$$A_<^+ = A_>^+ + ie^{-2k\gamma}e^{-2ik\phi}, \tag{2.2a}$$

$$A_<^- = A_<^- - ie^{-2k\gamma}e^{2ik\phi}, \tag{2.2b}$$

where

$$\gamma = \mathrm{Im}\int_a^{a+ib} V^{1/2}(z)\,dz, \tag{2.3a}$$

$$\phi = \mathrm{Re}\int_0^{a+ib} V^{1/2}(z)\,dz. \tag{2.3b}$$

In Eq. (2.3b) the contour from 0 to $a+ib$ is along the real axis from 0 to a and then a straight line from a to $a+ib$. The determination of $V^{1/2}$ is the positive square root along the real axis and its analytic continuation along the path from a to $a+ib$. The fact that the origin of coordinates 0 plays a particular role in the connection formulae from $+\infty$ to $-\infty$ is because the origin is also singled out in the definitions of Ψ^+ and Ψ^-.

Here a brief digression is useful. There are various methods for establishing the connection formulae (2.2). The traditional way is to use Stokes lines.[12,13] This approach, however, has been put on a systematic footing only in limited cases so far,[19] and its extension to singularities other than turning points is not obvious. A different, compact and quite general method has been introduced recently by Berry.[16] He uses a decomposition of the form (1.3), but with ψ^+ and ψ^- limited to the first term and with x-dependent coefficients A^+ and A^-. A set of coupled first-order equations are obtained by Berry for A^+ and A^- in which a change of independent variables is made from x to $k\int_0^x V^{1/2}(y)dy$. This system is then solved by a Born series in all order of multiple scattering. Term by term asymptotic expansions are performed on the multiple integrals and the resulting series is resummed to produce the complex WKB formulae.

For our purpose, the calculation of reflection coefficients, it is sufficient to consider the ratio of the amplitude A^+ and A^-. Defining the complex amplitude ratio

$$R = A^-/A^+, \tag{2.4}$$

we have the connection formula

$$R_< = (R_> - ie^{-2k\gamma}e^{2ik\phi})/(1 + iR_> e^{-2k\gamma}e^{-2ik\phi}). \tag{2.5}$$

III. THE STATISTICAL CASE

We now assume that the "potential" $V(x)$ is random, homogeneous, and calculate the reflection coefficient for a finite slab of the medium of length X large compared to the typical separation between successive real parts of turning points. The geometry of the problem is shown in Fig. 1. At the end $(x=0)$ of the slab there is no incoming wave, so that $R=0$. The problem is to find the statistics of $R(X)$ at the beginning $(x=-X)$, where there are both incoming and reflected waves. When x is moved from 0 to $-X$, each time x passes between a pair of complex conjugate turning points, the new value of R is given (to leading order in k) in terms of the preceding one by (2.5), with the parameters ϕ and γ given by Eqs. (2.3) appropriate for the corresponding turning points. Strictly speaking, the connection formulae are valid only from $+\infty$ to $-\infty$ for individual pairs of turning points. Nevertheless, they will still be valid, to leading order, if the typical (real) separation between successive pairs of turning points is large compared to $k^{-1}\gamma^{-1}$.

To characterize the diffusion process, we need to know the change $\delta R = R_< - R_>$ up to second order in the parameter $e^{-2k\gamma}$; this is because, after averaging, all first-order contributions to Fokker–Planck coefficients will vanish. Let us denote this change by $(\delta R)_{1\mathrm{TP}}$ for "one turning point." From Eq. (2.5), we obtain

$$(\delta R)_{1\mathrm{TP}} = -ie^{-2k\gamma}(e^{2ik\phi} + R^2 e^{-2ik\gamma}) - Re^{-4k\phi}(1 + R^2 e^{-4ik\phi}). \tag{3.1}$$

Here R stands for $R_>$.

We now perform a k-dependent rescaling of length:

$$X/l(k) = x_*, \quad R(X) = R^*(x_*), \tag{3.2}$$

where $l(k)$ will be determined later. When x_* is incremented by δx_*, the complex amplitude ratio undergoes a change δR^* given by (3.1), provided there is a pair of turning points in the interval $(x_*, x_* + \delta x_*)$. In order for $R^*(x_*)$ to become a diffusion process in the limit $k \to +\infty$, we must show that

$$\langle \delta R^*/\delta x_* \rangle, \quad \langle (\delta R^*)^2/\delta x_* \rangle, \quad \langle |\delta R^*|^2/\delta x_* \rangle \tag{3.3}$$

have finite limits as $k \to +\infty$, first, and then $\delta x_* \downarrow 0$, and also that higher-order coefficients tend to zero. The first- and second-order limiting coefficients are the Fokker–Planck coefficients.

To evaluate the various moments in (3.3), we need to know the distribution (per unit length) of (complex conjugate

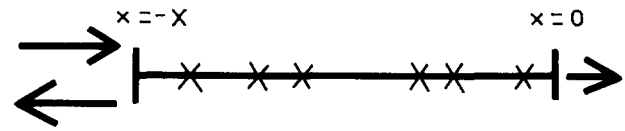


FIG. 1. A finite slab containing many (real parts of) turning points.

pairs of) turning points having a phase factor ϕ (within $d\phi$) and an amplitude factor γ (within $d\gamma$). Actually the phase argument appearing in (3.1) is $k\phi$, so that, for $k\to+\infty$, the phase (defined mod 2π) becomes uniformly distributed and independent of γ; hence all oscillating exponentials that possibly remain in δR^*, in $(\delta R^*)^2$, and in $|\delta R^*|^2$ average to zero.

To perform the remaining averages, we need to know the distribution per unit length of γ which we denote by $M(\gamma)$.

$M(\gamma)\,d\gamma\,\delta X$ = mean number of turning points in an interval δX with damping factor γ with $d\gamma$. (3.4)

In this section, we assume that $M(\gamma)$ is known. It will be shown in the next section how it can be evaluated from a Rice-type formula. Putting $\delta x = l(k)\delta x_*$, we obtain

$$\langle \delta R^*/\delta x_* \rangle = -l(k)\langle\langle e^{-4\gamma k}\rangle\rangle R^*, \tag{3.5a}$$

$$\langle (\delta R^*)^2/\delta x_* \rangle = -2l(k)\langle\langle e^{-4\gamma k}\rangle\rangle R^{*2}, \tag{3.5b}$$

$$\langle |\delta R^*|^2/\delta x_* \rangle = l(k)\langle\langle e^{-4\gamma k}\rangle\rangle(1+|R^*|^4). \tag{3.5c}$$

Here, the double brackets denote an average per unit length

$$\langle\langle e^{-4\gamma k}\rangle\rangle = \int_0^{+\infty} e^{-4\gamma k} M(\gamma)\,d\gamma. \tag{3.6}$$

It is now clear that finite Fokker–Planck coefficients are obtained in the relevant limit by taking

$$l^{-1}(k) = \langle\langle e^{-4\gamma k}\rangle\rangle. \tag{3.7}$$

It is also easily checked that the coefficients of order higher than 2 vanish.

Knowing the Fokker–Planck coefficients, we can write the diffusion equation for the probability distribution of the complex amplitude ratio. This is conveniently done in terms of variables u and θ defined by

$$R = u^{1/2}e^{i\theta}. \tag{3.8}$$

Note that u is the reflection coefficient for the energy. It is a simple matter to calculate the Fokker–Planck coefficients in the u and θ representation in terms of the previously known ones. We thereby obtain the diffusion equation for $\mathscr{P}(x_*, u, \theta)$, the joint probability density of u and θ

$$\frac{\partial}{\partial x_*}\mathscr{P}(x_*;u,\theta) = -\frac{\partial}{\partial u}((1-u)^2\mathscr{P}) + \frac{\partial^2}{\partial u^2}(u(1-u)^2\mathscr{P}) + \frac{1}{4}\frac{\partial^2}{\partial\theta^2}\left(\frac{(1+u)^2}{u}\mathscr{P}\right). \tag{3.9}$$

The "initial" condition is

$$\mathscr{P}(0; u, \theta) = \delta(u), \tag{3.10}$$

which expresses the fact that the complex amplitude ratio is zero when the slab length vanishes.

In the sequel, we shall be interested only in the marginal distribution of u, the energy reflection coefficient

$$P(x_*; u) = \int_0^{2\pi} \mathscr{P}(x_*; u, \theta)\,d\theta. \tag{3.11}$$

Integrating (3.9) over θ, we obtain an equation for P:

$$\frac{\partial}{\partial x_*}P(x_*; u) = -\frac{\partial}{\partial u}((1-u)^2P) + \frac{\partial^2}{\partial u^2}(u(1-u)^2P), \tag{3.12a}$$

$$P(0; u) = \delta(u). \tag{3.12b}$$

In a somewhat more compact form, the diffusion equation may be written

$$\frac{\partial P}{\partial x_*} = \frac{\partial}{\partial u}\left[u\frac{\partial}{\partial u}(1-u)^2P\right]. \tag{3.13}$$

We now proceed to solve this equation. We introduce the Laplace transform

$$\tilde{P}(p; u) = \int_0^{\infty} e^{-px_*}P(x_*; u)\,dx_*. \tag{3.14}$$

$\tilde{P}$ satisfies the following equation:

$$p\tilde{P}(p; u) - \delta(u) = \frac{\partial}{\partial u}\left[u\frac{\partial}{\partial u}(1-u)^2\tilde{P}(p; u)\right]. \tag{3.15}$$

This equation may be solved by a moment method. We note that the support of $\tilde{P}$ is in the interval [0, 1]; this merely expresses that the energy reflection coefficient is nonnegative and cannot exceed unity. We define the moments

$$c_n(p) = \int_0^1 u^n\tilde{P}(p; u)\,du, \quad n = 0, 1, \ldots. \tag{3.16}$$

Simple manipulations on (3.15) lead to the following recursion relations:

$$c_0(p) = 1/p, \tag{3.17a}$$

$$pc_n(p) = n^2[c_{n-1}(p) - 2c_n(p) + c_{n+1}(p)], \quad n \geqslant 1. \tag{3.17b}$$

From this, we can construct a continued fraction representation for $c_1(p)$, the Laplace transform of the mean energy reflection coefficient

$$c_1(p) = \cfrac{1}{p(p+2)} - \cfrac{1^2\cdot 2^2}{(p+2\cdot 1^2)(p+2\cdot 2^2)} - \cfrac{2^2\cdot 3^2}{(p+2\cdot 2^2)(p+2\cdot 3^2)} - \cdots. \tag{3.18}$$

This continuous fraction has one obvious singularity at $p = 0$ which just reflects the fact that the mean reflection coefficient tends towards a nonzero value (namely, unity) as $x_*\to\infty$. It may be shown that its other nearest singularity is at $p_* = -1/4$. As a consequence, the mean energy reflection coefficient behaves as follows for large x:

$$1 - \langle R^*(x_*)^2\rangle \sim \exp(-x_*/4) \quad \text{for } x_*\to+\infty \tag{3.19}$$

(with possible algebraic prefactors). The same exponential factor is obtained for all the higher moments of the energy reflection coefficient. We have thus shown that the WKB limit is not inconsistent with total reflection when the slab length increases.

IV. THE LOCALIZATION LENGTH

Putting together various pieces from the previous section, we find that the mean energy reflection coefficient, in the original variable X, has the following exponential behavior:

$$1 - \langle |R(X)|^2\rangle \sim \exp[-X/(4l(k))] \quad \text{for } X\to+\infty, \tag{4.1}$$

with

$$l^{-1}(k) = \int_0^{+\infty} e^{-4\gamma k}M(\gamma)\,d\gamma, \tag{4.2}$$

where $M(\gamma)$, the distribution of damping lengths, is defined by (3.4).

Hence, the localization length is given by

$$\delta(x) = 4l(k). \tag{4.3}$$

We recall that the damping length γ is expressible as an integral from the real axis up to the turning point at $a + ib$

$$\gamma = \mathrm{Im} \int_a^{a+ib} V^{1/2}(z)\, dz. \tag{4.4}$$

It appears not possible to go much further in calculating the localization length without making some specific assumption about the random "potential" $V(x)$. Henceforth, we shall be interested in the case where the fluctuations are *weak*. Gaussian fluctuations of the form $V(x) = 1 + \epsilon m(x)$ [with Gaussian $m(x)$] are not consistent with the assumed positivity of $V(x)$ everywhere on the real axis.

A simple model, consistent with positivity, is

$$V(x) = 1 + \epsilon m^2(x), \quad \epsilon \text{ small positive}, \tag{4.5}$$

where $m(x)$ is a zero-mean-value Gaussian process with analytic correlation function. We assume that the fluctuations are around unity, since a different choice can be reduced to the present one by rescaling the wavenumber k. We also normalize ϵ and x in such a way that $\langle m^2 \rangle = \langle m'^2 \rangle = 1$, which means, in particular, that we take as unit length a typical scale of the random fluctuations.

It is now easy to locate the most relevant turning points, i.e., those giving the smallest γ. They are found to be the turning points closest to the real axis. The latter may be obtained simply by Taylor-expanding near the real axis up to second order

$$m(x + iy) = m(x) + iym'(x) - \tfrac{1}{2} y^2 m''(x) + \cdots. \tag{4.6}$$

Substituting into (4.5) and demanding that the "potential" vanishes, we obtain, to leading order,

$$m(x) = 0, \quad y^2 m'^2(x) \simeq \frac{1}{\epsilon}, \quad \gamma \simeq \frac{\pi y}{4} \simeq \frac{\pi}{4} \frac{\epsilon^{-1/2}}{|m'(x)|}. \tag{4.7}$$

So the most relevant turning points are located near the real zeros of $m(x)$ having a very large derivative. As a consequence, the high-k behavior of the localization length is governed by the small-γ behavior of $M(\gamma)$, which in turn depends on the far tail of the conditional distribution of the absolute value of $m'(x)$ at those points where m vanishes. The situation here is very similar to what is discussed in Sec. II of Ref. 18 for the nonlinear Langevin equation. We shall therefore skip some redundant details. We denote by $p(0) = (2\pi)^{-1/2}$ the probability density of m at $m = 0$ and by $Q(w) = 2(2\pi)^{-1/2} e^{-w^2/2}$ the conditional probability density of m' at those points where $m = 0$. The localization length, as given by (4.2) and (4.3) is reexpressed, asymptotically for $k \to +\infty$ (at fixed ϵ), as

$$\delta^{-1}(k) \sim 4p(0) \int_0^{+\infty} \exp(-\pi\epsilon^{-1/2} k/w) w Q(w)\, dw. \tag{4.8}$$

The integral on the right-hand side is now evaluated asymptotically for $k \to +\infty$ by Laplace's method. We thereby obtain

$$\delta^{-1}(k) \sim 2^{5/2} z^{-1/2} \pi^{-1/6} k^{1/3} \epsilon^{-1/6} \exp(-\tfrac{3}{2}\pi^{2/3} k^{2/3} \epsilon^{-1/3}). \tag{4.9}$$

Note that the precise functional form depends here on the Gaussian assumption made for $m(x)$. To ensure consistency of the present analysis and of that of Sec. III, it is necessary that the localization length be large compared to the mean spacing between relevant zeros. We found that the latter behaves for $k \to +\infty$ as $\exp(\pi^{2/3} k^{2/3} \epsilon^{-1/3})$, which is indeed smaller than $\delta(k)$ given by (4.9).

We have thus found that: (i) the localization length is exponentially large with the wavenumber (but involves a fractional power of k); (ii) it is also exponentially large with $1/\epsilon$. This is in contrast to the limit studied in Refs. 6 and 7 where it is found that when k is fixed and $\epsilon \downarrow 0$, the localization length goes like $1/\epsilon^2$.

Finally, we wish to stress that in more than one dimension, backscattering from a random medium in the WKB limit may proceed by a very different mechanism, involving just transport along the rays obtained from a straightforward geometrical optics approximation. The rays are governed by a stochastic ODE[20]; their direction can angularly diffuse. In the limit of weak fluctuations, the direction of the rays may undergo a random walk, allowing appreciable backscattering over a characteristic reflection distance which is likely to vary proportionally to an inverse power of the fluctuation strength ϵ. It is not at all obvious that, in more than one dimension, the characteristic reflection distance will be the same as the localization distance; indeed, localization involves subtle interference phenomena which are not captured by geometrical optics.

[1] P. L. Sulem and U. Frisch, J. Plasma Phys. **8**, 217 (1972); P. L. Sulem, Physica **70**, 190 (1973).
[2] U. Frisch, C. Froeschlé, J. P. Scheidecker, and P. L. Sulem, Phys. Rev. A **8**, 1416 (1973).
[3] D. J. Thouless, "Percolation and localization," in *Ill Condensed Matter*, edited by R. Balian, R. Maynard, and G. Toulouse (North-Holland, Amsterdam, 1979), p. 1.
[4] F. Delyon, H. Kunz, and B. Souillard, J. Phys. A **16**, 25 (1983).
[5] V. Oseledec, Trans. Moscow Math. Soc. **19**, 197 (1968); D. Ruelle, "Ergodic Theory of Differential Dynamical Systems," Publ. Math. IHES **50**, 275 (1979).
[6] G. C. Papanicolaou, SIAM J. Appl. Math. **21**, 3 (1971).
[7] G. C. Papanicolaou and J. B. Keller, SIAM J. Appl. Math. **21**, 287 (1971).
[8] W. Kohler and G. C. Papanicolaou, J. Math. Phys. **14**, 1733 (1973).
[9] W. Kohler and G. C. Papanicolaou, J. Math. Phys. **15**, 2186 (1974).
[10] J. B. Keller, G. C. Papanicolaou, and J. Weilenmann, Comm. Pure Appl. Math. **26**, 497 (1973).
[11] C. Bender and S. Orszag, *Advanced Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[12] J. Heading, *An Introduction to Phase Integral Methods* (Methuen, New York, 1962).
[13] R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, New York, 1973).
[14] M. V. Berry and K. Mount, Rep. Prog. Phys. **35**, 315 (1972).
[15] N. Fröman and P. Fröman, *J. W. K. B. Approximation: Contribution to the Theory* (North-Holland, Amsterdam, 1965).
[16] M. V. Berry, "Semiclassically weak reflections above analytic and non-analytic potential barriers," to be published.
[17] J. L. Gautero and U. Frisch, C. R. Acad. Sci. **247**, 48 (1983).
[18] U. Frisch and R. Morf, Phys. Rev. A **23**, 2673 (1981).
[19] A. Voros, "The return of the quartic oscillator," Ann. Inst. H. Poincaré, Sér. A (1983), to be published.
[20] U. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).

Multifrequency inverse problem for the reduced wave equation with sparse data

V. H. Weston
Division of Math Science, Purdue University, West Lafayette, Indiana 47907

(Received 20 July 1983; accepted for publication 16 December 1983)

The inverse problem for the reduced wave equation $\Delta u + k^2 n^2(x)u = 0$ is considered where the quantity to be determined is the value of $n(x)$ in a compact domain D in R^3. The data consists of a finite set of measurements of the scattered field produced by different incident fields. The measurements are made at various points exterior to D and at possibly different frequencies. The mathematical problem involves solving a system of nonlinear functional equations. Conditions are developed which indicate when the measured body is close to a particular comparison body (a linearized perturbation is valid). A new higher order nonlinear iterative procedure is developed for the full nonlinear problem, which reduces to the usual solution in the linearized region. The method is illustrated by computational results for the one-dimensional case.

PACS numbers: 03.40.Kf, 42.20. − y

I. INTRODUCTION

The multifrequency inverse problem associated with the reduced wave equation

$$\Delta u + k^2 n^2(x)u = 0, \quad x \in R^3, \tag{1}$$

is considered here for the case of sparse data and an uncontrolled environment. In such an environment all that can be specified about the scattering object is that (i) it is located in a prescribed bounded region D and (ii) it is characterized by a real index of refraction $n(x)$. The region D has piecewise smooth (C^1) boundary. The index of refraction is sectionally continuous, i.e., it is continuous everywhere apart from a finite set of piecewise smooth surfaces across which it has finite jump discontinuities. The host medium exterior to D will have $n(x) \equiv 1$.

The data consists of a finite set of measured values of the scattered field at different locations external to the scatterer, and at different frequencies. Since the data set is sparse, the solution to the inverse problem is not unique. Additional constraints would have to be imposed to get uniqueness (such constraints can lead to more complicated and less tractable problems). In this connection the fixed frequency inverse problem with sparse data was analyzed where the imposed constraints consisted of finding the solution closest to a given object.[1,2] However, in this paper we will concentrate on generating a solution to the full nonlinear problem without imposing any artificial *a priori* constraints. It will be shown in a subsequent paper that the method developed here contains an intrinsic natural condition of stability, which makes it superior to present nonlinear techniques.

A brief outline of the paper is as follows: In Sec. II, the direct scattering problem is formulated in terms of the (Lipmann–Schwinger) integral equation. From this the inverse problem is expressed in terms of a system of nonlinear functional equations. In Sec. III, the linearized version of the inverse problem is discussed where the measured body is a small perturbation of a known (comparison) body. A nonlinear algorithm is developed for the full nonlinear inverse problem in Sec. IV. Here it is pointed out that the method has an additional advantage over previous methods in that the regularization process that must be used (for inverting almost singular matrices) is much simpler involving only one constraint (related to the error in the matrix). This is followed up in Sec. V, by a brief discussion of computational results for the one-dimensional case. To simplify analysis and give more clarity, a lot of the mathematical details are relegated to the three appendices.

Some of the notation that will be used throughout the paper is as follows: If $f(x)$ and $g(x)$ are real $\mathscr{L}_2(D)$ functions, then the corresponding inner product and norm are

$$(f,g) = \int_D f(x)g(x)\,dx,$$

$$\|f\| = \left(\int_D f(x)^2\,dx\right)^{1/2},$$

whereas if η is a column vector in R^{2N}, its norm will be given by

$$\|\eta\|_2 = (\eta_1^2 + \eta_2^2 + \cdots + \eta_{2N}^2)^{1/2}$$

and a square $2N$ matrix $H = \{H_{ij}\}$ will have corresponding norm denoted by $\|H\|_2$.

II. BASIC MATHEMATICAL PROBLEM

To simplify analysis, set

$$v(x) = n^2(x) - 1 \tag{2}$$

and note that $v(x) \equiv 0$ outside D. Let v_* and v_m denote the value of v corresponding to the comparison and measured body, respectively.

The data set will then consist of the set of N complex numbers $u_m^s(x_l, k_l)$, which is the measured value of the scattered field at a point x_l outside D The scattered field associated with each measurement is generated by a particular incident field $u^i(x, k_l)$ impinging upon the obstacle ($v = v_m$) at a frequency ω_l (corresponding to the wave number k_l).

Let $G_*(x, y; k_l)$ denote the Green's function for the reduced wave equation associated with the comparison body ($n = n_*$) and which satisfies the radiation condition. The total field $u(x, k_l)$ (where $u = u^i + u^s$) generated by the incident wave $u^i(x, k_l)$ impinging upon an arbitrary scatterer v (with support in D) will satisfy the integral equation.

 0022-2488/84/051382-09$02.50

$$u(x,k_l) = u_*(x,k_l) + k_l^2 \int_D G_*(x,y;k_l) \times [v(y) - v_*(y)]u(y,k_l)\,dy. \quad (3)$$

Here $u_*(x, k_l)$ is the total field produced by the same incident wave $u^i(x, k_l)$ impinging upon the comparison body $(v = v_*)$.

The scattered field at a point x_l exterior to D is then given by

$$u^s(x_l,k_l) = u_*^s(x_l,k_l) + k_l^2 \int_D G_*(x_l,y;k_l) \times [v(y) - v_*(y)]u(y,k_l)\,dy, \quad (4)$$

where the term in the integrand $u(y, k_l)$ is obtained by solving integral equation (3). The direct scattering problem is given by Eqs. (3) and (4).

The inverse problem consists of finding the value of $v(x)$ such that

$$u^s(x_l,k_l) = u_m^s(x_l,k_l), \quad l = 1,2,...,N, \quad (5)$$

where the left-hand side of system (5) is given by Eq. (4) and the right-hand side are measured values. Since the solution $u(y, k_l)$ of Eq. (3) is a function of $v - v_*$, expression (4) is a nonlinear function of $v - v_*$. The basic mathematical problem is thus reduced to solving a system of nonlinear complex functional equations for the real variable $v(x)$.

The usual nonlinear techniques like Newton's method cannot be applied to the nonlinear system because the Fréchet derivative of the nonlinear function does not have an inverse due to nonuniqueness.

III. LINEARIZED PROBLEM

If the actual measured body is sufficiently close to the comparison body so that v_m lies in the set v such that

$$\max_{\substack{k_l \\ l=1,2,...,N}} k_l^4 \iint_{D\times D} |G_*^2(x,y;k_l)|(v - v_*)^2\,dx\,dy \ll 1, \quad (6)$$

then the solution of integral equation (3) could be approximated by

$$u(x,k_l) \sim u_*(x,k_l),$$

and system (5) in turn approximated by the linear system

$$b_l = k_l^2 \int_D G_*(x_l,y;k_l)u_*(y,k_l)w(y)\,dy, \quad l = 1,2,...,N, \quad (7)$$

with

$$w = v - v_*, \quad (8)$$

where the complex number

$$b_l = u_m^s(x_l,k_l) - u_*^s(x_l,k_l) \quad (9)$$

is the difference between the measured value of the scattered field and the calculated value corresponding to v_*, at the point x_l and wave number k_l.

The system of N linear complex equations can be reduced to a system of $2N$ real equations. The complex quantities involved in the system are decomposed as follows:

$$k_l^2 G_*(x_l,y;k_l)u_*(y,k_l) = H_l(y) + iH_{l+N}(y), \quad (10)$$

$$b_l = B_l + iB_{l+N}. \quad (11)$$

System (7) can then be reduced to the system of $2N$ real equations

$$(H_l,w) = B_l, \quad l = 1,2,...,2N. \quad (12)$$

With x_l exterior to $\overline{D}$, both $G_*(x_l,y;k_l)$ and $u_*(y,k_l)$, and hence $H_l(y)$, will be continuous functions of y in $\overline{D}$, when v_* is a sectionally continuous function (see Appendix A).

If the measured body v_m is in the linearized region of the comparison body v_* [i.e., condition (6) holds for $v = v_m$], then the system of $2N$ real linear equations (12) provides a useful model for solution. The system as it stands is ill-posed, with a nonunique solution, and any resulting matrix that may be used to invert the system may be ill-conditioned. Moreover, due to errors in data, one would not want an exact solution of system (12), but a solution that satisfies the equation to within the errors of the data (errors in B_l). The usual techniques of regularization may be employed. In particular, the technique of Backus and Gilbert[3] is most suitable.

IV. NONLINEAR PROCESS

A. Conceptual approach

If one has a catalog of comparison bodies, then linear system (12) can be solved for each value of v_* belonging to a catalog of prescribed comparison bodies $\{v_*\}$, obtaining a solution w which is a function of v_*. Then the comparison body that is closest to the measured body is selected. If this comparison body is in the linearized region of the measured body, then the technique mentioned in the previous section or others may be employed to get the optimum solution. There are a number of criteria for closeness than can be used. The usual approach is to employ the following quantity:

$$\phi(v_*) = \sum_{l=1}^{N} |u_m^s(x_l,k_l) - u_*^s(x_l,k_l)|^2,$$

which is the sum of the squares of the differences in the measured and calculated scattered fields. This quantity by itself will not indicate when the linearization process is valid, but is useful to eliminate possible choices of comparison bodies.

The best choice from the theoretical viewpoint is to select the comparison value v_* such that

$$\min_{\substack{v_* \\ l=1,2,...,N}} \max_{k_l} k_l^4 \iint_{D\times D} |G_*^2(x,y;k_l)|w^2(v_*)\,dx\,dy. \quad (13)$$

If the minimum value is much less than unity, then the linearization process employing the value of v_* yielding the minimum would be valid. A better choice from the practical computational standpoint would be to choose v_* so that

$$\min_{v_*} \int_D w^2(v_*)\,dx. \quad (14)$$

This latter choice provides an obvious answer of nonuniqueness of linear system (12). Condition (14) immediately infers that one would want the least squares solution of system (12). The main difficulty in using condition (14) is that it will not signify directly whether the linearization process is valid. As it stands, it can be only used for comparison. However, an additional condition to go along with (14) will be given.

If either one does not have a catalog of comparison bodies to begin with, or else the comparison body belonging to

a catalog is not in the linear region of the measured body, a nonlinear iterative scheme will be used based upon the above concepts. The idea is to obtain a sequence $\{v_*^n\}$ of comparison bodies which minimizes the function in expression (14). Note that when $v_* = v_m$, then the absolute minimum is obtained.

B. Linear inversion

The iterative procedure (which will be developed in the next section), at each step, will require the inversion of a linear system corresponding to Eq. (12), with the solution subject to a condition of form (14). In order to get such a least squares solution [in the $\mathscr{L}_2(D)$ sense] of the linear system a fundamental assumption has to be made.

Assumption: The data points $\{x_l\}$ and wave numbers $\{k_l\}$ are chosen so that $\{H_l(x)\}_{l=1}^{2N}$ is a linearly independent set.

Since the problem of interest here is the sparse data case, this means that, in any numerical application, the number $2N$ will be much less than the number M, the dimension of the finite dimensional subspace used to approximate $\mathscr{L}_2(D)$, and any redundant set of measurements which violates the assumption will be culled out.

It will be shown that, due to the nature of the iteration process, the assumption must be checked out only at the initial step of the iteration process.

With the above asumption, the matrix $\{H_{ij}\}$ whose elements are given by

$$H_{ij} = (H_i, H_j) \tag{15}$$

will be positive definite, since then the quadratic form

$$\sum_{i,j=1}^{2N} c_i H_{ij} c_j = \int_D \left(\sum_{i=1}^{2N} c_i H_i(x) \right)^2 dx$$

will vanish only if $\Sigma_{i=1}^{2N} c_i H_i(x) \equiv 0$. Hence $\{H_{ij}\}$ will be invertible. Let $\tilde{H}_{ij}$ represent the elements of the inverse matrix,

$$\sum_{j=1}^{2N} H_{ij} \tilde{H}_{jk} = \delta_{ik}. \tag{16}$$

Set

$$\xi_i(x) = \sum_{j=1}^{2N} \tilde{H}_{ij} H_j(x); \tag{17}$$

then it can be shown that $\xi_j(x)$ and $H_j(x), j = 1,2,...,2N$, form a biorthogonal set

$$(\xi_i, H_j) = \delta_{ij}. \tag{18}$$

The least squares solution of system (12) is then given by

$$w(x) = \sum_{j=1}^{2N} B_j \xi_j(x) = \sum_{j=1}^{2N} \eta_j H_j(x) \tag{19}$$

where

$$\eta_j = (\xi_j, w) = \sum_{k=1}^{2N} \tilde{H}_{jk} B_k. \tag{20}$$

Note that $w(x)$ is a continuous function of x over the domain $\overline{D}$.

C. Iteration procedure

An iterative procedure based upon a descent type process applied to the nonlinear functional $f(v_*)$, where

$$f(v_*) = \int_D w^2(v_*)\, dx \tag{21}$$

will be used. A minimizing sequence $\{v_*^n\}$ will be sought such that

$$f(v_*^{n+1}) < f(v_*^n). \tag{22}$$

This will require the Taylor expansion

$$f(v_* + \delta v_*) = f(v_*) + (F(v_*), \delta v_*) + \tfrac{1}{2}(\mathfrak{F}'(v_*)\delta v_*, \delta v_*) + \cdots,$$

where $F(v_*)$ is the gradient of $f(v_*)$ and $\mathfrak{F}'$ the Fréchet derivative of F

$$\delta F = \mathfrak{F}'\, \delta v = (F'(v_*), \delta v_*). \tag{23}$$

Here $F'(v_*)$ denotes the kernel of the integral operator $\mathfrak{F}'$.

Explicit expressions for these quantities can be obtained using expansion (19) for the solution $w(v_*)$.

First note that $f(v_*)$ can be represented by the following quadratic form:

$$f(v_*) = \sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j. \tag{24}$$

Using the result given in Appendix B that the Fréchet differential of B_j is given by

$$\delta B_j = -(H_j, \delta v_*),$$

it follows that

$$\delta f = -2 \sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} (H_j, \delta v_*) + \sum_{i,j=1}^{2N} B_i (\delta \tilde{H}_{ij}) B_j \tag{25a}$$

$$= -2(w \delta v_*) - 2(q, \delta v_*), \tag{25b}$$

where, from Appendix C, it is shown that

$$q(x) = \sum_{p=1}^{2N} \eta_p \int_D H'_p(x,z) w(z)\, dz. \tag{26}$$

Here $H'_p(x,z)$ is defined as follows:

$$\delta H_p = \int_D H'_p(x,z) \delta v_*(z)\, dz \tag{27}$$

and its precise form is given by Eq. (B4). Thus the gradient F is given by

$$-\tfrac{1}{2} F = w(x) + q(x), \tag{28}$$

which will be, at least, a continuous function over $\overline{D}$. From Eq. (25) the second Fréchet differential of $f(v_*)$ is given by

$$\delta^2 f = 2 \sum_{i,j=1}^{2N} (H_i, \delta v_*) \delta \tilde{H}_{ij} (H_j, \delta v_*) - 4 \sum_{i,j=1}^{2N} (H_i, \delta v_*) \delta \tilde{H}_{ij} B_j - 2 \sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} (\delta H_j, \delta v_*) + \sum_{i,j=1}^{2N} B_i (\delta^2 \tilde{H}_{ij}) B_j. \tag{29}$$

The explicit form for $F'(x,y)$ may be derived (see Appendix C)

$$\tfrac{1}{2} F'(x,y) = \sum_{l,p=1}^{2N} \left\{ R_l(x) \tilde{H}_{lp} R_p(y) - \eta_p \eta_l \int_D H'_p(x,z) H'_l(y,z)\, dz \right\}$$

$$-\sum_{p=1}^{2N}\eta_p\left\{H'_p(x,y)+\int_D H''_p(x,y,z)w(z)\,dz\right\}, \tag{30}$$

where

$$R_p(x) = H_p(x)+\int_D\left\{H'_p(x,z)w(z)+\sum_{m=1}^{2N}\eta_m H'_m(x,z)H_p(z)\right\}dz. \tag{31}$$

Here the differential of H' is given by

$$\delta H'(x,y)=\int_D H''(x,y,z)\,\delta v_*(z)\,dz, \tag{32}$$

and its precise form is given by Eqs. (B5) and (B6) in Appendix B.

From Eq. (30) it is seen that the operator $\mathfrak{F}'$ has the following decomposition:

$$\mathfrak{F}'=\mathfrak{R}-\mathfrak{B}-\tfrac{1}{2}\mathfrak{P}^2-\mathfrak{C}, \tag{33}$$

where $\mathfrak{R}$ is a positive operator with kernel

$$a(x,y)=2\sum_{l,p=1}^{2N}R_l(x)\tilde{H}_{lp}R_p(y) \tag{34a}$$

and $\mathfrak{P}$ and $\mathfrak{C}$ are self-adjoint integral operators with respective kernels

$$b(x,y)=2\sum_{p=1}^{2N}\eta_p H'_p(x,y), \tag{34b}$$

$$c(x,y)=2\sum_{p=1}^{2N}\eta_p\int_D H''_p(x,y,z)w(z)\,dz. \tag{34c}$$

For further analysis it is convenient to define

$$\beta=(\mathfrak{R}F,F)/(F,F), \tag{35}$$

$$\gamma=-((\mathfrak{P}+\tfrac{1}{2}\mathfrak{P}^2+\mathfrak{C})F,F)/(F,F), \tag{36}$$

where F is the gradient (28). It should be noted that

$$\gamma\leqslant\tfrac{1}{2}-(\mathfrak{C}F,F)/(F,F).$$

Some important properties of F and $\mathfrak{F}'$ are summarized as follows.

Lemma: In the neighborhood of the absolute minimum point of f, the gradient has the behavior

$$-\tfrac{1}{2}F=w+O(\|w\|^2), \tag{37}$$

$$\|q\|/\|w\|\sim O(\|w\|). \tag{38}$$

Proof: From Eqs. (20), (21), and (24), it follows that

$$\|w\|^2=\sum_{i,j=1}^{2N}B_i\tilde{H}_{ij}B_j=\sum_{i,j=1}^{2N}\eta_i H_{ij}\eta_j;$$

hence it may be deduced that

$$\|\boldsymbol{\eta}\|_2\leqslant\|w\|/\lambda_{\min}^{1/2}, \tag{39}$$

$$\|\mathbf{B}\|_2\leqslant\|w\|/\lambda_{\max}^{1/2}, \tag{40}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of the positive definite matrix H. From Eq. (26) it is immediately seen that

$$\|q\|\sim O(\|w\|^2),$$

and the results for F follow from Eq. (28).

Theorem: $\mathfrak{F}'$ is a compact operator with polar kernel given by Eq. (30) which maps $\mathscr{L}_2(D)$ into $C(D)$. It has a decomposition of the form

$$\mathfrak{F}'=\mathfrak{R}-\mathfrak{B}-\tfrac{1}{2}\mathfrak{B}^2-\mathfrak{C}$$

where $\mathfrak{R}$ is a positive operator. In the neighborhood of the point $w\equiv 0$, the behavior of $\mathfrak{F}'$ is given by its kernel

$$F'(x,y)=2\sum_{i,j=1}^{2N}H_i(x)\tilde{H}_{ij}H_j(y)+O(\|w\|), \tag{41}$$

and in particular

$$(\mathfrak{F}'F,F)/(F,F)\sim 2+O(\|w\|). \tag{42}$$

Proof: It follows from Eq. (C 12) in Appendix C that $\mathfrak{F}'$ has a polar kernel, and from Eqs. (34a), (34b), (34c) together with Eqs. (39), (40) that $\mathfrak{B}\sim 0(\|w\|)$, $\mathfrak{C}\sim 0(\|w\|^2)$, and the kernel $a(x,y)$ of $\mathfrak{R}$ has the behavior

$$a(x,y)=2\sum_{l,p=1}^{2N}H_l(x)\tilde{H}_{lp}H_p(y)+O(\|w\|).$$

This yields expression (41). Since

$$-\tfrac{1}{2}F(x)=w(x)+O(\|w\|^2)$$

$$=\sum_{j=1}^{2N}\eta_j H_j(x)+O(\|w\|^2),$$

it follows from Eq. (35), that $\beta=2+O(\|w\|)$ and Eq. (36) that $\gamma\sim O(\|w\|)$. The results given by Eq. (42) then follow.

Since the operator $\mathfrak{F}'$ does not possess an inverse at $w\equiv 0$ [it is seen from Eq. (41), that $\mathfrak{F}'$ has finite rank at $w\equiv 0$], Newton's method for a minimizing sequence cannot be employed. Instead the descent approach (Vainberg[4]) given by the sequence

$$v_*^{n+1}-v_*^n=-\omega_n\alpha_n F(v_*^n) \tag{43}$$

will be employed. The real positive coefficient α_n in the descent method is obtained by[5,6] minimizing the function

$$g(t)=f(v_*^n-tF(v_*^n))$$

of the positive real variable t. However, it appears that common iteration techniques such as the Curry step-length specification to determine $\min t=\alpha_n$ are too complicated to apply here. This would require the solution of the direct scattering problem at each step. The simplest choice then is to choose α_n by minimizing the quadratic approximation to $g(t)$, giving

$$\alpha_n=\|F(v_*^n)\|^2/(\mathfrak{F}'F(v_*^n),F(v_*^n)). \tag{44}$$

Unfortunately, expression (44) cannot be directly used and has to be modified. There are two problems with (44). The derivation of (44) omits the contribution from the cubic term corresponding to the higher order Fréchet derivative which may dominate the quadratic term. In this case the value given by (44) may be too large. To take care of this, a relaxation factor ω_n is included in expression (43). A suitable value of $\omega_n<1$ would take care of this problem. However far more serious is the fact that the operator $\mathfrak{F}'$ is not a positive definite operator. Expression (44) may be negative! The inverse of the expression by the right-hand side of Eq. (44) can be written as the sum of two terms β_n and γ_n defined, respectively, by Eqs. (35) and (36) (with the subscript n denoting v_* replaced by v_*^n), where β_n is positive and γ_n is positive or negative. For negative values of γ_n it is better to use the expression

$$1/\alpha_n=\beta_n=(\mathfrak{R}F(v_*^n),F(v_*^n))/\|F(v_*^n)\|^2 \tag{45}$$

in place of Eq. (44).

In the neighborhood of the absolute minimum point $w\equiv 0$, β_n is the dominant term with $\gamma_n \to 0$. In fact, from Eq. (42) it is seen that

$$1/\alpha_n \sim 2 + O(\|w\|). \tag{46}$$

This provides a useful tool to indicate when the descent sequence has approached the linearized region of the solution.

The optimum choice of ω_n and α_n still has to be investigated. A particular choice for these quantities based upon preliminary numerical computations is given in a later section of this paper.

There still remains to be considered the question of the invertibility of the matrix $\{H_{ij}\}$ at each iteration step. It is assumed throughout that for each v_*^n the corresponding set $\{H_l(x)\}_{l=1}^{2N}$ is linearly independent over $\overline{D}$. All that is needed to achieve this is to select the data points x_l and wave number k_l so that $\{H_l(x)\}_{l=1}^{2N}$ is linearly independent for the initial value v_*^1 of the minimizing sequence $\{v_*^n\}$. The reason for this is that the quantity to be minimized, $f(v_*)$, involves the inverse matrix $\{\tilde{H}_{ij}\}$ [as indicated by Eq. (24)]; hence the effect of changes in the inverse matrix given by $\{\delta\tilde{H}_{ij}\}$ is included in the process for obtaining v_*^{n+1} from v_*^n. In expression (25) for the differential δf the first term on the right-hand side essentially is due to change in data, and the second term $\delta\tilde{H}_{ij}$ is due to the change in the inverse matrix. The term $q(x)$ in expression (28) for the gradient of f represents the change of the inverse matrix. Thus for step sizes (values of $\|v_*^{n+1} - v_*^n\|$) such that the Taylor expansion is valid, the matrix $\{\tilde{H}_{ij}\}$ will remain bounded.

The descent process will in theory terminate when a stationary point is reached. The stationary points are those values of v_* such that

$$-\tfrac{1}{2}F(v_*) \equiv w + q = 0. \tag{47}$$

The stationary point of main interest here is the one that gives the absolute minimum of $f(v_*)$, namely, $w\equiv 0$, which implies that $B_l = 0$ for $l = 1,2,\ldots,2N$. The other stationary points (which could be reached by a minimizing sequence) correspond to a local minimum or saddle point. Since q is the order of $\|w\|^2$ for small $\|w\|$, it is easily seen that a minimizing sequence is in the neighborhood of the stationary point of main interest $w\equiv 0$ if

$$\|w + \tfrac{1}{2}F(v_*)\|/\|w\| = \|q\|/\|w\| \ll 1$$

and in the neighborhood of the other stationary points if

$$\tfrac{1}{2}\|F(v_*)\|/\|w\| = \|w + q\|/\|w\| \ll 1.$$

In practice the descent approach may be terminated when either (i) the value v_*^n yields a computed value of the scattered field which differs from the measured scattered field by a number less than measurement error, i.e., $\Sigma_{l=1}^{2N} B_l^2 < \epsilon^2$ or (ii) when the descent approach has reached the linear region of solution corresponding to the absolute minimum. Once in the linear region of the solution, other techniques may be employed like that of Backus and Gilbert to get the optimum smoothest structure. The following result can be used to determine when the descent approach has approached the linearized region of the solution.

Theorem: A necessary condition for the descent approach to have reached the linear region of the solution is that

$$\|q\|/\|w\| \ll 1, \tag{48}$$

$$(\mathfrak{F}'F,F)/(F,F) \sim 2. \tag{49}$$

D. Regularization

The critical juncture of the procedure is in the solution of linear system (12). For the initial step (involving the initial choice of v_*) it was assumed that the data points $\{x_l\}$ and wave numbers $\{k_l\}$ are chosen so that $\{H_l(x)\}_{l=1}^{2N}$ is a linearly independent set of continuous functions over D; and any redundant set of measurements which violates this assumption would be culled out. As pointed out, theoretically this only has to be assured of initially, since the descent process will tend to retain the linear independence of $\{H_l\}$ for each v_*^n.

Recall that the minimum $\mathscr{L}_2(D)$ norm solution of system (12) is given by substituting expression (19)

$$w(x) = \sum_{j=1}^{2N} \eta_j H_j(x)$$

into system (12). This yields the algebraic system of equations for $\boldsymbol{\eta}$

$$H\boldsymbol{\eta} = \mathbf{B}. \tag{50}$$

In the numerical computation of the solution of system (50) constraints have to be imposed that involve some of the following:

(i) errors in H;
(ii) errors in the data or $\{B_j\}$;
(iii) bounds on the solution $\boldsymbol{\eta}$.

The errors in the matrix elements H_{ij} are the most serious. These errors arise from two sources, numerical errors (due to computational procedures, round-off, etc.) and model error due to the approximation of the space of continuous functions over D by a finite-dimensional space. With H a positive definite matrix, the errors in H will distort the component of the solution belonging to the eigenspace spanned by the eigenfunctions corresponding to the small eigenvalues. If the error in the matrix H is denoted by $E = \{E_{ij}\}$, then the numerical solution of system (50) will be required to the following accuracy:

$$\|H\boldsymbol{\eta} - \mathbf{B}\|_2 \leqslant \|E\|_2 \|\boldsymbol{\eta}\|_2, \tag{51}$$

where $\|\boldsymbol{\eta}\|_2 = \Sigma_{j=1}^{2N} \eta_j^2$.

No constraint need be placed upon the bound of the solution $\boldsymbol{\eta}$ for the iterative scheme developed here. This is in contrast to the schemes employed by Coen, Mei, and Angelakos[7] and to the Levenberg–Marquardt algorithm (Moré,[8] Levenberg,[9] and Marquardt[10]), which require such a constraint due to a condition like (6) being imposed on their iteration process.

Errors in the data (provided they are not too large) can be ignored initially in the nonlinear iterative scheme. They become important in the linear region where v_* is close to v_m. Generally, since no bounds are placed on $\|\boldsymbol{\eta}\|_2$, the error term $\|E\boldsymbol{\eta}\|_2$ may be larger than the error in the data.

With only the constraint (51) being considered, the system of equations can be regularized (Miller[11]) by replacing them by

$$(H + \varepsilon I)\boldsymbol{\eta} = \mathbf{B}, \tag{52}$$

where $\varepsilon = \|E\|_2$.

The regularization procedure is more important in the initial steps of the iteration procedure since the descent process tends to increase the values of the smallest eigenvalues of the positive definite matrix H as a function v_*^n at each step. This is seen as follows. If $\lambda_1, \lambda_2, \ldots, \lambda_{2N}$ are the eigenvalues of H with $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_{2N}$ the corresponding set of orthonormal eigenvectors, the solution of system (52) is given by

$$\boldsymbol{\eta} = \sum_{i=1}^{2N} \frac{(\mathbf{B}\cdot\boldsymbol{\phi}_i)\boldsymbol{\phi}_i}{\lambda_i + \varepsilon},$$

and the function that is minimized by the descent process can be expressed in the form

$$f(v_*) = \int_D w(v_*)^2\, dy = \sum_{i=1}^{2N} \frac{\lambda_i (\mathbf{B}\cdot\boldsymbol{\phi}_i)^2}{(\lambda_i + \varepsilon)^2}.$$

When the regularized system (52) is used, the expressions (28) and (30) for the gradient $F(x)$ and the derivative $F'(x,y)$ are modified. Apart from terms the order of ε which can be neglected for small ε, the major change occurs in the first term on the right-hand side of Eq. (30), where the matrix coefficient $\tilde{H}_{lp}$ is replaced by the corresponding coefficient of the inverse matrix $(H + \epsilon I)^{-1}$.

For regularization techniques applied to optical problems, see Ref. 12.

E. Constrained iterative process

The descent procedure can be modified to give a solution satisfying the constraint

$$n^2(x) \geqslant 1, \quad x \in D.$$

This can be achieved by replacing the iterate

$$v_*^{n+1}(x) = v_*^n(x) - \omega_n \alpha_n F(v_*^n, x) \tag{53}$$

by $v_*^{n+1}(x) = 0$, at all points x in D where the right-hand side of Eq. (53) is negative. This process is equivalent to adding a function $\psi(x)$ to the gradient F so that

$$v_*^{n+1} = v_*^n - \omega_n \alpha_n (F(v_*^n, x) + \psi(x)) \geqslant 0$$

essentially modifying the direction of $v_*^{n+1} - v_*^n$ from that of the negative gradient to that of a deviation from it. The descent process will still be valid provided that the deviated direction is not more than 90°, i.e.,

$$(F + \psi, F) > 0,$$

and the relaxation factor ω_n or α_n is properly chosen.

V. COMPUTATIONAL RESULTS

In order to illustrate the application of the method, some preliminary computational results for the one-dimensional problem are presented here. More details of the computational aspects will be presented elsewhere including eventual computational results for higher dimensions.

In the one-dimensional case the obstacle (or slab in this case) will be located in a region D where D is the interval $-1 \leqslant x \leqslant 1$. Results are presented here for the case where the actual obstacle is inside the domain D. In particular, the scattering obstacle will be given by

$$n = 1, \quad -1 \leqslant x - 0.05,$$
$$n = 2 + 20x, \quad -0.05 \leqslant x \leqslant 0,$$
$$n = 2, \quad 0 \leqslant x \leqslant 1.$$

Since $n \equiv 1$ outside D, the actual obstacle lies in the interval $-0.05 \leqslant x \leqslant 1$.

The factors α_n and ω_n used in the computational algorithm corresponding to Eq. (43) are essentially the same as given in the body of the paper for the case when $\gamma_n \geqslant 0$. In this case $\omega_n = 1$ and α_n is given by Eq. (44). However, for the case where $\gamma_n < 0$, these factors are modified as follows:

(i) for $\|q\|/\|w\| \geqslant 0.9$,

$$\omega_n \alpha_n = (\beta_n + \gamma_n)^{-1}, \quad \text{for } -\beta_n/2 \leqslant \gamma_n < 0,$$
$$\omega_n \alpha_n = 2/\beta_n, \quad \text{for } \gamma_n < -\beta_n/2;$$

(ii) for $\|q\|/\|w\| < 0.9$,

$$\omega_n \alpha_n = 1/\beta_n, \quad \gamma_n < 0.$$

The computational algorithm was modified by using the constraint $n \geqslant 1$ described in the previous section. The regularization factor ε was taken to be $\varepsilon = 0.000\,01$.

Computations were carried out using "measured" (actually computed) values of the backscattered field for a set of different frequencies. Different choices of initial values of $n(x)$ were used. Briefly, some of the results are as follows. For data at the wave numbers $k = \pi/4, \pi/2, 3\pi/4, \pi$, the algorithm converged (as far as the actual iterations were carried out) to the correct solution, when the initial choice for $n(x)$ was (i) $n(x) = 1.25$, $-1 \leqslant x \leqslant 1$, and (ii) $n(x) = 1.5$, $-1 \leqslant x \leqslant 1$. However, the algorithm converged to a different solution when the initial value of $n(x)$ was $n(x) = 1.75$, $-1 \leqslant x \leqslant 1$, but with the addition of more data (backscattered field at different frequencies) the algorithm tended to local minimum instead of indicating that it was not a true solution. One interesting thing is that if the set of four frequencies were reduced to $k = \pi/8, \pi/4, \pi/2, 3\pi/4$, the algorithm converged to the correct solution when the initial value of $n(x)$ was $n(x) = 1.75$, $-1 \leqslant x \leqslant 1$. Some of the computational results are presented in Fig. 1.

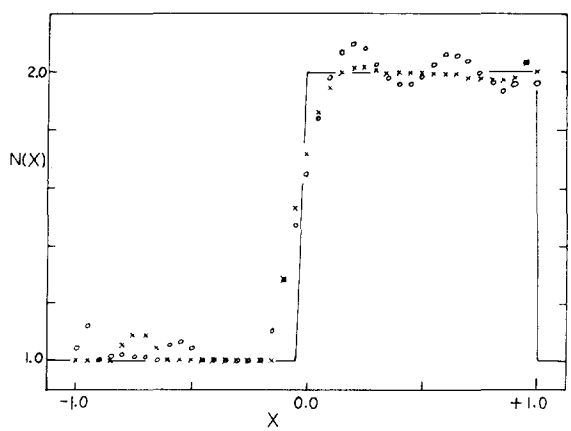


FIG. 1. Computed and actual value of the index of refraction $N(X)$: (—) actual profile of scattering body; (o o o) computed profile for 20 iterations starting with initial value $n(x) = 1.5$ for $-1 < x < 1$, with measurements at $k = \pi/4, \pi/2, 3\pi/4, \pi$; (× × ×) computed profile for 20 iterations starting with initial value $n(x) = 1.75$ for $-1 < x < 1$, with measurements at $k = \pi/8, \pi/4, \pi/2, 3\pi/4$.

VI. COMMENTS

The inverse-scattering algorithm works quite well for the sparse data case based upon initial testing. Since the algorithm is sensitive to the number of small eigenvalues of the initial matrix $\{H_{ij}\}$ [the value of $\{H_{ij}\}$ derived from the initial choice of $n(x)$ in the interation procedure], the algorithm may or may not be useful for the large data case. This remains to be seen. Here the regularization parameter ε will play a much more important role than in the sparse data case.

In the application of the iteration procedure here, no *a priori* knowledge was used to choose the initial value of $n(x)$. It would be extremely useful to see if a good initial choice for $n(x)$ can be made based upon physical grounds such as the Born approximation, and when such choices of initial $n(x)$ lead to the correct solution.

The technique developed here can be abstracted and applied to other inverse scattering problems, either scalar or vector. At present it is being applied to the inverse problem associated with acoustic scattering (reduced wave equation) by an inpenetrable object, characterized by a Dirichlet boundary condition on its surface.

ACKNOWLEDGMENT

The research reported herein was supported herein was supported in part through O. N. R. Grant N14-83-K-0038.

APPENDIX A

Here, some details on the nature of the solution to the nonhomogeneous reduced wave equation

$$\Delta u + k^2 n^2(x) u = -(n^2 - 1) f(x) \tag{A1}$$

are presented. The function $f(x)$ is taken to be continuous, and $n(x)$ has the sectionally continuous properties stated in the introduction. To be more precise, $n(x)$ is bounded everywhere and is identical to unity in the region exterior to $\overline{D}$. The compact region D is decomposable into a finite number N' of regions D_i, each with piecewise smooth boundaries δD_i, such that $n(x)$ is continuous in each closed domain $\overline{D}_i$.

With the free-space Green's function given by

$$G_0(x,y) = e^{ik|x-y|}/4\pi|x-y|.$$

Equation (A1) can be expressed in terms of the integral equation

$$u(x) - k^2 \int_D G_0(x,y) v(y) u(y)\, dy = \int_D G_0(x,y) v(y) f(y)\, dy, \tag{A2}$$

where $v(x) = n^2(x) - 1$. With the decomposition

$$\int_D G_0(x,y) v(y) u(y)\, dy = \sum_{i=1}^{N'} \int_{D_i} G_0(x,y) v(y) u(y)\, dy$$

and the fact that with $v(x)$ belonging to $C(\overline{D})$ the operator $\mathfrak{G}_i$ given by

$$\mathfrak{G}_i u = \int_{D_i} G_0(x,y) v(y) u(y)\, dy$$

maps $C(\overline{D}_i)$ into $C^1(R^3)$, it is seen that the solution $u(x)$ of Eq. (A2), when it exists, will belong to $C^1(R^3)$. Since the integral operator in Eq. (A2) (an integral equation of the second kind) is compact, uniqueness implies existence of solution. Because the solution for the corresponding homogeneous equation belongs to $C^1(R^3)$, uniqueness follows [proof identical to Leis[13] for the case where $n(x)$ is continuous everywhere]. As an immediate result, it is seen that the scattered field $u^s(x)$ produced by the incident wave $u^i(x)$ belongs to $C^1(R^3)$. This follows on taking $f(x) = k^2 u^i(x)$ in Eq. (A1).

In addition, the properties of the Green's function $G(x,x_0;k,v)$ associated with Eq. (1) may be obtained. With the decomposition

$$G(x,x_0;k,v) = G_0(x,x_0) + k^2 \int_D G_0(x,z) G_0(z,x_0) v(z)\, dz + g(x,x_0), \tag{A3}$$

it can be shown that the term $g(x,x_0)$ satisfies the integral equation (A2) with

$$f(x) = k^4 \int_D G_0(x,z) G_0(z,x_0) v(z)\, dz. \tag{A4}$$

With the decomposition of the integral expression over D into integrals over D_i, it can be shown that $f(x)$ given by Eq. (A4) belongs to $C(R^3)$. Thus it follows from symmetry and previous arguments that the second and third terms on the right-hand side of Eq. (A3) belong to $C(R^3)\times C(R^3)$ and $C^1(R^3)\times C^1(R^3)$, respectively.

APPENDIX B: EXPLICIT EXPRESSIONS FOR δB_j, H_j', AND H_j''

As a preliminary, the differentials δG and δu, where u is the total field, are required. Using the differential equation for $G(x,y;k,v)$

$$\Delta G(x,y;k,v) + k^2(v+1) G(x,y;k,v) = -\delta(x-y)$$

and the similar equation for $G(x,y;k,v+\delta v)$, it can be shown that

$$G(x,y;k,v+\delta v) = G(x,y;k,v) + k^2 \int_D G(x,z;k,v) G(y,z;k,v+\delta v)\, \delta v(z)\, dz,$$

from which it follows that

$$\delta G(x,y;k,v) = k^2 \int_D G(x,z;k,v) G(y,z;k,v)\, \delta v(y)\, dy. \tag{B1}$$

From integral equation (3) for the total field

$$u(x,k;v+\delta v) = u(x,k;v) + k^2 \int_D G(x,y;k,v) u(y,k;v+\delta v)\, \delta v(y)\, dy,$$

it follows that

$$\delta u(x,k;v) = k^2 \int_D G(x,y;k,v) u(y,k;v)\, \delta v(y)\, dy. \tag{B2}$$

Now, from Eq. (9) and (11), B_l is given explicitly by

$$B_l + iB_{l+N} = u_m^s(x_l,k_l) - u^s(x_l,k_l,v);$$

hence, on using Eq. (B2), it is seen that

$$\delta(B_l + iB_{l+N}) = -k_l^2 \int_D G(x_l,y;k_l,v) u(y,k_l;v)\, \delta v(y)\, dy,$$

which, on using Eq. (10), yields

$$\delta B_l = -(H_l,\delta v), \quad l = 1,2,...,2N. \tag{B3}$$

From Eq. (10) it is seen that

$$\delta(H_l + iH_{l+N}) = k_l^2 u(y,k_l;v)\,\delta G(x_l,y;k_l,v) + k_l^2 G(x_l,y;k_l,v)\,\delta u(y,k_l;v);$$

hence, on using Eqs. (B1) and (B2), it follows that

$$\delta H_l = \int_D H'(x,y,v)v(y)\,dy,$$

where

$$H'_l(x,y;v) + iH'_{l+N}(x,y;v) = k_l^4 G(x,y;k_l,v)[u(y,k_l;v)G(x_l,x;k_l,v) + u(x,k_l;v)G(x_l,y;k_l,v)]. \tag{B4}$$

For v sectionally continous function, the terms in the square brackets will be continuous functions of x and y over $\overline{D}$. Hence, from the remarks made in Appendix A, H'_l as a kernel of an integral operator will map $C(\overline{D})$ into $C^1(\overline{D})$.

One can proceed in a similar position and show that if

$$\delta H'_l = \int_D H''_l(x,y,z;v)\,\delta v(z)\,dz,$$

then

$$H''_l(x,y,z;v) + iH''_{l+N}(x,y,z;v) = Q(x,y,z) + Q(y,z,x) + Q(z,x,y), \tag{B5}$$

where

$$Q(x,y,z) = k_l^6 u(x,k_l;v)G(y,z;k_l,v) \times [G(x,z;k_l,v)G(x_l,y;k_l,v) + G(x,y;k_l,v)G(x_l,z;k_l,v)]. \tag{B6}$$

Since the points x_l are exterior to $\overline{D}$, it follows from the results in Appendix A on the behavior of the scattered field $u^s(x)$ and the Green's function that $H''_l(x,y,z;v)$ has the general form

$$H''_l(x,y,z;y) = \frac{A(x,y,z)}{|x-y||y-z|} + \frac{A(y,z,x)}{|y-z||z-x|} + \frac{A(z,x,y)}{|z-x||x-y|}, \tag{B7}$$

where $A(x,y,z)$ is a continous function of x, y, and z over $\overline{D}$.

APPENDIX C: EXPLICIT EXPRESSIONS FOR $F(x)$, $F'(x,y)$

From Eq. (16), it can be deduced that the differential of the inverse matrix $\{\tilde{H}_{ij}\}$ is given by

$$\delta\tilde{H}_{ij} = -\sum_{k,p=1}^{2N} \tilde{H}_{ip}\,\delta H_{pk}\,\tilde{H}_{kj}, \tag{C1}$$

where from Eq. (15)

$$\delta H_{pk} = (\delta H_p,H_k) + (H_p,\delta H_k). \tag{C2}$$

From Eq. (20) and (C1) it follows that

$$\sum_{i,j=1}^{2N} B_i\,\delta\tilde{H}_{ij}B_j = -\sum_{k,p=1}^{2N} \eta_p\,\delta H_{pk}\,\eta_k,$$

which reduces to

$$= -2\sum_{p=1}^{2N} \eta_p\,(\delta H_p, w) \tag{C3}$$

on using Eq. (19). Inserting this into Eq. (25), the differential of f takes the form

$$-\tfrac{1}{2}\delta f = (w,\delta v_*) + \sum_{p=1}^{2N} \eta_p(\delta H_p,w) = (w+q,\delta v_*), \tag{C4}$$

where on using Eq. (27)

$$q(x) = \sum_{p=1}^{2N} \eta_p \int_D H'_p(x,z)\delta v_*(z)\,dz. \tag{C5}$$

To compute the second differential of f, the differential of η_j is needed. From Eq. (20), $\delta\eta_j$ is given by

$$\delta\eta_j = \sum_{k=1}^{2N} [B_k\,\delta\tilde{H}_{jk} + \tilde{H}_{jk}\,\delta B_k]$$

which on using Eqs. (B3) and (C2) reduces to

$$\delta\eta_j = -\sum_{p=1}^{2N} \tilde{H}_{jp}\left[\sum_{k=1}^{2N} \eta_k\,\delta H_{pk} + (H_p,\delta v_*)\right] = -\sum_{p=1}^{2N} \tilde{H}_{jp}(R_p,\delta v_*), \tag{C6}$$

where from Eqs. (C2) and (19)

$$R_p(x) = H_p(x) + \int_D \{H'_p(x,z)w(z) + \sum_{k=1}^{2N} \eta_k H'_k(x,z)H_p(z)\}\,dz. \tag{C7}$$

From expression (C4) for δf, the second differential is seen to have the form

$$-\tfrac{1}{2}\delta^2 f = (\delta w,\delta v_*) + \sum_{p=1}^{2N} \delta\eta_p(\delta H_p,w) + \sum_{p=1}^{2N} \eta_p[(\delta^2 H_p,w) + (\delta H_p,\delta w)]. \tag{C8}$$

Using the result obtained from Eq. (19)

$$\delta w = \sum_{j=1}^{2N} (H_j\,\delta\eta_j + \eta_j\,\delta H_j) \tag{C9}$$

expression (C8) reduces to

$$-\tfrac{1}{2}\delta^2 f = \sum_{j=1}^{2N} \delta\eta_j\left[(H_j,\delta v_*) + (\delta H_j,w) + \sum_{p=1}^{2N} \eta_p(\delta H_p,H_j)\right] + \sum_{p=1}^{2N} \eta_p\left[(\delta H_p,\delta v_*) + (\delta^2 H_p,w) + \sum_{j=1}^{2N} \eta_j(\delta H_p,H_j)\right] = \sum_{j=1}^{2N} \delta\eta_j(R_j,\delta v_*) + \sum_{j,p=1}^{2N} \eta_p\eta_j(\delta H_p,\delta H_j) + \sum_{p=1}^{2N} \eta_p[\delta H_p,\delta v_*) + (\delta^2 H_p,w)]. \tag{C10}$$

Finally, inserting expression (C6) into (C15) and using the definition

$$\delta^2 f = ((F',\delta v),\delta v),$$

it is seen that

$$\tfrac{1}{2}F'(x,y) = \sum_{j,p=1}^{2N} \Big\{ R_j(x)\tilde{H}_{jp}R_p(y) - \eta_j\eta_p \int_D H'_p(x,z)H'_j(y,z)\,dz \Big\} - \sum_{p=1}^{2N} \eta_p \Big\{ H'_p(x,y) + \int_D H''_p(x,y,z)w(z)\,dz \Big\}. \tag{C11}$$

Using the result that $\{H_l(x)\}_{l=1}^{2N}$, $w(x)$ and $q(x)$ are continuous functions over $\bar{D}$, $H'_l(x,y)$ is a polar kernel, and $H''_l(x,y,z)$ has the general form given by Eq. (B7), it can be shown that $F'(x,y)$ is a polar kernel with general form

$$F'(x,y) = \gamma(x,y)/|x-y|, \tag{C12}$$

where $\gamma(x,y)$ is $C(\bar{D})\times C(\bar{D})$.

[1]V. H. Weston, "Inverse problem for the reduced wave equation with fixed incident field. II," J. Math. Phys. **22**, 2523–9 (1981).

[2]V. H. Weston, "Inverse problem for the reduced wave equation with fixed incident field. III," J. Math. Phys. **24**, 828–33 (1983).

[3]G. Backus and F. Gilbert, "Uniqueness in the Inversion of Inaccurate Gross Earth Data," Philos. Trans. R. Soc. London Ser. A **226**, 123–92 (1970).

[4]M. M. Vainberg, *Variational Method and Method of Monotone Operators in the Theory of Non-Linear Equations* (Wiley, New York, 1973).

[5]J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Non-Linear Equations in Several Variables* (Academic, New York, 1970).

[6]W. C. Rheinboldt, *Methods for Solving Systems of Non-Linear Equations*, Regional Conference Series in Applied Mathematics (SIAM, Philadelphia, 1974).

[7]S. Coen, K. Mei, and D. J. Angelakos, "Inverse Scattering Technique Applied to Remote Sensing Layered Media," IEEE Trans. **AP-29**, 298–306 (1981).

[8]J. J. Moré, *The Levenberg–Marquardt Algorithm Implementation and Theory*, Lecture Notes in Mathematics Vol. 30, Proceedings (Springer-Verlag, Berlin, 1978), pp. 105–16.

[9]K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," Quart. J. Appl. Math. **2**, 164–8 (1944).

[10]D. W. Marquardt, "An Algorithm for Least Squares Estimation of Non-Linear Parameters," SIAM J. Appl. Math. **11**, 431–41 (1963).

[11]K. Miller, "Least Squares Methods for Ill-Posed Problems with a Prescribed Bound," SIAM J. Math. Anal. **1**, 52–74 (1970).

[12]M. Bertero, C. DeMol, and E. R. Pike, Opt. Acta **29**, 727, 1599 (1982).

[13]R. Leis, *Vorlesungen über Partielle Differential-Gleichungen Zweiter Ordnung* (Bibliographisches Institut, Mannheim, 1967).

Linear response theory revisited. IV. Applications

Pagiotis Vasilopoulos[a] and Carolyn M. Van Vliet
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Quebec H3C 3J7, Canada and Department of Electrical Engineering, University of Florida, Gainesville, Florida 32611

(Received 27 July 1983; accepted for publication 4 November 1983)

The one-body linear response expressions (diagonal and nondiagonal) for the dc electrical conductivity, obtained in a previous paper, are applied to specific situations. In the case of "no collisional current" we evaluate the relaxation times which enter an expression (diagonal) for the longitudinal magnetoconductivity; in the case of "collisional current" another expression (diagonal) for the transverse magnetoconductivity is evaluated. The calculations are carried out in the framework of electron–phonon interaction in crystalline materials; various kinds of phonons are considered. The formula for "collisional current" is also used for an analytic evaluation of the phonon assisted hopping conductivity in crystalline and amorphous materials. The results are in harmony with those of the literature or are new. Further, the nondiagonal expression leads to a result for the oscillatory Hall effect, which, in contrast with previous results, is independent of the interaction (e.g., interaction with impurities).

PACS numbers: 03.60. + w, 72.10. − d, 72.10.Di

1. INTRODUCTION

In the first paper of this series,[1] referred to as LRTI, the Kubo–Green formulas, which relate transport coefficients to certain forms of the correlation function of fluctuations about an equilibrium state, were discussed, and a reinterpretation of linear response theory was given. In particular, it was argued that in Kubo's theory proper no dissipation occurs as reflected by zero entropy production and the Heisenberg form for the time-dependent operator $B(t)$ of the system. Dissipative behavior was introduced by writing the system Hamiltonian as $H = H^0 + \lambda V$, where H^0 represents the motion proper and λV the interaction causing randomizing transitions between the eigenstates of H^0 (e.g., electron–phonon interaction). The application of the van Hove limit, $\lambda \to 0$, $t \to \infty$, $\lambda^2 t$ finite, led to an entirely different time behavior of the reduced operators $B^R(t)$. In the subdynamics of H^0 the reduced operators show clearly relaxation

$$K_d^R(t) = e^{-\Lambda_d t} K_d^R(0); \tag{1.1}$$

here $K_d^R(0) = K_d \equiv K_d^S$ is the Schrödinger operator, and "d" denotes the diagonal part in the representation of H^0. Λ_d is the master superoperator in Liouville space, defined by

$$\Lambda_d K = -\sum_{\gamma\gamma'} |\gamma\rangle\langle\gamma| [W_{\gamma'\gamma} \langle\gamma'|K|\gamma''\rangle - W_{\gamma\gamma'}\langle\gamma|K|\gamma\rangle], \tag{1.2}$$

where $|\gamma\rangle$ are the eigenstates of H^0, with eigenvalues ϵ_γ, and where the transition rate $W_{\gamma\gamma'}$ is given by the golden rule (the Van Hove limit is equivalent to the first Born approximation)

$$W_{\gamma\gamma'} = \frac{2\pi\lambda^2}{\hbar} |\langle\gamma|V|\gamma''\rangle|^2 \delta(\epsilon_\gamma - \epsilon_{\gamma'}) = W_{\gamma'\gamma}. \tag{1.3}$$

In the second paper of the series,[2] referred to as LRTII, it was shown that the linear response formulae of LRTI in the Van Hove limit could be obtained without previous knowledge of the Kubo–Green formulae. To that purpose, Zwanzig's projection operator technique was applied to the von Neumann equation for the total density operator. The total Hamiltonian was $H = H^0 + \lambda V - AF(t)$, where $-AF(t)$, the field Hamiltonian, represents the coupling of the system to an external field, with $F(t)$ being a generalized force and A the conjugate extensive operator. The "projection" was followed by the application of the Van Hove limit, and this led to a many-body inhomogeneous master equation (diagonal), which contains not only the relaxation terms of the Pauli master equation but also the coupling to the external field. The solution of this master equation gave the new many-body response formulae identical with those of LRTI. At the same time the solution of the nondiagonal master equation obtained in LRTII led to the nondiagonal many-body response formulas for the susceptibility and conductivity.

In the third paper,[3] referred to as LRTIII, a reduction was made of the many-body results of LRTI and LRTII to one-body results. To that purpose, the Hamiltonian H^0 was considered to represent a fermion and a boson gas and λV the interaction between them being of binary nature. In second quantization formalism it is

$$H^0 = \sum_\zeta \mathbf{n}_\zeta \epsilon_\zeta + \sum_\eta \mathbf{N}_\eta E_\eta, \tag{1.4}$$

$$\lambda V = \sum_{\zeta'\zeta''\eta'\eta''} c_{\zeta''}^\dagger a_{\eta''}^\dagger (\zeta''\eta''|\lambda|\zeta'\eta') a_{\eta'} c_{\eta}, \tag{1.5}$$

$$|\gamma\rangle = |\{n_\zeta\}, \{N_\eta\}\rangle = |\{n\}\rangle \otimes |\{N\}\rangle. \tag{1.6}$$

In these expressions $\{|\zeta\rangle\}$, $\{|\eta\rangle\}$ denote the sets of one-particle states for fermions and bosons with eigenvalues ϵ_ζ and E_η, respectively. $\mathbf{n}_\zeta = c_\zeta^\dagger c_\zeta$ and $\mathbf{N}_\eta = a_\eta^\dagger a_\eta$ are the number operators with eigenvalues n_ζ and N_η, while the c's and the a's are the creation and annihilation operators for fermions and bosons, respectively.

As for the field Hamiltonian $-A \cdot F(t)$, in the case of an externally applied electric field $E(t)$, it is $\mathbf{F}(t) = q\mathbf{E}(t)$ and $\mathbf{A} = \Sigma_i(\mathbf{r}_i - \langle\mathbf{r}_{eq}\rangle)$, where q is the charge of the carriers, $\mathbf{r}_i$ their positions, and $\langle\mathbf{r}_i\rangle_{eq}$ their positions prior to the switching on of the electric field. The electrical current Schrödinger operator is $\mathbf{J} = q\Sigma_i \mathbf{v}_i/\Omega = q\dot{A}/\Omega$, where Ω is the volume of the sample.

[a] Presently at the Max Planck Institut fur Festköperforschung, Stuttgart, FRG.

 0022-2488/84/051391-13$02.50

The diagonal form of the reduced current operator (denoted by the superscript R) of a variable $\dot{B}$ is [cf. LRTII Eqs. (4.28) and (4.29)]

$$J^R_{B,d} \equiv \dot{B}^R_d = -\Lambda_d B_d + (\dot{B})_d, \tag{1.7}$$

where the time-dependence for $(\dot{B}^R(t))_d$ is again given by (1.1). Then, the many-body formula for the diagonal part of the conductivity, denoting by Greek subscripts the tensor components, reads ($\beta = 1/kT$)

$$\sigma^d_{\mu\nu}(i\omega) = \beta\Omega \int_0^\infty dt\, e^{-i\omega t}\, \mathrm{Tr}[\rho_{\mathrm{eq}} J^R_{dv} J^R_{d\mu}(t)], \tag{1.8}$$

with

$$\mathbf{J}^R_d = \frac{q}{\Omega}\left[-\Lambda_d \sum_i (\mathbf{r}_i - \mathbf{r}^{\mathrm{eq}}_i)_d + \sum_i \mathbf{v}_{id}\right]; \tag{1.9}$$

see Eqs. (2.5) and (2.6) of LRTIII. The reduction of the many-body expression (1.8) [as well of the expression (1.18); see below] to a one-body expression was made in LRTIII with the help of a few theorems connecting the many-body aspects, expressed by the superoperator Λ_d (or the master operator M), with the one-body aspects, represented by the Boltzmann operator [see LRTIII, Eqs. (2.12), (2.16), (2.30), and (2.32)]; the latter is defined, in function space, by

$$\mathscr{B}_\xi f(\xi) = \sum_{\xi'} \{w_{\xi\xi'} f(\xi)[1 - f(\xi')] - w_{\xi'\xi} f(\xi')[1 - f(\xi)]. \tag{1.10}$$

The fermion transition rates $w_{\xi\xi'}$ are given by

$$w_{\xi''\xi'} = \sum_{\eta'\eta''} Q(\xi'',\eta'';\xi',\eta')\langle N_{\eta''}(1 + N_{\eta'})\rangle_{\mathrm{eq}}$$

$$\approx \sum_{\eta'\eta''} Q(\xi'',\eta'';\xi',\eta')\langle N_{\eta''}\rangle_{\mathrm{eq}}(1 + \langle N_{\eta'}\rangle_{\mathrm{eq}}); \tag{1.11}$$

as usual, it has been assumed that the bosons remain at equilibrium. The latter equality is based on the truncation rule of LRTII, Eq. (8.1) and the Q's are the binary transition rates [see LRTII, Eq. (8.18)]. Using the equilibrium Bose–Einstein distribution, one finds

$$w_{\xi''\xi'} = w_{\xi'\xi''} e^{\beta(\epsilon_{\xi''} - \epsilon_{\xi'})} \tag{1.12}$$

[cf. LRTIII, Eq. (2.15)].

The first term of (1.9) or (1.7) represents the "collisional" current of LRTII and LRTIII, the second term the "ponderomotive" current. In the absence of collisional current, i.e., when $(\xi|\mathbf{r}_i - \mathbf{r}^{\mathrm{eq}}_i|\xi) = 0$ [see the definition of Λ_d, Eq. (1.2)], the many body conductivity (1.8) reduces to [cf. LRTIII, Eq. (2.55)]

$$\sigma^d_{\mu\nu}(i\omega) \approx -\frac{q^2}{\Omega}\sum_\xi \frac{\partial\langle n_\xi\rangle_{\mathrm{eq}}}{\partial\epsilon_\xi} \frac{v_{\nu\xi}v_{\mu\xi}}{i\omega + 1/\tau(\epsilon_\xi)}, \tag{1.13}$$

where the relaxation time $\tau(\epsilon_\xi)$ is defined by Eq. (2.53) of LRTIII, i.e.,

$$\mathscr{B}^0_\xi v_{\mu\xi} = \sum_{\xi'} w_{\xi\xi'}(v_{\mu\xi} - v_{\mu\xi'}) = \frac{1}{\tau(\epsilon_\xi)} v_{\mu\xi}, \tag{1.14}$$

v_ξ is the velocity matrix element $(\xi|\mathbf{v}|\xi)$. Equation (1.14) is valid for near-elastic collisions as can be seen from (1.10) by writing $w_{\xi\xi'} \approx w_{\xi'\xi}$.[4] More interesting, however, formulas (1.13) and (1.14) remain unaltered when we deal with nondegenerate systems and inelastic collisions. In this case (1.13) has been obtained by solving the quantum Boltzmann equation of LRTII; it then turns out[5,6] that one has to evaluate the quantity $\mathscr{B}_\xi\langle n_\xi\rangle_{\mathrm{eq}} v_{\mu\xi}$. Using the definition of $\mathscr{B}_\xi$, Eq. (1.10), and taking $\langle n_\xi\rangle_{\mathrm{eq}} \approx e^{-\beta(\epsilon_\xi - \epsilon_F)}$, one finds, on account of (1.12),

$$\mathscr{B}_\xi\langle n_\xi\rangle_{\mathrm{eq}} v_{\mu\xi} = \langle n_\xi\rangle_{\mathrm{eq}} \sum_{\xi'} w_{\xi\xi'}(v_{\mu\xi} - v_{\mu\xi'}) = \langle n_\xi\rangle_{\mathrm{eq}} \mathscr{B}^0_\xi v_{\mu\xi}. \tag{1.15}$$

The dc limit of (1.13) reads ($\mu = \nu = z, j = -qv/\Omega$)

$$\sigma_{zz}(0) \equiv \sigma_{zz} = -\Omega \sum_\xi \frac{\partial\langle n_\xi\rangle_{\mathrm{eq}}}{\partial\epsilon_\xi} \tau(\epsilon_\xi) j_{z\xi} j_{z\xi}. \tag{1.16}$$

When there is only collisional current, i.e., when the second term of (1.9) vanishes. The expression (1.8) reduces in the dc limit ($\mu = \nu = x$), to

$$\sigma^d_{xx}(0) = \sigma_{xx} = \frac{\beta q^2}{\Omega} \sum_{\xi\xi'} \langle n_\xi\rangle_{\mathrm{eq}}(1 - \langle n_{\xi'}\rangle_{\mathrm{eq}}) W_{\xi\xi'}(X_\xi - X_{\xi'})^2, \tag{1.17}$$

where $X_\xi = (\xi|x|\xi)$, cf. LRTIII, Eq. (2.84). This formula is the extended Adams–Holstein result[7] for processes involving inelastic binary collisions. It was first given by Argyres and Roth.[8] It has also been obtained to within a proportionality constant, for amorphous materials, by Capek[9] and by Zvagin.[10] Likewise, it can be derived by solving the quantum Boltzmann equation of LRTII.[11,6]

The many-body nondiagonal part of the conductivity reads

$$\sigma^{\mathrm{nd}}_{\mu\nu}(i\omega) = \Omega \int_0^\infty dt\, e^{-i\omega t} \int_0^\beta d\beta'\, Tr[\rho_{\mathrm{eq}} J^R_{\mathrm{nd}\nu}(-i\hbar\beta') J^R_{\mathrm{nd}\mu}(t)]. \tag{1.18}$$

Here $J^R_{\mathrm{nd}} \equiv \Sigma q v_{\mathrm{nd}}/\Omega$, there being no collisional current for the nondiagonal part [compare Eq. (1.7)]. Further,

$$J^R_{\mathrm{nd}}(t) = e^{iH^0 t/\hbar} J_{\mathrm{nd}} e^{-iH^0 t/\hbar} \tag{1.19}$$

and

$$(J^R(-i\hbar\beta))_{\mathrm{nd}} = e^{\beta H^0} J_{\mathrm{nd}} e^{-\beta H^0}; \tag{1.20}$$

see LRTIII, Eqs.(3.3), (3.4), and (3.5). The one-body result, obtained from (1.18), reads [compare LRTIII, Eq. (3.21)]

$$\sigma^{\mathrm{nd}}_{\mu\nu}(i\omega) = \Omega\hbar \sum_{\xi'\xi''}{}' \langle n_{\xi'}\rangle_{\mathrm{eq}}(1 - \langle n_{\xi''}\rangle_{\mathrm{eq}}) \times (\xi'|j_\nu|\xi'')(\xi''|j_\mu|\xi') \frac{1 - e^{-\beta(\epsilon_{\xi''} - \epsilon_{\xi'})}}{\epsilon_{\xi''} - \epsilon_{\xi'}} \times \left[i\mathscr{P}\frac{1}{\epsilon_{\xi''} - \epsilon_{\xi'} - \hbar\omega} + \pi\delta(\epsilon_{\xi''} - \epsilon_{\xi'} - \hbar\omega)\right], \tag{1.21}$$

where $\mathscr{P}$ denotes the principal part and where the prime on Σ means $\xi' \neq \xi''$; this result can also be obtained[6] from the solution of the nondiagonal Boltzmann equation of LRTIII.

The reason for this somewhat lengthy introduction was to present partly the line of development from LRTI to LRTIII as well as all the necessary formulas for the calculations of the subsequent sections.

The paper is organized as follows: in the next section, dealing with longitudinal magnetoresistance, we will evaluate the relaxation times entering the expression (1.16) for electron–phonon interactions; various kinds of phonons will be considered. In the following section, using formula (1.17), we will evaluate the transverse magnetoconductivity for electron–phonon collisions; using the same formula, we will finish this section by calculating the hopping conductivity in crystalline and amorphous materials. In the last section the oscillatory Hall effect will be dealt with. A study of the two (one)-dimensional quantized Hall effect will be undertaken in a future paper.

2. NO COLLISIONAL CURRENT; LONGITUDINAL MAGNETOCONDUCTIVITY

We consider the case where in addition to the electric field a magnetic field is present. In the Landau gauge the one-particle Hamiltonian reads

$$h^0 = (\mathbf{p} + e\mathbf{A})^2/2m, \quad \mathbf{A} = (0,Bx,0), \tag{2.1}$$

the magnetic field B being in the z direction. The one-particle eigenstates and eigenvalues are

$$|\zeta) = \phi_N(x + \lambda^2 k_y) e^{ik_y y} e^{ik_z z}/A^{1/2}, \tag{2.2}$$

$$\epsilon_\zeta \equiv \epsilon_{Nk_yk_z} = (N + 1/2)\hbar\omega_0 + \hbar^2 k_z^2/2m, \quad N = 0,1,2,\cdots, \tag{2.3}$$

where $\omega_0 = |q|B/m$ is the cyclotron frequency, ϕ_N represents the harmonic oscillator wave function, and λ is the radius of the Landau orbit, $\lambda^2 = \hbar/m\omega_0$ whose center is $x_0 = \hbar k_y/m^*\omega_0$, and A is the area L_yL_z, the linear dimensions being L_x, L_y, L_z. For crystals, we replace the mass m of the carriers (electrons) by the effective mass m^*. In the representation (2.2) the following matrix elements have been computed by Kahn and Frederikse[12]:

$$(\zeta|x|\zeta') = x^0\delta_{NN'}\delta_{kk'} + (\hbar/2m^*\omega_0)^{1/2}[(N+1)^{1/2}\delta_{N',N+1} + (N)^{1/2}\delta_{N',N-1}]\delta_{kk'}, \tag{2.4}$$

$$(\zeta|y|\zeta') = (L_y/2)\delta_{NN'}\delta_{kk'} = y^{eq}\delta_{NN'}\delta_{kk'}, \tag{2.5}$$

$$(\zeta|v_z|\zeta) = \hbar k_z/m^*, \tag{2.6}$$

$$(\zeta|v_x|\zeta') = i(\hbar\omega_0/2m^*)^{1/2} \times [-(N+1)^{1/2}\delta_{N',N+1} + (N)^{1/2})\delta_{N',N-1}]\delta_{kk'}, \tag{2.7}$$

$$(\zeta|v_y|\zeta') = (\hbar\omega_0/2m^*)^{1/2}[(N+1)^{1/2}\delta_{N',N+1} + (N)^{1/2}\delta_{N',N-1}]\delta_{kk'}, \tag{2.8}$$

where $\delta_{kk'} = \delta_{k_yk_y'}\delta_{k_zk_z'}$. As for the density of states $N(t)$, we find with

$$N(\epsilon) = \sum_{Nk_yk_z} \delta(\epsilon - \epsilon_{Nk_yk_z}) \tag{2.9'}$$

the result (spin included)

$$N(\epsilon) = \frac{\Omega\hbar\omega_0}{(2\pi)^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} \sum_N^{N_{\max}} \frac{1}{\sqrt{\epsilon - (N + \frac{1}{2})\omega_0}}, \tag{2.9}$$

where, in transforming the sum over k into an integral, we assumed periodic boundary conditions with limits for x, $-L_x/2$ and $L_x/2$, and for k_y, $-L_x/2\lambda^2$ and $L_x/2\lambda^2$; we also used the property of the delta function (2.22) (see below). $N_{\max}$ is the largest N for which the radicant is positive.

From (2.7) and (2.8) it is immediately seen that there is no diagonal ponderomotive current corresponding to the second term of (1.9) either in the x or in the y direction, but only in the z direction. We also note that

$$(\zeta|z - z_{eq}|\zeta) = 0, \tag{2.10}$$

which means that there is no collisional contribution to the current in the z direction, along which we will take the electric field in this section. We can now use, for the evaluation of the longitudinal magnetoconductivity formula (1.16) which applies to the case of "no collisional current," of LRTIII, i.e., ponderomotive current only. For the relaxation time $\tau(\epsilon_\zeta) = \tau(\epsilon_z)$ we need the transition rate $w_{\zeta\zeta'}$. It is given for electron–phonon interaction by[3,13]

$$w_{\zeta\zeta'} = \sum_{\mathbf{q}}[Q(\zeta,\mathbf{q}\to\zeta')\langle N_{\mathbf{q}}\rangle_{eq} + Q(\zeta\to\zeta',\mathbf{q})(1 + \langle N_{\mathbf{q}}\rangle_{eq}], \tag{2.11}$$

with

$$Q(\zeta,\mathbf{q}\to\zeta') = (2\pi/\hbar)|F(q)|^2|\mathscr{I}_{N,N'}|^2\delta_{k',k+q}\delta(\epsilon_\zeta - \epsilon_{\zeta'} + E_{\mathbf{q}}), \tag{2.12}$$

$$Q(\zeta\to\zeta',\mathbf{q}) = (2\pi/\hbar)|F(q)|^2|\mathscr{I}_{N',N}|^2\delta_{k',k-q}\delta(\epsilon_\zeta - \epsilon_{\zeta'} - E_{\mathbf{q}})$$

and

$$|\mathscr{I}_{NN'}|^2 = \frac{N'!}{N!}e^{-x}x^{N-N'}[L_N^{N-N'}(x)]^2, \quad N' \leqslant N, \tag{2.13}$$

$$|\mathscr{I}_{N'N}|^2 = \frac{N!}{N'!}e^{-x}x^{N'-N}[L_N^{N'-N}(x)], \quad N \leqslant N'.$$

The first term of (2.11) corresponds to the absorption of a phonon with wave vector $\mathbf{q}$ and energy $E_{\mathbf{q}}$, the second to the emission. The symbol $\delta_{k',k\pm q}$ is an abbreviation for $\delta_{k_y',k_y\pm q_y}\delta_{k_z',k_z\pm q_z}$ and the quantity x is equal to $(\lambda^2/\hbar^2)q_1^2 = (\lambda^2/\hbar^2)(q_x^2 + q_y^2)$. $\langle N_{\mathbf{q}}\rangle_{eq}$ is the equilibrium number of phonons and $F(\mathbf{q})$ is an interaction function depending on the model (deformation potential, etc.). All these quantities depend on the kind of phonons considered. Finally, $L_N^{N'-N}(x)$ is a Laguerre polynomial and N, N' are integers corresponding to Landau levels.

Since we want to evaluate the magnetoconductivity for nondegenerate carriers we can use the expression (1.14) to evaluate the relaxation time. In the lhs of (1.14), we substitute the expressions (2.6) and (2.11)–(2.13), and we carry out the sum over k_y' and k_z'. In addition, in the absorption term we put

$$N' - N = M, \quad M = 0,1,2,..., \tag{2.14a}$$

and in the emission term

$$N' - N = -M, \quad M = 0,1,2,.... \tag{2.14b}$$

For the arguments of the delta functions we use (2.3) and for the sum over $\mathbf{q}$, in cylindrical coordinates, we write

$$\sum_{\mathbf{q}} \to \frac{\Omega}{8\pi^3}\int d^3q = \frac{\Omega}{4\pi^2}\int dq_z \int_0^\infty dx. \tag{2.15}$$

We then find, with $\langle N_{\mathbf{q}} \rangle_{\text{eq}} = N_0$,

lhs of (1.14)

$$= -\frac{e}{\pi\hbar^2\lambda^2}\sum_M \int dq_z\, |F(q)|^2 q_z \times\Big\{ -N_0\delta\Big(q_z^2 + 2k_z q_z + \frac{2m^*}{\hbar^2}(M\hbar\omega_0 - E_q)\Big)\frac{N!}{(N+M)!}I_N\Big) + (1+N_0) \times\delta\Big(q_z^2 - 2k_z q_z - \frac{2m^*}{2}(M\hbar\omega_0 - E_q)\Big) \times\frac{(N-M)!}{N!}I_{N-M}\Big\}, \tag{2.16}$$

where

$$I_N = \int_0^\infty e^{-x} x^M [L_N^M(x)]^2\, dx \tag{2.17}$$

and where, in view of the approximations that will follow, we pulled the factors N_0, $1+N_0$, $|F(q)|^2$ and the delta functions in front of the integrals over x. For $M>0$ the above expression, upon using the values of the integrals I_N from the Appendix [cf. (A1) and (A2)], reduces to

lhs of (1.14)

$$= -\frac{e^2}{\pi\hbar^2\lambda^2}\sum_M \int dq_z\, |F(q)|^2 q_z \times\Big\{ -N_0\delta\Big(q_z^2 + 2k_z q_z + \frac{2m^*}{\hbar^2}(M\hbar\omega_0 - E_q)\Big) + \frac{(N-M)!}{N!}(1+N_0) \times\delta\Big(q_z^2 - 2k_z q_z - \frac{2m^*}{\hbar^2}(M\hbar\omega_0 - E_q)\Big)\Big\}. \tag{2.18}$$

As for the rhs of (1.14) from (2.6) we get

$$\text{rhs of (1.14)} = -\frac{1}{\tau(\epsilon_z)}\frac{e\hbar k_z}{\Omega m^*}. \tag{2.19}$$

To proceed further, we have to consider various kinds of phonons; we will consider longitudinal phonons in the deformation potential scheme.

A. Undamped case

In this subsection we will not consider any collision broadening effects; these effects will be dealt with in the next subsection.

1. Acoustical phonons

As usual, $E_{\mathbf{q}} \approx \hbar u_0 q$, u_0 being the sound velocity, and

$$|F(q)|^2 = c'q = (c^2/2\rho u_0)q, \tag{2.20}$$

where ρ is the density and c the deformation potential constant. Further, expanding the exponential in $N_0 = (e^{\beta E_{\mathbf{q}}} - 1)^{-1}$ up to third order, we find

$$|F(q)|^2 N_0 \approx \frac{c'}{\beta\hbar u_0}\frac{1}{1+\frac{1}{2}\beta\hbar u_0 q + \cdots}, \qquad |F(q)|^2(1+N_0) \approx \frac{c'}{\beta\hbar u_0}\frac{1}{1-\frac{1}{2}\beta\hbar u_0 q + \cdots}. \tag{2.21}$$

In (2.18) and (2.21) we take $q \approx q_z$. This approximation is justified as follows: in evaluating (1.16) we will need the density of states which diverges [see Eqs. (2.9) and (2.3)] for $K_z \to 0$, i.e., for small q_z; however, due to the factor $j_{z\zeta}j_{z\zeta}$ [see (1.16)] this divergence disappears. Moreover, since the electric field and the current are in the z direction, we expect the largest contribution to the current from processes involving large momentum transfer in the z direction, i.e., processes with large q_z and consequently small $q_\perp$. Actually the situation is more complex, but, as indicated by Barker[14] and by Peterson[15] for the nondegenerate case, the approximation is justified.

For the integral over q_z we use the following property of the delta function[16]:

$$\delta[g(x)] = \sum_j \frac{1}{|g'(x_j)|}\delta(x - x_j), \tag{2.22}$$

where $g'(x)$ is the derivative of $g(x)$ and x_j its zeros. We then easily find, upon using (2.19),

$$\frac{1}{\tau(\epsilon_z)} = \frac{\Omega m^* cc'}{e\hbar k_z \pi\hbar^2\lambda^2\alpha_1}\sum_M\Big\{\frac{1}{(\alpha_2^2 - 2\alpha_3 M)^{1/2}} \times\frac{1}{1-\alpha_1-\alpha_2+\frac{1}{2}\alpha_1^2\alpha_3 M}\Big\}\frac{(N-M)1!}{N!} \times\frac{1}{(\alpha_2^2+2\alpha_3 M)^{1/2}}\frac{1}{1+\alpha_1\alpha_2-\frac{1}{2}\alpha_1^2}\Big\} \times(\alpha_2+\alpha_1\alpha_3 M), \tag{2.23}$$

where $\alpha_1 = \beta\hbar u_0$, $\alpha_2 = k_z - m^* u_0/\hbar$, and $\alpha_3 = m^*\omega_0/\hbar$. This expression for $\tau(\epsilon_z)$ is too complicated to substitute in (1.16). If we keep only the linear term in q in the expansion for N_0, i.e., if we take $N_0 \approx 1 + N_0 \approx (\alpha_1 q)^{-1}$, then we find

$$\frac{1}{\tau(\epsilon_z)} = \frac{2\Omega c'\alpha_3\alpha_2}{\pi\hbar^3\alpha_1 k_z}\sum_M\Big\{\frac{1}{(\alpha_2^2-2\alpha_3 M)^{1/2}} + \frac{(N-M)!/N!}{(\alpha_2^2+2\alpha_3 M)^{1/2}}\Big\}. \tag{2.24}$$

In the quantum limit $M=0$ and (2.18) is greatly simplified; upon substituting the values of α_1, α_2,α_3 we find ($c = E_1/\Omega^{1/2}$)

$$\frac{1}{\tau(\epsilon_z)} = \frac{1}{2\pi}\Big(\frac{2m^*}{\hbar^2}\Big)^{3/2}\frac{E_1^2 kT}{\hbar\rho u_0^2}\hbar\omega_0\epsilon_z^{-1/2}, \tag{2.25}$$

where $\epsilon_z = \hbar^2 k_z^2/2m^*$. This result remains the same even if the collisions are elastic; it is easily seen by putting $u_0 = 0$ only in α_2, appearing in (2.24). In the literature[15,17] we find the result (2.25) to within a constant: $1/2\pi$ is replaced by $1/4\pi$.

If the collisions are elastic, $E_{\mathbf{q}} \approx 0$ and for $M>0$, we find

$$\frac{1}{\tau(\epsilon_z)} = \frac{1}{2\pi}\Big(\frac{2m^*}{\hbar^2}\Big)^{3/2}\frac{E_1^2 kT}{\hbar\rho u_0^2} \times\hbar\omega_0\sum_N[\epsilon_\zeta - (N+\tfrac{1}{2})\hbar\omega_0]^{-1/2}. \tag{2.26}$$

This result is different from that of Argyres[18] only in that $1/2\pi$ replaces $1/4\pi$.

2. Piezoelectrical phonons

The only difference from the procedure of Sec. 2A1 up to (2.24) is that $|F(q)|^2 = P/q$, where P is a piezoelectric constant; we find

$$\frac{1}{\tau(\epsilon_z)} = \frac{m^*\Omega P\alpha_2}{\pi\hbar^3\alpha_1 k_z}\sum_M \frac{1}{M}\left\{\frac{1}{(\alpha_2^2 - 2\alpha_3 M)^{1/2}} - \frac{(N-M)!/N!}{(\alpha_2^2 + 2\alpha_3 M)^{1/2}}\right\}, \quad M \neq 0. \tag{2.27}$$

3. Optical phonons

As usual, we take $E_q \approx E = \text{const}$ and

$$|F(q)|^2 = \frac{\hbar^2 D^2}{2\Omega\rho E} = D', \tag{2.28}$$

where D is a constant; we then find

$$\frac{1}{\tau(\epsilon_z)} = \frac{\hbar D^2}{8\pi\rho E}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\hbar\omega_0\sum_M\left\{\frac{N_0}{\sqrt{\epsilon_z - (M\omega_0 - E)}} + \frac{(N-M)!}{N!}\frac{1+N_0}{\sqrt{\epsilon_z + (M\hbar\omega_0 - E)}}\right\}. \tag{2.29}$$

If the phonon energy is such that $N - M \gg 1$, then $(N-M)!/N!$ may be approximated by 1 and (2.29) becomes identical with the result of Peterson[15] and of Kharus and Tsidil'kovskii.[19] This approximation, however, is redundant in the quantum limit, where $M = 0$.

4. Polar optical phonons

The only difference from Sec. 2A3 is that

$$|F(q)|^2 = \frac{A}{q^2} \approx \frac{A}{q_z^2}, \tag{2.30}$$

where A is a constant; the result is

$$\frac{1}{\tau(\epsilon_z)} = \frac{\Omega A m^{*1/2}}{2\sqrt{2}\pi\hbar^2}\hbar\omega_0\sum_M\frac{1}{M\hbar\omega_0 - E}\times\left\{\frac{-N_0}{\sqrt{\epsilon_z - (M\hbar\omega_0 - E)}} + \frac{(N-M)!}{N!}\frac{1+N_0}{\sqrt{\epsilon_z + (M\hbar\omega_0 - E)}}\right\}; \tag{2.31}$$

this expression is much simpler than that given by Kharus and Tsidil'kovskii.[19]

We now go back to (1.16) and transform the sum into an integral; using the density of states (2.9), we find, with $\epsilon_z = \epsilon - (N + 1/2)\hbar\omega_0$,

$$\sigma_{zz} = -\frac{e^2}{2\pi^2 m^*}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\hbar\omega_0 \times\int_{\hbar\omega_0/2}^{\infty}\frac{\sum_N[\epsilon_\xi(N+1/2)\hbar\omega_0]^{1/2}}{1/\tau(\epsilon_z)}\frac{\partial\langle n_\xi\rangle}{\partial\epsilon_\xi}d\epsilon_\xi. \tag{2.32}$$

To proceed further, we have to substitute the various relaxation times. (2.32) is identical with the results of Argyres[18] and of Peterson.[15] For acoustical phonons these authors have carried out the complete calculation with $1/\tau(\epsilon_z)$ given by (2.26); for optical phonons the calculation has been done by Peterson[15] and by Kharus and Tsidil'kovskii[19] with $(N-M)!/N! \approx 1$ in (2.29). Finally, an approximate result was given for polar optical phonons.[19]

B. Collision broadening

As indicated by formula (2.26), the experimentally observed oscillations of the magnetoresistance are due to the oscillating relaxation times [see the various expressions for $1/\tau(\epsilon_z)$]. Note also that the relaxation time due to simultaneous scattering processes is given, as usual, by

$$\frac{1}{\tau(\epsilon_z)} = \sum_i\frac{1}{\tau_i(\epsilon_z)}, \tag{2.33}$$

where i indicates a scattering process, e.g., scattering by optical phonons, impurities, etc. However, the experimentally observed broadening of the oscillations of the magnetoresistance is not shown nor suggested by the above relaxation time formulas. A simple and heuristic way to include broadening in the calculations is to replace the delta functions, appearing in (2.18) by Lorentzians of constant width Γ and shift Δ. In general, Γ and Δ depend on the energy but the results one arrives at are similar.[20] Thus, using the representation of the delta function

$$\delta(\epsilon_k - \epsilon) = \lim_{s\to 0^+}\frac{1}{\pi}\operatorname{Im}(\epsilon_k - \epsilon - is)^{-1} \tag{2.34}$$

and letting $\epsilon_k \to \epsilon_k + \Delta - i\Gamma$, the above-mentioned Lorentzian has the standard form

$$\frac{1}{\pi}\frac{\Gamma}{(\epsilon_k - \epsilon + \Delta)^2 + \Gamma^2}. \tag{2.35}$$

1. Acoustical phonons

We take $E_q \approx \hbar u_0 q_z$ and $N_0 \approx 1 + N_0 \approx (\beta\hbar u_0 q)^{-1}$; on account of (2.35), we get from (2.18)

lhs of (1.14)

$$= -\frac{2Bc'\Gamma}{\pi^2\hbar^3\lambda^2 u_0\beta}\sum_M\int_{-\infty}^{\infty}dq_z\, q_z \times\frac{1}{[(2Bk_z - \hbar u_0)q_z + Bq_z^2 + C']^2 + \Gamma^2} + \frac{(N-M)!/N!}{[-(2Bk_z - \hbar u_0)q_z + Bq_z^2 - C']^2 + \Gamma^2}, \tag{2.36}$$

where $B = \hbar^2/2m^*$ and $C' = M\hbar\omega_0$; for simplicity we took $\Delta = 0$. In the absorption term we put $y = q_z + (k_z - \hbar u_0/2B)$, in the emission term $y = q_z - (k_z - \hbar u_0/2B)$. The integrals over y can be found from tables,[21] and the result for the absorption term is

$$\int_{-\infty}^{\infty}dq_z\, q_z\{\cdots\} = -\left(k_z - \frac{\hbar u_0}{2\beta}\right)\frac{\pi}{2B\Gamma}\operatorname{Re}\left\{E' + i\frac{\Gamma}{B}\right\}^{-1/2}, \tag{2.37}$$

where $E' = (k_z - \hbar u_0/2B)^2 - C'/B$; for the emission term we find (2.37) multiplied by $-(N-M)!/N!$. Combining this with (2.19) we find for the relaxation time

$$\frac{1}{\tau(\epsilon_z)} = \frac{m^* c'\Omega}{2\pi\hbar^4\lambda^2 u_0\beta}\frac{k_z - \hbar u_0/2B}{k_z}\times\sum_M\left(1 + \frac{(N-M)!}{N!}\right)\operatorname{Re}\left\{E' + i\frac{\Gamma}{B}\right\}^{-1/2}. \tag{2.38}$$

2. Piezoelectrical phonons

The procedure is identical; using the approximations of Sec. 2B1, we find

$$\frac{1}{\tau(\epsilon_z)} = \frac{m^*P\Omega}{2\pi\hbar^4\lambda^2 u_0\beta}\,\frac{1}{k_z(k_z-\hbar u_0/2B)} \times\sum_M\left(1+\frac{(N-M)!}{N!}\right)\mathrm{Re}\left\{E'+i\frac{\Gamma}{B}\right\}^{-1/2}. \quad (2.39)$$

3. Optical phonons

Using the approximations of Sec. 2A, we find

$$\frac{1}{\tau(\epsilon_z)} = \frac{m^*D'\Omega}{2\pi\hbar^3\lambda^2}\sum_M\left[N_0+\frac{(N-M)!}{N!}(1+N_0)\right] \times\mathrm{Re}\left\{E''+i\frac{\Gamma}{B}\right\}^{-1/2}, \quad (2.40)$$

where $E''=k_z^2-C/B$, $C=M\hbar\omega_0-E$. (2.34) is in agreement with the result of Barker[20] except for the factor $(N-M)!/N!$, if it is approximated by 1 then, as shown by Barker,[20] who used Poisson's formula for the sum $\sum_M \mathrm{Re}\{E''+i\Gamma/B\}^{-1/2}$, (2.40) leads to a good agreement with experimental results of Stradling and Wood.[22]

4. Polar optical phonons

As in Sec. 2A4, we easily find ($q^2\approx q_z^2$)

$$\frac{1}{\tau(\epsilon_z)} = \frac{m^*A\Omega}{2\pi\hbar^3\lambda^2}\,\frac{1}{k_z^2}\sum_M\left[N_0+\frac{(N-M)!}{N!}(1+N_0)\right] \times\mathrm{Re}\left\{E''+i\frac{\Gamma}{B}\right\}^{-1/2}. \quad (2.41)$$

3. COLLISIONAL CURRENT

We are interested in the electrical conductivity in crystalline materials in the presence of a magnetic field perpendicular to the electric field which is taken in the x direction. As stated at the beginning of the previous section, there is no diagonal ponderomotive current [see (2.7)]; there is only a diagonal collisional current as is seen from (2.4) and formula (1.17) for the transverse magnetoconductivity $[X_\zeta=(\zeta|x|\zeta)]$.

The situation is the same, i.e., there is only collisional current, in the case of the hopping conductivity because, as indicated by Mott and Davis,[23] the diagonal matrix element of the velocity operator between localized states vanishes. Thus, in both cases formula (1.17) applies. In what follows we will evaluate the transverse magnetoconductivity in crystalline materials for scattering by phonons (longitudinal phonons only in the deformation potential scheme) and the phonon assisted hopping conductivity in crystalline and amorphous materials.

A. Crystalline materials, Landau states

The Landau states and eigenvalues are given by (2.2) and (2.3). The transition rate $w_{\zeta\zeta'}$ is given by (2.11)–(2.13). From (2.4) it is seen that

$$X_\zeta=(\zeta|x|\zeta)=-\lambda^2k_y. \quad (3.1)$$

This means that the factor $(X_\zeta-X_{\zeta'})^2$, appearing in (1.17), varies like $(k_y-k_y')^2=q_y^2$, due to the Kronecker deltas of (2.12). When (2.11)–(2.13) and (3.1) are substituted in (1.17), it is easily seen that the only dependence of σ_{xx} on k_y, k_y' comes from the factor $(X_\zeta-X_{\zeta'})^2$, i.e., $\sigma_{xx}\sim q_y^2$. But due to the cylindrical symmetry σ_{yy} must vary like q_x^2. Thus, $\sigma_{xx}=\frac{1}{2}(\sigma_{xx}+\sigma_{yy})$ will depend on $\frac{1}{2}(q_x^2+q_y^2)=\frac{1}{2}q_\perp^2$ (see Ref. 5). Carrying out the summation over k_y', we then get from (1.17) ($\langle n_\zeta\rangle_{\mathrm{eq}}=f_\zeta$)

$$\sigma_{xx} = \frac{\pi e^2\beta\lambda^2}{\hbar\Omega} \times\sum_{N,N'}\sum_{k_y,k_z,k_z',\mathbf{q}} x|F(q)|^2|\mathscr{I}_{N,N'}|^2 f_{N,k_z}(1-f_{N',k_z'}) \times\left[N_0\delta(\epsilon_{N,k_z}-\epsilon_{N',k_z'}+E_{\mathbf{q}})\delta_{k_z',k_z+q_z} +(1+N_0)\delta(\epsilon_{N,k_z}-\epsilon_{N',k_z'}-E_{\mathbf{q}})\delta_{k_z',k_z-q_z}\right]; \quad (3.2)$$

x, as before, is equal to $\lambda^2q_\perp^2/2$. As in the previous section, we will consider various kinds of phonons.

1. Acoustical phonons

a. Inelastic collisions: We take $N_0\approx 1+N_0\approx(\beta\hbar u_0q)^{-1}$ and use (2.20); further, we take

$$E_{\mathbf{q}}\approx\hbar u_0q\approx u_0q_\perp; \quad (3.3)$$

this approximation is based on the fact that the density of states (2.9) diverges for $k_z\to 0$, which means that the largest contribution to σ_{xx} comes from processes involving small momentum transfer q_z in the direction of the applied magnetic field.

We now change $-q_z$ to q_z in the second term of (3.2) and use (2.22) for the integrals over q_z; besides, to arrive at a "reasonable" integral over x, we neglect, in view of (3.3), q_z in the factor $f_{N',k_z'}$. Carrying out the sum over k_z' and the integral over q_z, we find

$$\sigma_{xx} = \frac{m^*e^2c'}{2\pi\hbar^4u_0}\sum_{N,N'}\sum_{k_y,k_z}f_{N,k_z}(1-f_{N',k_z})\int_0^\infty dx\,x|\mathscr{I}_{N,N'}|^2 \times\left\{\frac{1}{(k_z^2-2M/\lambda^2+2m^*u_0q_\perp/\hbar)^{1/2}} +\frac{1}{(k_z^2+2M/\lambda^2-2m^*u_0q_\perp/\hbar)^{1/2}}\right\}, \quad (3.4)$$

where we use (2.14). The summations over N, N' extend up to those values of N, N' for which the radicants are positive.

In (3.4), we put $q_\perp=(2x/\lambda^2)^{1/2}$, $c_a=k_z^2-2M/\lambda^2$, $c_e=k_z^2+2M/\lambda^2$ and $b=(2m^*u_0/\hbar)(2/\lambda^2)^{1/2}$. The ratio $b/c_{a,e}$ is $u_0(2m^*\hbar\omega_0)^{1/2}/\epsilon_z\mp M\hbar\omega_0$, i.e., much smaller than 1 for $M=0$; for $M\neq 0$ we assume that $\epsilon_z-M\hbar\omega_0$ is such that b/c_a is smaller than 1. Using then the expansion

$$(1+y)^{-1/2}=1-\tfrac{1}{2}y+\tfrac{3}{8}y^2-\tfrac{15}{16}y^3+\cdots,\quad y<1, \quad (3.5)$$

we can write (3.4) in the form

$$\sigma_{xx} = \frac{m^*e^2c'}{2\pi\hbar^4u_0}\sum_{N,N'}\sum_{k_y,k_z}f_{N,k_z}(1-f_{N',k_z}) \times\left\{c_a^{-1/2}\int_0^\infty dx \times\left(1-\frac{1}{2}\frac{b}{c_a}x^{1/2}+\frac{3}{8}\left(\frac{b}{c_a}\right)^2x+\cdots\right)x|\mathscr{I}_{N,N'}|^2 +c_e^{-1/2}\int_0^\infty dx\right.$$

$$\times\left(1+\frac{1}{2}\frac{b}{c_e}x^{1/2}+\frac{3}{8}\left(\frac{b}{c_a}\right)^2x+\cdots\right)x|\mathscr{I}_{N,N'}|^2\Big\}. \tag{3.6}$$

The integrals over x are hard to evaluate in closed expressions: they can be performed either numerically or by means of the hypergeometric function.[21] This is necessary only for the $x^{m/2}$ terms with m odd; for m even, as shown in the Appendix, the integrations can be done exactly. If we assume that the magnetic field is so strong that the phonons cannot induce transitions between the various levels N, N', i.e., if we can take $N\approx N'$, then $M\approx 0$, $c_a=c_e=k_z^2$, the terms $x^{m/2}$ with m odd cancel and (3.6) becomes

$$\sigma_{xx}=\frac{m^*e^2c'}{\pi\hbar^4u_0}\sum_N\sum_{k_y,k_z}f_{N,k_z}(1-f_{N,k_z})|c_a^{-1/2}|\int_0^\infty dx$$
$$\times\left(1+\frac{3}{8}\left(\frac{b}{c_a}\right)^2x+\frac{35}{128}\left(\frac{b}{c_a}\right)^4x^2+\cdots\right)x|\mathscr{I}_{N,N}|^2. \tag{3.7}$$

The integrals over x are given in the Appendix, and we will consider only three terms in the expansion.

(i) Degenerate case: Since $f_{N,k_z}=\langle n_{N,k_z}\rangle_{\rm eq}$ is given by the Fermi–Dirac distribution function, we make the usual approximation

$$\beta f_{N,k_z}(1-f_{N,k_z})\approx\delta(\epsilon_z-\epsilon_F'),\quad \epsilon_F'=\epsilon_F-(N+\tfrac{1}{2})\hbar\omega_0. \tag{3.8}$$

With (3.8), the density of states (2.9) and the results for the integrals over x we find [cf. (A6), (A7), (A9)]:

$$\sigma_{xx}=2\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho u_0^2\hbar^4\beta}$$
$$\times\sum_N\left\{\frac{2N+1}{\epsilon_F'}+\frac{3b'^2}{4}\frac{3N^2+3N+1}{\epsilon_F'^3}\right.$$
$$\left.+\frac{35}{64}b'^4\frac{(2N+1)(5N^2+5N+3)}{\epsilon_F'^5}+\cdots\right\}, \tag{3.9}$$

where $b'=(\hbar^2b/2m^*)=u_0(2m^*\hbar\omega_0)^{1/2}$. In the quantum limit $N=0$ and the first term of (3.9) gives twice the result of Kubo *et al.*[24]

(ii) Nondegenerate case: As usual, we take

$$f_{N,k_z}(1-f_{N,k_z})\approx f_{N,k_z}=e^{\beta(\epsilon_F'-\epsilon_z)}; \tag{3.10}$$

we then find

$$\sigma_{xx}=2\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho u_0^2\hbar^4}\sum_N e^{\beta\epsilon_F'}\int_0^\infty d\epsilon_z\,e^{-\beta\epsilon_z}$$
$$\times\left\{\frac{2N+1}{\epsilon_z}+\frac{3}{4}b'^2\frac{3N^2+3N+1}{\epsilon_z^3}\right.$$
$$\left.+\frac{35}{64}b'^4\frac{(2N+1)(5N^2+5N+3)}{\epsilon_z^5}+\cdots\right\}. \tag{3.11}$$

From tables of integrals,[21] we find

$$\int_0^\infty x^{\nu-1}(x+c)^{-\rho}e^{-\mu x}\,dx$$
$$=\frac{1}{\sqrt{\pi}}\left(\frac{c}{\mu}\right)^{\nu-1/2}e^{c\mu}\Gamma(\nu)K_{1/2-\nu}\left(\frac{c\mu}{2}\right) \tag{3.12}$$
$$[|\arg c|<\pi,\ \mathrm{Re}\,\mu>0,\ \mathrm{Re}\,\nu>0],$$

where $\Gamma(\nu)$ is the gamma function and $K_\nu(x)$ the modified Bessel function. The integrals appearing in (3.11) can be written as

$$I=\int_0^\infty d\epsilon_z\,\frac{e^{-\beta\epsilon_z}}{\epsilon_z^m}$$
$$=\lim_{c\to 0}\int_0^\infty d\epsilon_z\,\frac{e^{-\beta\epsilon_z}}{\epsilon_z^{1/2}(\epsilon_z+c)^{m-1/2}},\quad m=1,3,5,\cdots. \tag{3.13}$$

With $\Gamma(\frac{1}{2})=\sqrt{\pi}$, $K_0(x)\sim\ln(2/\gamma x)$ ($\gamma=1.781$), and (3.12) we get from (3.11) (we don't take the limit $c\to 0$, though $c\ll 1$)

$$\sigma_{xx}\approx 2\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho u_0^2\hbar^4}\sum_N e^{\beta[\epsilon_F-(N+1/2)\hbar\omega_0]}\ln\left(\frac{4}{\gamma c\beta}\right)$$
$$\times\{(2N+1)+\tfrac{3}{4}b'^2(3N^2+3N+1)$$
$$+\tfrac{35}{64}b'^4(2N+1)(5N^2+5N+1)+\cdots\}. \tag{3.14}$$

As in (i) the first term of (3.14) in the quantum limit ($N=0$) gives twice the result of Kubo *et al.*,[24] if we take $c=(2e/\gamma)^{1/2}(\hbar u_0/\lambda)\approx 1.23b$; the other terms represent "corrections." The summation over N can easily be performed but the result is complicated. Assuming $N_{\max}\to\infty$, we find for the first term[6]

first term of (3.14)

$$\approx\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\hbar^4\rho u_0^2}e^{\beta\epsilon_F}\frac{\coth(\beta\hbar\omega_0/2)}{\sinh(\beta\hbar\omega_0/2)}\ln\left(\frac{4}{\gamma c\beta}\right). \tag{3.15}$$

b. Elastic collisions: If we repeat the calculations with $E_{\mathbf{q}}\approx 0$, then, instead of (3.4), we find

$$\sigma_{xx}=\frac{m^*e^2c'}{2\pi\hbar^4u_0}\sum_{N,N'}\sum_{k_y,k_z}f_{N,k_z}(1-f_{N,k_z})\int_0^\infty dx\,x|\mathscr{I}_{N,N'}|^2$$
$$=\left\{\frac{1}{(k_z^2-2M/\lambda^2)^{1/2}}+\frac{1}{(k_z^2+2M/\lambda^2)^{1/2}}\right\}. \tag{3.16}$$

In arriving at (3.16) we replaced $1-f_{N',k_z'}$ by $1-f_{N,k_z}$ due to the delta function $\delta(\epsilon_{N,k_z}-\epsilon_{N',k_z'})$. The result for the integral over x is found in the Appendix [see (A4)]; with that (3.16) becomes

$$\sigma_{xx}=\frac{m^*e^2c'}{2\pi\hbar^4u_0}\sum_{N,M}\sum_{k_y,k_z}f_{N,k_z}(1-f_{N,k_z})$$
$$\times\left\{\frac{2N+M+1}{(k_z^2-2M/\lambda^2)^{1/2}}+\frac{2N-M+1}{(k_z^2+2M/\lambda^2)^{1/2}}\right\}. \tag{3.17}$$

(i) Degenerate case: With (3.8) and (2.9), (3.17) becomes

$$\sigma_{xx}=\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho u_0^2\hbar^4\beta}\sum_{N,M}\frac{1}{\sqrt{\epsilon_F'}}$$
$$\times\left\{\frac{2N+M+1}{\sqrt{\epsilon_F'-M\hbar\omega_0}}+\frac{2N-M+1}{\sqrt{\epsilon_F'+M\hbar\omega_0}}\right\}. \tag{3.18}$$

Note that when $M=0$, (3.18) becomes the first term of (3.9).

(ii) Nondegenerate case: (3.10), (2.9), and (3.16) give

$$\sigma_{xx}=\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho' u_0^2\hbar^4}\sum_{N,M}e^{\beta\epsilon_F'}\int_0^\infty d\epsilon_z$$

$$\times \frac{e^{-\beta\epsilon_z}}{\sqrt{\epsilon_z}} \left\{ \frac{2N+M+1}{\sqrt{\epsilon_z - M\hbar\omega_0}} + \frac{2N-M+1}{\sqrt{\epsilon_z + M\hbar\omega_0}} \right\}. \quad (3.19)$$

The integral for the emission form is given by (3.12); the integral over the absorption term can easily be evaluated with a change of variables, $x = \epsilon_z - M\hbar\omega_0$. The result is

$$\sigma_{xx} = \frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^3\rho u_0^2\hbar^4} \sum_{N,M} e^{[\epsilon_F - (N+1/2)\hbar\omega_0]} \times \{(2N+M+1)e^{-\alpha M} + (2N-M+1)e^{\alpha M}\}K_0(\alpha M), \quad (3.20)$$

where $\alpha = \beta\hbar\omega_0/2$. The sum over M can also be performed.[6] For $M = 0$, we find the first term of (3.14).

2. Optical phonons

From (2.28), (3.2), and (A4) of the Appendix, we get

$$\sigma_{xx} = \frac{m^{*2}e^2D'\Omega\omega_0\beta}{4\pi^3\hbar^4} \sum_{N,N'} \int d\epsilon_z \; f_{N,k_z}(1 - f_{N',k_z}) \times \left\{ \frac{N_0(2N+M+1)}{\sqrt{\epsilon_z}\sqrt{\epsilon_z + E - M\hbar\omega_0}} + \frac{(1+N_0)(2N-M+1)}{\sqrt{\epsilon_z}\sqrt{\epsilon_z - (E - M\hbar\omega_0)}} \right\}. \quad (3.21)$$

(i) Degenerate case: If N and N' are large, we may take $N' \approx N$ in f_{N',k_z}; then, upon using (3.8), we find

$$\sigma_{xx} = \frac{m^{*2}e^2D'\Omega\omega_0\beta}{4\pi^3\hbar^4} \sum_{N,N'} \left\{ \frac{N+N'+1}{\sqrt{\epsilon_F - (N+1/2)\hbar\omega_0}} \times \frac{N_0}{\sqrt{\epsilon_F - (N+1/2)\hbar\omega_0 + E}} + \frac{1+N_0}{\sqrt{\epsilon_F - (N+1/2)\hbar\omega_0 - E}} \right\}. \quad (3.22)$$

(ii) Nondegenerate case: If we use (3.10), the integrals over ϵ_z take the same form as those encountered previously [see (3.12) and (3.19)]; they can be easily evaluated.[6] The result is

$$\sigma_{xx} = \frac{m^{*2}e^2D'\Omega\beta\omega_0}{4\pi^3\hbar^4} \sum_{N,M} e^{\beta[\epsilon_F - (N+1/2)\hbar\omega_0]} \times \{N_0(2N+M+1)e^z + (1+N_0)(2N-M+1)e^{-z}\}K_0(|z|), \quad (3.23)$$

where $z = \beta(E - M\hbar\omega_0)/2$. For $z \to 0$, i.e., when $M \to \omega_L/\omega_0$, ω_L being the phonon frequency, we get a logarithmic divergence, since $K_0(|z|) \approx -\ln(|z|)$.

3. Polar optical phonons

The procedure is identical with that of (3.12), the only differences being the use of (2.30) with $q^2 \approx q_\perp^2$ instead of (2.28) and the use of (A1) and (A2) for the integrals over x

(i) Degenerate case: We find

$$\sigma_{xx} = \frac{m^*e^2A\Omega}{(2\pi\hbar)^3} \sum_{N,N'} \frac{1}{\sqrt{\epsilon_F - (N+1/2)\hbar\omega_0}} \times \left\{ \frac{N_0}{\sqrt{\epsilon_F - (N'+1/2)\hbar\omega_0 + E}} + \frac{N'!}{N!} \frac{1+N_0}{\sqrt{\epsilon_F - (N'+1/2)\hbar\omega_0 - E}} \right\}. \quad (3.24)$$

The quantum limit version of (3.24) ($N = N' = 0$) is much simpler than the result of Efros.[25] For $N'!/N! \approx 1$ we can perform the summations over N and N' according to a method by Adams and Holstein.[7] Writing $\epsilon_F/\hbar\omega_0 = I + \delta + 1/2$, where I is an integer and where $0 \leqslant \delta \leqslant 1$, we have

$$\sum_N^I \frac{1}{\sqrt{\epsilon_F - (N+1/2)\hbar\omega_0}} = \frac{1}{\sqrt{\hbar\omega_0}} \left\{ 2\left(\frac{\epsilon_F}{\hbar\omega_0}\right)^{1/2} + \phi \right\}, \quad (3.25)$$

$$\sum_{N'}^{I'} \frac{1}{\sqrt{\epsilon_F - (N'+1/2)\hbar\omega_0 \pm E}} = \frac{1}{\sqrt{\hbar\omega_0}} \left\{ 2\left(\frac{\epsilon_F - E}{\hbar\omega_0}\right)^{1/2} + \phi \right\},$$

where $\phi = -2(\frac{1}{2} + \delta)^{1/2} + \delta^{-1/2} + (1+\delta)^{-1/2}$. For $\epsilon_F \gg E$, the two sums become identical and (3.24) takes the simple form

$$\sigma_{xx} = \frac{m^*e^2A\Omega}{(2\pi\hbar)^3\hbar\omega_0} (2N_0+1) \left\{ 2\left(\frac{\epsilon_F}{\hbar\omega_0}\right)^{1/2} + \phi \right\}^2. \quad (3.26)$$

This expression diverges for $\delta \to 0$. Efros,[25] following a different procedure, neglects the third term in ϕ and finds a logarithmic divergence:

$$\sigma_{xx} \propto \ln \delta(1-\delta).$$

(ii) Nondegenerate case: The result has been obtained by Charbonneau[5] using the same formalism and is

$$\sigma_{xx} = \frac{m^*e^2A\Omega\beta}{(2\pi\hbar)^3} \sum_{N,M} e^{\beta[\epsilon_F - (N+1/2)\hbar\omega_0]} \times \left[N_0e^z + \frac{(N-M)!}{N!}(1+N_0)e^{-z} \right] K_0(|z|), \quad (3.27)$$

with z, as in Sec. 3A2, equal to $\beta(E - M\hbar\omega_0)/2$. Equation (3.27) exhibits the same logarithmic divergence, obtained in the literature,[26,20] as (3.23).

4. Collision broadening

The unpleasant logarithmic divergences of the previous subsection [cf. (3.20), (3.23), (3.27)] disappear by taking into account collision broadening of the energy levels or by going beyond the Van Hove limit (i.e., the first Born approximation). Collision-broadening effects are more important at high temperatures, and we will consider them, in what follows, in the simple way of Sec. 2B only for the nondegenerate case.

a. Optical phonons: We start from (3.2) in which we change $-q_z$ to q_z in the emission term and perform the sum over k_z'. We substitute (2.28) and use (3.10). The delta functions are replaced by the Lorentzians (2.35) and for the integrals over x we use (A4); with the density of states (2.9) we find

$$\sigma_{xx} = \frac{e^2\beta D'\Omega\Gamma}{4\pi^4\hbar\lambda^2} \sum_{N,M} e^{\beta[\epsilon_F - (N+1/2)\hbar\omega_0]} \int_{-\infty}^{\infty} dk_z e^{-\beta B k_z^2}$$

$$\times\int_{-\infty}^{\infty} dq_z \frac{N_0(2N+M+1)+(1+N_0)(2N-M+1)}{(2Bq_zk_z+Bq_z^2+C)^2+\Gamma^2} \tag{3.28}$$

with $B=\hbar^2/2m^*$ and $C=\pm M\hbar\omega_0\mp E-\Delta$, the upper signs corresponding to absorption, the lower to emission. The integral over q_z is transformed by $y=q_z+k_z$ and the result, apart from the constants, is[21]

$$I(k_z)=(\pi/2B\Gamma)\,\mathrm{Re}\{E'+i\Gamma/B\}^{-1/2}, \tag{3.29}$$

where $E'=k_z^2-C/B$. Then, the integral over k_z, with $k_z^2=x$, becomes

$$I=\mathrm{Re}\int_0^{\infty} dx\left\{x-\frac{(C+i\Gamma)}{B}\right\}^{-1/2} e^{-\beta Bx}x^{-1/2}$$
$$=\mathrm{Re}\exp\left[\frac{\beta}{2}(-C+i\Gamma)\right]K_0\left[\frac{\beta}{2}(-C+i\Gamma)\right]. \tag{3.30}$$

The result for σ_{xx} is

$$\sigma_{xx}=\frac{e^2\beta D'\Omega m^*\omega_0}{4\pi^3\hbar^4}\sum_{N,M} e^{\beta[\epsilon_F-(N+1/2)\hbar\omega_0]}$$
$$\times\{N_0(2N+M+1)+(1+N_0)(2N-M+1)\}$$
$$\times\mathrm{Re}\exp\left[\frac{\beta}{2}(-C+i\Gamma)\right]K_0\left[\frac{\beta}{2}(-C+i\Gamma)\right]. \tag{3.31}$$

b. Polar optical phonons: Following the steps of Sec. 3A4a, we find, with (2.30) and (A1), (A2),

$$\sigma_{xx}=\frac{m^*e^2\beta A\Omega}{(2\pi\hbar)^3}\sum_{n,M} e^{\beta[\epsilon_F-(N+1/2)\hbar\omega_0]}$$
$$\times\left\{N_0+\frac{(N-M)!}{N!}(1+N_0)\right\}$$
$$\times\mathrm{Re}\exp\left[\frac{\beta}{2}(-C+i\Gamma)\right]K_0\left[\frac{\beta}{2}(-C+i\Gamma)\right]. \tag{3.32}$$

This expression is essentially the same as that given by Barker.[14] To make a comparison with (3.27) we define[26] $\delta^P=\omega_L/N_0{}^P$, where ω_L is the phonon frequency and P the largest integer contained in ω_L/ω_0. The divergent terms of (3.27) are those with $P=M$ and $\delta^P\to 0$ or $P=M-1$ and $\delta^P\to 1$. But for these values the expression (3.32) does not diverge. With $K_0(x)\approx -\ln x$, $x\ll 1$, (3.32) behaves, its oscillatory part represented by the divergent terms as ($\delta^P=\delta$)

$$\ln|\tfrac{1}{2}\beta(-C+i\Gamma)|=\ln\{(\beta\hbar\omega_0/2)[\delta^2+(\Gamma/\hbar\omega_0)^2]\}. \tag{3.33}$$

The same remark applies also to (3.31) when compared with (3.23). An harmonic analysis of (3.33), using the Poisson sum formula, done by Barker[14] leads to a good agreement with the experimental results.[22,27]

c. Acoustical phonons (elastic collisions): The calculations for inelastic collisions become very complicated, therefore we limit ourselves to elastic collisions. Proceeding as above, we find ($E_q\approx 0$)

$$\sigma_{xx}=\frac{m^{*2}e^2E_1^2\omega_0}{(2\pi)^2\rho u_0^2\hbar^4}\sum_{N,M} e^{\beta[\epsilon_F-(N+1/2)\hbar\omega_0]}(2N+1)$$
$$\times\,\mathrm{Re}\exp\left[\frac{\beta}{2}(-C+i\Gamma)\right]K_0\left[\frac{\beta}{2}(-C+i\Gamma)\right], \tag{3.34}$$

where $C=\pm M\hbar\omega_0-\Delta$. The divergence of (3.20) disappears again. The sum over N can be performed.[6]

B. Hopping conductivity

In doped semiconductors at low temperatures the number of the ionized impurities and the density of carriers in the conduction band become very small. However, transport effects are observed (e.g., residual conductivity), and they are attributed to the formation of an impurity band; this means that if the impurity number is not very small, the wave functions of the electrons bound to impurities have a certain overlap and the electrons can tunnel from one impurity site to another. However, if the impurity number is very small, the overlap of the electronic wave functions becomes negligible, i.e., the electronic states become localized. Yet, transport effects are observed, and they are due to the hopping of the electrons from one localized state to another, the necessary energy for the jump being provided by the acoustical phonons since the temperatures are very low. The situation is the same in amorphous materials with localized states.

As mentioned earlier, there is only a collisional contribution to the current and for the dc conductivity formula (1.17) applies. In what follows, limiting ourselves to very low temperatures, we will evaluate the dc conductivity for both crystalline and amorphous materials. It will be shown that in the first case (1.17) leads to an activated type of conduction, $\sigma=\sigma_0 e^{-\epsilon/kT}$, whereas in the second case to the variable range hopping law of Mott and Davis.[23]

1. Crystalline materials

We adopt the model of Miller and Abrahams[28] for nearest-neighbor hopping between donor sites in the presence of acceptors, i.e., "compensation"; for details see their paper. Starting with formula (1.17), we see that

$$(X_\zeta-X_{\zeta'})^2=\{(\zeta|x|\zeta)-(\zeta'|x|\zeta')\}^2=\bar{R}_{\zeta\zeta'}^2, \tag{3.35}$$

where $\bar{R}_{\zeta\zeta'}$ is the mean hopping length in the direction of the electric field, taken along the x axis. In general, $\bar{R}_{\zeta\zeta'}$ depends mainly on the phonon energy $E_{\mathbf{q}}=\epsilon_{\zeta'}-\epsilon_\zeta$, but, for nearest-neighbor hopping, according to the model, it will be approximated by the mean distance between impurities (n type)

$$\bar{R}_{\zeta\zeta'}\approx R_D=(3/4\pi N_D)^{1/3}, \tag{3.36}$$

where N_D is the number of the donors, assumed sufficiently small for hopping to occur.

The transition rate for electron–phonon interaction $w_{\zeta\zeta'}$ is given by (2.11) and (2.12); in the latter formulae $\mathscr{I}_{N,N'}$ and $\mathscr{I}_{N',N}$ are to be replaced by $(\zeta'|e^{i\mathbf{q}\cdot\mathbf{r}}|\zeta)$ and $(\zeta'|e^{-i\mathbf{q}\cdot\mathbf{r}}|\zeta)$, respectively. For acoustical phonons, in the deformation potential model, $w_{\zeta\zeta'}$ has been calculated variationally by Miller and Abrahams,[28] the result for the absorption term, when the effective mass is anisotropic, is

$$w_{\zeta\zeta'}=\lambda R_D^{3/2}|\Delta|N_0 e^{-2R_D/a}, \tag{3.37}$$

where

$$\lambda=\left(\frac{E_1^2}{\pi\rho u_0^2\hbar^4}\right)\left(\frac{2e^2}{3\kappa_0 a^2}\right)^2\left(\frac{\pi a}{4\alpha n^2}\right)^{1/2},$$

Δ is the phonon energy, and N_0 is the equilibrium number of phonons. Miller and Abrahams have done this calculation

for Si and Ge in the effective mass approximation; κ_0 is the dielectric constant, α equals $(a/b)^2 - 1$, a and b are the transverse and longitudinal orbits of the ellipsoids, and n is the number of the ellipsoids. For the emission term N_0 is replaced by $1 + N_0$.

We now consider the case where almost all the states below the Fermi energy are occupied and the ones above it are not, so that the jumps (or hops) occur above the Fermi energy (we consider very low temperatures). The jump from a state above the Fermi energy to a state at or just below the Fermi energy has been eliminated in deriving (1.17) through detailed balance (see LRTIII). All this means that we may take $1 - \langle n_{\zeta'} \rangle_{eq} \approx 1$ in (1.17), neglect the emission term of $w_{\zeta\zeta'}$ and use the approximation (3.8) to write

$$\langle n_\zeta \rangle_{eq} \approx \delta(\epsilon_\zeta - \epsilon_F)/\beta(1 - \langle n_\zeta \rangle_{eq}). \tag{3.38}$$

For the low temperatures of interest here the number of phonons N_0 in (3.37) will be taken as $N_0 \approx e^{-\beta E_q} = e^{-\beta(\epsilon_{\zeta'} - \epsilon_\zeta)}$. As for the densities of states $N(\epsilon_\zeta)$, $N(\epsilon_{\zeta'})$, when transforming the sum in (1.17) to an integral, they will be assumed smoothly varying functions and will be replaced by their value at the Fermi energy (compare Ref. 28).

We now substitute (3.36), (3.37), and (3.38) in (1.17) and carry out the integration over ϵ_ζ, after transformation of the sum to an integral; with $x = \epsilon_{\zeta'} - \epsilon_F$, we find

$$\sigma_{xx} \approx \frac{2e^2\lambda}{\Omega} R_D^{7/2} e^{-2R_D/a} \{N(\epsilon_F)\}^2 \int x e^{-\beta x}\, dx. \tag{3.39}$$

In order for the electrons to overcome the electrostatic attraction in hopping from one donor site to another, the acoustic phonon energy must have a certain minimum value, say ϵ_3. On the other hand, the maximum phonon energy $\epsilon_\zeta + \epsilon$ cannot exceed the Debye energy. Thus, $\epsilon_3 \leqslant x \leqslant \epsilon_3 + \epsilon$, and this is shown schematically in Fig. 1.

With these limits for x we easily find, for $\beta > 1$, the final result

$$\sigma_{xx} = \frac{2e^2\lambda(\beta\epsilon_3 + 1)}{\beta^2\Omega} R_D^{7/2} e^{-2R_D/a} \{N(\epsilon_F)\}^2 e^{-\beta\epsilon_3}. \tag{3.40}$$

This result is of the activation energy type,[28,29] $\sigma = \sigma_0 e^{-\epsilon_3/kT}$; note that we have not used any percolation arguments in deriving (3.40); the treatments of Miller and Abrahams[28] and of Pollak,[29] though with more details, especially with regard to the activation energy, are much more complicated. As for the prefactor, σ_0, there is not, to our knowledge, general agreement in the literature.

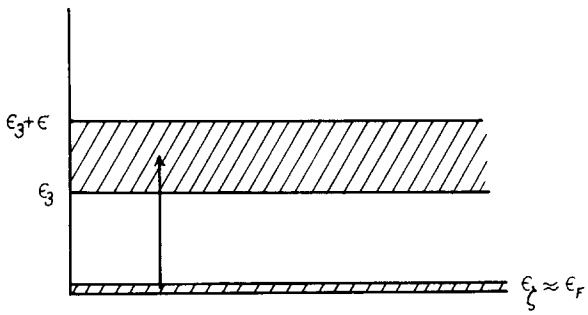


FIG. 1. Energies allowed for hopping transition.

2. Amorphous materials

In strongly disordered materials the conductivity is either of the activation energy type for nearest-neighbor hopping and relatively high temperatures, or it follows Mott's law for variable range hopping: $\sigma = \sigma_0 e^{-(T_0/T)^{1/4}}$ at relatively low temperatures, see Ref. 23; the latter, experimentally verified, is expected to occur at very low temperatures, close to zero. In this case, according to Mott's ideas, the number of phonons available might not be enough to supply the energy necessary for the electron to jump to the nearest available site. The electron must jump among the various available sites, to the one which differs in energy from the original site by the amount of phonon energy. This means that the hopping might not occur between nearest neighbors, as assumed in Sec. 3B1, but between more distant neighbors, i.e., we have a variable range hopping. The situation is depicted schematically in Fig. 2.

Put in another way, the above ideas mean that the mean hopping length is a sensitive function of the phonon energy, i.e., $\overline{R}_{\zeta\zeta'} = f(\epsilon_{\zeta'} - \epsilon_\zeta)$. Now as far as calculations are concerned, we have again to evaluate $\overline{R}^2_{\zeta\zeta'}$ and $w_{\zeta\zeta'}$. The situation, however, when compared to the one of the previous subsection, seems hopeless since the wave functions are not known. One is bound to make further approximations. As suggested by the result of Miller and Abrahams, the transition rate $w_{\zeta\zeta'}$ will be assumed[23,30] to behave as

$$w_{\zeta\zeta'} \approx \gamma_0 e^{-2\alpha\overline{R}_{\zeta\zeta'} - \beta(\epsilon_{\zeta'} - \epsilon_\zeta)}, \tag{3.41}$$

where α^{-1} is a localization length and where we neglected again the emission term of $w_{\zeta\zeta'}$; γ_0 is a constant depending on the strength of the electron–phonon interaction, the density of phonons and very weakly on the energies ϵ_ζ, $\epsilon_{\zeta'}$, and $\overline{R}_{\zeta\zeta'}$.

To proceed further, we have to evaluate $\overline{R}^2_{\zeta\zeta'}$. Mott's way of evaluating it has been criticized and we will use the more rigorous result of Apsley and Hughes,[31] also obtained by Brenig *et al.*[32] The idea is to write the exponential in (3.41) as $-\mathscr{R}$, $\mathscr{R}$ being the "distance" in the four-dimensional space shown above. By means of a model Apsley and

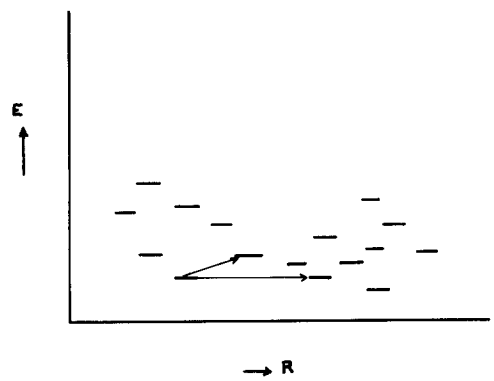


FIG. 2. Hopping processes between statistically distributed states, localized in energy and position. The jump probability is determined by the distance in space and energy.

Hughes[31] show that the mean hopping length for "nearest-neighbor" hopping in the four-dimensional sense $\mathscr{R}_{nn}$ is given by

$$\mathscr{R}_{nn} = C(\epsilon_{\zeta'} - \epsilon_{\zeta})^{1/4}, \tag{3.42}$$

where $C = \Gamma(5/4)(3a'^3\beta/\pi s)^{1/4}$; s is the number of localized states per atom and $a' = 2\alpha a$, a being the interatomic distance.

The densities of states, in the model of Apsley and Hughes, are taken constant and equal to $8\alpha^3 s$. We now use (3.42) to find $\overline{R}_{\zeta\zeta'}$ and substitute the result, together with (3.35), (3.41), and (3.38), into (1.17); again $1 - \langle n_{\zeta'}\rangle_{\text{eq}} \approx 1$. After the integration over ϵ_{ζ}, we find ($x = \epsilon_{\zeta'} - \epsilon_F$)

$$\sigma_{xx} \approx \frac{\gamma_0 e^2}{2\Omega\alpha^2}(8\alpha^3 s)^2 \int (Cx^{1/4} - \beta x)^2 e^{-Cx^{1/4}}\, dx. \tag{3.43}$$

In analogy with the previous subsection we take the limits of x to be ϵ_3' and $\epsilon_3' + \epsilon'$, the two limits being ascribed to the same physical picture as before. Putting $x^{1/4} = \omega$ in (3.43), we find integrals of the type

$$I(n) = \int \omega^n e^{-C\omega}\, d\omega, \quad n = 5, 8, 11. \tag{3.44}$$

The integration is elementary and, for $\beta\epsilon' \geqslant 1$, we find

$$I(n) = \left\{ \frac{\epsilon_3'^{1/4}}{C} + \sum_{k=1}^{n} \frac{n(n-1)\cdots(n-k+1)}{C^{k+1}} \epsilon_3'^{1/4} \right\} e^{-C\epsilon_3'^{1/4}}. \tag{3.45}$$

The final result for σ_{xx} is

$$\sigma_{xx} \approx (2\gamma_0/\Omega)(8\alpha^2 s e)^2 \{C^2 I(5) + \beta^2 I(11) - 2\beta C I(8)\}. \tag{3.46}$$

Since $C \propto \beta^{1/4}$ (3.45) shows that (3.46) can be written as $\sigma = \sigma_0 e^{-(T_0/T)^{1/4}}$ with a complicated prefactor σ_0; that is, we find Mott's well-known law without making some of his already criticized assumptions[30,31] and without using percolation arguments.

4. THE QUANTUM MECHANICAL HALL EFFECT

It has long been realized that the diagonal matrix elements of the current, when the states are the Landau states (2.2), vanish [see Eqs. (2.7) and (2.8)] so no Hall effect results. A critique of a previous approach,[12] which included the electric field in the unperturbed Hamiltonian h^0, was made in LRTIII, Sec. 3.2. For the Hall effect we need the conductivity component ζ_{yx}; as states in LRTIII the collisional contribution is zero since $(\sigma|y - y_{\text{eq}}|\zeta) = 0$ [see Eqs. (2.83)]. Consequently, for the Hall effect we have only a nondiagonal ponderomotive contribution to the current, as formula (1.21) shows.

The approach of LRTIII was to start with the expression (1.21) ($\mu = y$, $\nu = x$); it was then shown with a little algebra and the use of Eqs. (2.7) and (2.8) for the matrix elements of the current that the dc version of (1.21) is given by [see LRTIII, Eq. (3.32)]

$$\sigma_{yx}^{\text{nd}}(0) = \sigma_{yx} = \frac{e}{2B\Omega} \sum_{k,N} (N+1) \times \{\langle\eta_N\rangle_{\text{eq}}(1 - \langle\eta_{N+1}\rangle_{\text{eq}})(1 - e^{-\beta\hbar\omega_0}) - \langle\eta_{N+1}\rangle_{\text{eq}}(1 - \langle\eta_N\rangle_{\text{eq}})(1 - e^{\beta\hbar\omega_0})\}. \tag{4.1}$$

This result is exact. For nondegenerate carriers (4.1) gave the ordinary Hall effect for strong magnetic fields, $\rho_{yx} \sim -B/en_0$, see Eq. (3.37) of LRTIII. At the same time it was shown that $\sigma_{xx}^{\text{nd}}(0)$ vanishes for nondegenerate carriers; see Eq. (3.39).

In the present article we will investigate the oscillatory Hall effect for degenerate statistics starting from (4.1); then, starting from an analogous to (4.1) expression for $\sigma_{xx}^{\text{nd}}(0)$ of LRTIII we will evaluate $\sigma_{xx}^{\text{nd}}(0)$ for degenerate statistics.

A. Oscillatory Hall effect

To simplify the notation, we put $\beta\hbar\omega_0 = x$ and

$$\langle\eta_N\rangle_{\text{eq}} = f_N = 1/(e^{\beta[\epsilon_z + (N+1/2)\omega_0 - \epsilon_F]} + 1),$$
$$\epsilon_z = \hbar^2 k_z^2 .2m^*. \tag{4.2}$$

For degenerate statistics we use the approximation (3.8) in order to write

$$f_N(1 - f_{N+1}) = \frac{f_N}{f_{N+1}} f_{N+1}(1 - f_{N+1}) \approx \frac{f_N}{\beta f_{N+1}} \delta(\epsilon - \epsilon_F''), \tag{4.3}$$

where $\epsilon_F'' = \epsilon_F - (N + 1/2)\hbar\omega_0$, and a similar expression for the second term of (4.1), involving $\epsilon_F' = \epsilon_F - (N - 1/2)\hbar\omega_0$. With the density of states (2.9), we then easily find

$$\begin{aligned}\sigma_{yx} &= \frac{2e}{B\beta(2\pi)^2\lambda^2}\left(\frac{2m^*}{\hbar^2}\right)^{1/2} \sum_N (N+1) \int_0^\infty \frac{d\epsilon_z}{\sqrt{\epsilon_z}} \\ &\times \left\{ (1 - e^{-x})\delta(\epsilon_z - \epsilon_F'') \frac{e^{\beta(\epsilon_z - \epsilon_F'')} + 1}{e^{\beta(\epsilon_z - \epsilon_F')} + 1} \right. \\ &\left. - (1 - e^{x})\delta(\epsilon_z - \epsilon_F') \frac{e^{\beta(\epsilon_z - \epsilon_F')} + 1}{e^{\beta(\epsilon_z - \epsilon_F'')} + 1} \right\} \\ &= \frac{2e \tanh(x/2)}{B\beta(2\pi)^2\lambda^2}\left(\frac{2m^*}{\hbar^2}\right)^{1/2} \sum_N \frac{N+1}{\sqrt{\hbar\omega_0}} \\ &\times \left\{ \frac{1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - (N+1)}} \right. \\ &\left. + \frac{1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - N}} \right\},\end{aligned} \tag{4.4}$$

since $\beta(\epsilon_F'' - \epsilon_F') = -\beta\hbar\omega_0 = -x$. Now the summation over N extends up to that value of N for which the radicants are positive; if this number is I for the second term it is $I - 1$ for the first. By writing the two sums explicitly we can easily show that

$$\sum_{N=0}^{I-1} \frac{N+1}{\sqrt{A - N - 1}} + \sum_{N=0}^{I} \frac{N+1}{\sqrt{A - n}} = \sum_{N=0}^{I} \frac{2N+1}{\sqrt{A - N}}, \tag{4.5}$$

where $A = \epsilon_F/\hbar\omega_0 - 1/2$. We thus arrive at the final result for σ_{yx} ($\lambda^2 = \hbar/m^*\omega_0$)

$$\sigma_{yx} = \frac{e \tanh(\beta\hbar\omega_0/2)}{4\pi^2\beta B}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} (\hbar\omega_0)^{1/2} \times \sum_{N=0}^{I} \frac{2N+1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - N}}. \tag{4.6}$$

This expression shows a clear oscillatory behavior since the quantity $\epsilon_F/\hbar\omega_0 - 1/2 - N$, depending on the value of the magnetic field B ($\omega_0 = eB/m^*$), can become large or small when $\epsilon_F/\hbar\omega_0 - 1/2$ is far or close to the value of a Landau level N. It is easily seen that this behavior is due to the oscillatory behavior of the density of states (2.9). This kind of oscillatory behavior has been observed experimentally by Frederikse and Hosler[33] and by Pavlov *et al.*[34]

For a semiconductor with spherical energy surfaces the Hall coefficient R_H in a semiclassical treatment is related to the conductivity components σ_{yx}, σ_{xx} through[35,36] (the magnetic field is in the z direction)

$$R_H = -\frac{1}{B}\frac{\sigma_{yx}}{\sigma_{yx}^2 + \sigma_{xx}^2} \tag{4.7}$$

with

$$\sigma_{yx} = \frac{ne^2}{m^*\omega_0}\frac{(\omega_0\tau)^2}{1+(\omega_0\tau)^2}, \tag{4.8}$$

$$\sigma_{xx} = \frac{ne^2}{m^*}\frac{\tau}{1+(\omega_0\tau)^2}; \tag{4.9}$$

here τ is a relaxation time and n is the carrier density. For strong magnetic fields ($\omega_0\tau \gg 1$) it is easily seen that $\sigma_{xx} \ll \sigma_{yx}$ and

$$R_H \approx -1/B\sigma_{yx}. \tag{4.10}$$

To see the connection between (4.6) and (4.8) and to write R_H in terms of the carrier (electron) concentration n, we evaluate the Fermi level (the factor 2 below stands for spin)

$$n = \frac{N}{\Omega} = \frac{2}{\Omega}\int_{1/2\hbar\omega_0}^{\infty} N(\epsilon)f(\epsilon)\,d\epsilon, \tag{4.11}$$

where $N(\epsilon)$ is the density of states and $f(\epsilon)$ the Fermi–Dirac distribution function. Since the temperatures are very low, possibly a few degrees Kelvin, we may take $f(\epsilon) \approx 1$ for $kT \ll \epsilon_F$, $\epsilon \leqslant \epsilon_F$, and obtain

$$n = \frac{1}{2\pi^2}\left(\frac{2m^*\hbar\omega_0}{\hbar^2}\right)^{3/2}\sum_N\left(\frac{\epsilon F}{\hbar\omega_0} - \frac{1}{2} - N\right)^{1/2}. \tag{4.12}$$

If we introduce the Fermi level without magnetic field $\epsilon_F^0 = (\hbar^2/2m^*)(3\pi^2 n)^{2/3}$, we can write (4.12) in the form

$$\frac{2}{3}\left(\frac{\epsilon_F^0}{\hbar\omega_0}\right)^{3/2} = \sum_N\left(\frac{\epsilon F}{\hbar\omega_0} - \frac{1}{2} - N\right)^{1/2}. \tag{4.13}$$

With this expression (4.6) takes the form

$$\sigma_{yx} = \frac{ne^2}{m^*\omega_0}\frac{\tanh(\beta\hbar\omega_0/2)}{2\beta\hbar\omega_0}\frac{3}{2}\left(\frac{\hbar\omega_0}{\epsilon_F^0}\right)^{3/2} \times\sum_N\frac{2N+1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - N}}; \tag{4.14}$$

the connection of (4.14) with (4.8) is apparent. As for the Hall coefficient (4.10) and (4.14) give

$$R_H \approx -\frac{1}{ne}\frac{2\beta\hbar\omega_0}{\tanh(\beta\hbar\omega_0/2)}\frac{2}{3}\left(\frac{\epsilon_F^0}{\hbar\omega_0}\right)^{3/2} \times\left(\sum_N\frac{2N+1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - N}}\right)^{-1}; \tag{4.15}$$

the quantum limit version ($N = 0$) of (4.15) is extremely simple:

$$R_H \approx -\frac{1}{ne}\frac{2\beta\hbar\omega_0}{\tanh(\beta\hbar\omega_0/2)}\frac{8}{9}\left(\frac{\epsilon_F^0}{\hbar\omega_0}\right)^3. \tag{4.16}$$

The main difference of our result (4.15), leaving aside its simplicity, from those of the literature[12,37] is that it does not depend on any interaction whereas those of the literature do, e.g., interaction with impurities. The only interaction entering our result is that with a periodic lattice, expressed by the effective mass m^*. Thus (4.16) is a basic general oscillatory Hall effect, *not found before.*

However, as far as agreement with experiment is concerned, our result, too, exhibits the same weakness as those of the literature,[12,37] i.e., the agreement is poor. An order of magnitude agreement is obtained in our case only for fields stronger than roughly 15 kG. One might think that this is due to neglect of σ_{xx} in the formula for R_H; however, as shown by Bastin *et al.*[37] for scattering by impurities $\sigma_{xx} \approx 0.1\sigma_{yx}$, so that the result remains almost the same. It is therefore likely that there is also a specific oscillatory Hall effect depending on the type of interaction, for which one must go beyond the first Born approximation (not allowed by our present formalism). A further study in this direction is being considered.

B. Evaluation of $\sigma_{xx}^{nd}(0) \equiv \sigma_{xx}$

An exact formula analogous to (4.1) was derived in LRTIII [Eq. (3.38)] for the component $\sigma_{xx}^{\rm nd}(0)$; it reads

$$\sigma_{xx} = \frac{ei}{2B\Omega}\sum_{k,N}(N+1)\{\langle\eta_N\rangle_{\rm eq}(1-\langle\eta_{N+1}\rangle_{\rm eq})(1-e^{-\beta\hbar\omega_0}) + \langle\eta_{N+1}\rangle_{\rm eq}(1-\langle\eta_N\rangle_{\rm eq})(1-e^{\beta\hbar\omega_0})\}, \tag{4.17}$$

which differs from (4.1) by the factor i and by the sign of the two contributions. Repeating verbatim the steps of Sec. 4A, which led from (4.1) to (4.6), we obtain

$$\sigma_{xx}^{\rm nd}(0) = \sigma_{xx} = -\frac{ie\tanh(\beta\hbar\omega_0/2)}{4\pi^2\beta B}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} \times(\hbar\omega_0)^{1/2}\sum_{N=0}^{I}\frac{1}{\sqrt{\epsilon_F/\hbar\omega_0 - 1/2 - N}}. \tag{4.18}$$

For very low frequencies ($\omega \to 0$) this imaginary result may contribute to the dielectric constant $\epsilon(\omega)$ of degenerate metals or semiconductors since

$$\epsilon(\omega) = \epsilon^0(\omega) + 4\pi i\sigma(\omega)/\omega. \tag{4.19}$$

ACKNOWLEDGMENTS

We are indebted to M. Charbonneau for preliminary results on the transverse magnetoresistance effect and for fruitful cooperation during the first stages of this work.

We also thank the National Sciences and Engineering Research Council of Canada for their support under Grant #A9522, as well as FCAC Québec.

APPENDIX

In this appendix, we give the values of certain integrals involving Laguerre polynomials, which are encountered in the text; from tables[21] we have been able to find only the first three of them, and we evaluate the remaining ones explicitly:

$$\int_0^\infty e^{-x}x^M [L_N^M(x)]^2\,dx = \frac{\Gamma(M+N+1)}{N!}$$

$$= \frac{(M+N)!}{N!}, \quad M>0, \tag{A1}$$

$$\int_0^\infty e^{-x}x^M [L_{N-M}^M(x)]^2\,dx = \frac{\Gamma(M+N-M+1)}{N!}$$

$$= 1, \quad M>0, \tag{A2}$$

$$\int_0^\infty e^{-x}[L_N(x)]^2\,dx = (-1)^N P_N(-1) = 1, \quad M=0, \tag{A3}$$

where $P_N(x)$ is a Legendre polynomial.

(a) $$I = \int_0^\infty x^{M+1}e^{-x}[L_N^M(x)]^2\,dx$$

$$= (2N+M+1)\frac{(N+M)!}{N!}. \tag{A4}$$

Proof: We can write I as follows:

$$I = \lim_{s\to 1}\left\{-\frac{\partial}{\partial s}\int_0^\infty x^M e^{-sx}[L_N^M(x)]^2\,dx\right\}.$$

Setting $sx = z$, we have

$$I = \lim_{s\to 1}\left\{-\frac{\partial}{\partial s}\int_0^\infty z^M e^{-z}\frac{1}{s^{M+1}}[L_N^M(z/s)]^2\,dx\right\}.$$

We take the derivative explicitly and use the relation[21]

$$z\frac{d}{dz}L_N^M(z) = NL_N^M(z) - (N+M)L_{N-1}^M(z); \tag{A5}$$

further, we use the orthogonality properties of the Laguerre polynomials and (A1). We then easily find (A4). For $M=0$, (A4) gives

$$I_1 = \int_0^\infty xe^{-x}[L_N(x)]^2\,dx = 2N+1. \tag{A6}$$

(b) The integral $I_n = \int_0^\infty x^n e^{-x}[L_N(x)]^2\,dx$ can be evaluated in exactly the same way; we have

$$I_n = \lim_{s\to 1}\left\{(-1)^n\frac{\partial^n}{\partial s^n}\int_0^\infty e^{-sx}[L_N(x)]^2\,dx\right\};$$

we then put $sx = z$ and use (A5) and (A4). We easily find (A6),

$$I_2 = 2(3N^2+3N+1) \tag{A7}$$

and

$$I_3 = 2(2N+1)(5N^2+5N+3). \tag{A8}$$

[1]K. M. van Vliet, J. Math. Phys. **19**(6), 1345 (1978).
[2]K. M. van Vliet, J. Math. Phys. **20**(12), 2573 (1979).
[3]M. Charbonneau, K. M. van Vliet, and P. Vasilopoulos, J. Math. Phys. **23**(2), 318 (1982).
[4]Involved also in (1.14) is the approximation $\tau(\epsilon_\xi)\approx\tau(\epsilon_{\xi'})$. However, this approximation is considerably better than the usual "relation time approximation"; read the discussion in LRTIII following Eq. (2.54).
[5]M. Charbonneau, Ph.D. thesis, University of Montreal (in preparation).
[6]P. Vasilopoulos, Ph.D. thesis, University of Montreal, 1982.
[7]E. N. Adams and T. D. Holstein, J. Phys. Chem. Solids **10**, 254 (1959).
[8]P. N. Argyres and L. M. Roth, J. Phys. Chem. Solids **12**, 89 (1959).
[9]V. Čápek, J. Phys. C **8**, 479 (1975).
[10]I. P. Zvagin, Phys. Status Solidi b **101**, 9 (1980).
[11]M. Charbonneau and K. M. van Vliet, Phys. Status Solidi (b) **101**, 509 (1980).
[12]A. H. Kahn and P. R. Frederikse, Solid State Phys. **9**, 257 (1959).
[13]R. C. Enck, A. S. Saleh, and H. Y. Fan, Phys. Rev. **182**, 790 (1969).
[14]J. R. Barker, J. Phys. C **5**, 1657 (1972).
[15]R. L. Peterson, Phys. Rev. B **6**, 3756 (1972).
[16]C. Cohen-Tounoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley-Interscience, New York, 1977), Vol. 2.
[17]P. N. Argyres and E. N. Adams, Phys. Rev. **104**, 900 (1956).
[18]P. N. Argyres, J. Phys. Chem. Solids, **4**, 19 (1958).
[19]G. I. Kharus and I. M. Tsidil'kovskii, Sov. Phys. Semiconductors **5**, 534 (1971).
[20]J. R. Barker, J. Phys. C **6**, 880 (1973).
[21]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[22]R. A. Stradling and R. A. Wood, J. Phys. C **1**, 1711 (1970).
[23]N. F. Mott and E. A. Davis, *Electronic Processes in Noncrystalline Materials* (Clarendon, Oxford, 1979).
[24]R. Kubo, S. J. Miyake, and N. Hashitsume, Solid State Phys. **17**, 269 (1964).
[25]A. L. Efros, Fiz. Tverd. Tela (Leningrad) **3**, 2848 (1961) [Sov. Phys. Solid State **3**, 2079 (1962)].
[26]V. L. Gurevich and Yu. A. Firsov, Zh. Eksp. Teor. Fiz. **40**, 198 (1961) [Sov. Phys. JETP **13**, 137 (1961)].
[27]R. A. Stradling and R. A. Wood, J. Phys. C **3**, L94 (1970).
[28]A. Miller and E. Abrahams, Phys. Rev. **120**, 745 (1960).
[29]M. Pollak, J. Non-Cryst. Solids **11**, 1 (1972).
[30]V. Ambegaokar, B. I. Halperin, and I. S. Langer, Phys. Rev. B **4**, 2613 (1971).
[31]N. Apsley and H. P. Hughes, Phil. Mag. **30**, 963 (1974).
[32]W. Brenig, G. Döhler, and P. Wölfe, Z. Phys. **246**, 1 (1971).
[33]H. P. R. Frederikse and W. R. Hosler, Phys. Rev. **108**, 880 (1958).
[34]S. T. Pavlov, R. V. Parfen'ev, Yu. A. Firsov, and S. S. Shalyt, Zh. Eksp. Teor. Fiz. **48**, 1565 (1965) [Sov. Phys. JETP **21**, 1049 1(965)].
[35]J. Ziman, *Electrons and Phonons* (Clarendon, Oxford, 1960).
[36]C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).
[37]A. Bastin, C. Lewiner, O. Betbeder-Matibet, and P. Nozieres, J. Phys. Chem. Solids **32**, 1811 (1971).

Quantum harmonic analysis on phase space

R. Werner
Fachbereich Physik, Universität Osnabrück, Postfach 4469, D-4500 Osnabrück, West Germany

(Received 26 April 1983; accepted for publication 29 July 1983)

Relative to an irreducible representation of the canonical commutation relations, convolutions between quantum mechanical operators and between functions and operators are defined, for which the usual Weyl transform acts as a Fourier transform. Basic properties of these operations are developed in close analogy to harmonic analysis on $\mathbb{R}^{2n}$. Using the quantum version of Wiener's approximation theorem, a natural one-to-one correspondence between the closed, phase-space translation invariant subspaces of classical and quantum observables is established.

PACS numbers: 03.65. – w

I. INTRODUCTION

In recent years there has been a renewed interest in the connections between classical mechanics and nonrelativistic quantum mechanics. Such studies have been less concerned with quantization rules, i.e., with an explanation of the structure of quantum mechanics in classical terms, than with the use of quantum-classical analogies for proving results within quantum mechanics. For example, the Berezin–Lieb inequalities[1] provide bounds for quantum mechanical partition functions in terms of their classical counterparts, and the number of bound states of a quantum system is estimated by the phase space volume available to the classical system.[2] Thus phase space intuition becomes a guiding principle for the further development of quantum mechanics.

The purpose of this paper is to demonstrate parallels between classical harmonic analysis on $\mathbb{R}^{2n}$ (phase space) and certain structures in quantum mechanics. The resulting formalism is easy to handle and suggests quantum analogs of many classical theorems, some of which will be proven in this note by "reduction to the classical case." Some connections of this type, particularly between the Weyl and Fourier transforms, are rather well known,[3,4] but to the author's knowledge only special cases of the convolutions defined in Sec. III have so far been introduced (in particular, the convolution with "coherent states" [5,6]). The systematic and unrestricted use of these convolutions will play a crucial role for the transfer of structure between classical and quantum theories and, especially, for "Wiener's approximation theorem" and its consequences developed in Sec. IV. Theorem 4.1 is an important tool for studying notions of physical distinguishability of quantum states and corresponding classical notions. This problem, which was the author's original motivation for the present work, will be analyzed in a separate publication.[7]

II. PRELIMINARIES AND NOTATION

In the following exposition, the close parallels between quantum and classical harmonic analysis will be emphasized by placing the classical and quantum objects side by side in a direct sum construction: If $\mathscr{A}_0$ denotes one of the function spaces of classical harmonic analysis and $\mathscr{A}_1$ denotes its quantum mechanical counterpart, we shall often study directly the space $\mathscr{A} = \mathscr{A}_0 \oplus \mathscr{A}_1$. The operations of the theory, like phase space translations, convolutions, and Fourier transforms, will be defined on such spaces $\mathscr{A}$. In this manner pairs and quadruples of definitions, theorems, and proofs can be stated in a very condensed form. The formulations for $\mathscr{A}_0$ and $\mathscr{A}_1$ can be regained by introducing the subscripts "0" and "1" everywhere and reading them as "classical" and "quantum." The obvious disadvantage of this notation is, of course, that I have to deviate from the standard notations. For example, the space of "observables" $\mathscr{L}^\infty(X,dx) \oplus \mathscr{B}(\mathscr{H})$ will be written as $\mathscr{R}^\infty = \mathscr{R}_0^\infty \oplus \mathscr{R}_1^\infty$. I hope that this notation will serve less to confuse the reader than to enhance the intuition for the correspondences between classical and quantum mechanics.

The basic objects of the theory are a symplectic (real) vector space X (phase space) of dimension $2N < \infty$, whose symplectic form will be denoted by $\{\cdot,\cdot\}: X \times X \to \mathbb{R}$, and an irreducible representation of the Weyl commutation relations over X in a Hilbert space $\mathscr{H}$, i.e., a strongly continuous map E_1 taking $x \in X$ to a unitary operator $E_1(x)$ such that $\{E_1(x) | x \in X\}' = \mathbb{C}1$ and

$$E_1(x)E_1(y) = e^{i\{x,y\}/2}E_1(x+y) \quad \text{for all } x,y \in X .$$

By von Neumann's uniqueness theorem,[8] E_1 is determined up to unitary equivalence and we may set: $X = \mathbb{R}^N \times \mathbb{R}^N$ = momentum ⊕ configuration space, $\{(p,q),(p',q')\} = p\cdot q' - q\cdot p'$, $\mathscr{H} = \mathscr{L}^2(\mathbb{R}^N, dq^N)$, and $(E_1(p,q)\psi)(\xi) = \exp(i(p\cdot q/2) + ip\cdot\xi)\psi(\xi+q)$ for $(p,q) \in X$ and $\psi \in \mathscr{H}$. The Lebesgue measure on X will be denoted by dx and normalized so that in any such representation $dx = (2\pi\hbar)^{-N} dp_1 \cdots dq_N$. (This choice is suggested by Lemma 3.1) I shall set $\hbar = 1$ in what follows.

The space of observables will be denoted by $\mathscr{R}^\infty := \mathscr{L}^\infty(X,dx) \oplus \mathscr{B}(\mathscr{H}) \equiv \mathscr{R}_0^\infty \oplus \mathscr{R}_1^\infty$. This W^*-algebra carries a trace denoted tr, which coincides with the usual trace of $\mathscr{B}(\mathscr{H})$ on $\mathscr{R}_1^\infty$ and with $\int dx$ on $\mathscr{R}_0^\infty$, i.e., $\mathrm{tr}(f \oplus A) = \int dx\, f(x) + \mathrm{tr}\, A$ for $f \in \mathscr{R}_0^\infty$ and $A \in \mathscr{R}_1^\infty$, provided both summands exist. For $1 \leqslant p < \infty$, we set $\mathscr{R}^p = \mathscr{R}_0^p \oplus \mathscr{R}_1^p \equiv \mathscr{L}^p(X,dx) \oplus \mathscr{T}^p(\mathscr{H})$, so that $\mathscr{R}^p$ is the closure of $\{A \in \mathscr{R}^\infty | \mathrm{tr}(|A|^p) < \infty\}$ in the norm $\|A\|_p := (\mathrm{tr}|A|^p)^{1/p}$. For $1 \leqslant p < \infty$ and $p^{-1} + q^{-1} = 1$, $\mathscr{R}^q$ is the dual space of $\mathscr{R}^p$. (Compare Ref. 9, p. 43.) In particu-

 0022-2488/84/051404-08$02.50

lar, $\mathscr{R}^1$ is the predual of $\mathscr{R}^\infty$, the canonical bilinear form $\langle\cdot,\cdot\rangle:\mathscr{R}^1\times\mathscr{R}^\infty\to\mathbb{C}$ being given by $\langle T,A\rangle=\mathrm{tr}(TA)$. For the norm in $\mathscr{R}^\infty$, $\|\cdot\|$ and $\|\cdot\|_\infty$ will be used interchangeably.

The phase space translations are represented by $\alpha:X\to\mathrm{Aut}(\mathscr{R}^\infty)$, given by $(\alpha_x f)(y)=f(y-x)$ for $f\in\mathscr{R}_0^\infty$ and by $\alpha_x A=E_1(x)AE_1(-x)$ for $A\in\mathscr{R}_1^\infty$. Because E_1 is a representation of X up to a factor, the representation relation $\alpha_x\alpha_y=\alpha_{x+y}$ holds on $\mathscr{R}^\infty$. By $\|\cdot\|_p$-continuity, α can be extended from $\mathscr{R}^1\cap\mathscr{R}^\infty$ to $\alpha_x:\mathscr{R}^p\to\mathscr{R}^p$ for $1\leqslant p<\infty$. On $\mathscr{R}^p(p<\infty)$ α is strongly continuous, but not on $\mathscr{R}^\infty$. The classical analogs of the "exponential operators" $E_1(x)\in\mathscr{R}_1^\infty$ are the exponential functions $E_0(x)\in\mathscr{R}_0^\infty$ given by $E_0(x)(y):=e^{i\{x,y\}}$. Thus for $k=0,1$ the relation $\alpha_x(E_k(y))=e^{i\{x,y\}}E_k(y)$ holds.

G will denote the group of linear transformations of X, leaving the symplectic form $\{\cdot,\cdot\}$ invariant. Since $\{E_1(gx)\}_{x\in X}$ again satisfies the Weyl relations, von Neumann's theorem asserts the existence of unitary operators $\{U_g\}_{g\in G}$ such that $U_gE_1(x)U_g^*=E_1(gx)$. One easily checks that U represents G up to a factor. Hence we may define a representation $\beta:G\to\mathrm{Aut}(\mathscr{R}^\infty)$ by setting $(\beta_g f)(x)=f(g^{-1}x)$ for $f\in\mathscr{R}_0^\infty$ and $\beta_g A=U_gAU_g^*$ for $A\in\mathscr{R}_1^\infty$. Again β_g is strongly continuous on $\mathscr{R}^p$ for $p<\infty$ but not on $\mathscr{R}^\infty$. For $k=0,1$ we have $\beta_gE_k(x)=E_k(gx)$ and $\beta_g\alpha_x=\alpha_{gx}\beta_g$, so that α and β together are a representation of the affine symplectic group. The transformation β_g corresponding to $g:x\mapsto(-x)$ will be used very frequently and will be denoted by β_-. It is implemented on $\mathscr{R}_1^\infty$ by the unitary operator $(U_-\psi)(\xi)=\psi(-\xi)$.

The space $\mathscr{B}=\mathscr{B}_0\oplus\mathscr{B}_1$ of complex valued functions on $\{0,1\}\times X$ will be the range of the Fourier transform, analytical properties being postulated as needed. $\mathscr{B}^2\subset\mathscr{B}$ will denote the subspace of square-integrable functions. In Sec. IV the space $\mathscr{C}=\mathscr{C}_0\oplus\mathscr{C}_1=\{A\in\mathscr{R}^\infty\,|\lim_{x\to0}\|\alpha_xA-A\|_\infty=0\}$ of strongly continuous elements for α, will be important. $\mathscr{C}$ is a C^*-subalgebra of $\mathscr{R}^\infty$, and $\mathscr{C}_0$ is just the space of bounded uniformly continuous functions on X. Whenever sums of subscripts 0,1 occur, they are to be taken mod 2, i.e., $1+1=0$.

III. CONVOLUTIONS AND FOURIER TRANSFORMS

We can now state the lemma which, in spite of its elementary character, is the cornerstone of the entire theory:

Lemma 3.1: Let $T_1,T_2\in\mathscr{R}_1^1$, i.e., be trace class operators. Then the function $x\mapsto\mathrm{tr}(T_1(\alpha_xT_2))$ is integrable and

$$\int dx\,\mathrm{tr}(T_1(\alpha_xT_2))=(\mathrm{tr}\,T_1)(\mathrm{tr}T_2)\,.$$

Proof: By bilinearity the proof is immediately reduced to the case $T_1,T_2\geqslant0$ and by spectral decomposition to the case that $T_i=P_{\psi_i}$ are one-dimensional projections. Let $\varphi_q(\xi):=\overline{\psi_2(\xi)}\psi_1(\xi+q)$. Then $\varphi_q\in\mathscr{L}^1(\mathbb{R}^N,d\xi^N)$ and $\langle\psi_2,E(p,q)\psi_1\rangle=e^{i\{p,q\}/2}\int d\xi^N e^{ip\cdot\xi}\varphi_q(\xi)$ $=e^{i(p\cdot q)/2}(2\pi)^{N/2}\hat{\varphi}_q(p)$, where $\hat{\varphi}_q$ denotes the Fourier transform of φ_q. Hence by the Plancherel formula:

$$\begin{aligned}\int dx\,\mathrm{tr}(T_1(\alpha_xT_2))&=(2\pi)^{-N}\int dp^Ndq^N\,|\langle\psi_2,E(p,q)\psi_1\rangle|^2\\&=\int dq^Ndp^N\,|\hat{\varphi}_q(p)|^2\\&=\int dq^N\,d\xi^N\,|\varphi_q(\xi)|^2\\&=\int dq^N\,d\xi^N\,|\psi_2(\xi)|^2|\psi_1(\xi+q)|^2\\&=\|\psi_1\|^2\|\psi_2\|^2=1\,.\end{aligned}$$ □

Various forms of this lemma exist in the literature, often formulated for the special case that T_1 or T_2 is a pure or even a coherent state. For our purposes it is important to note that its classical analog for $T_1,T_2\in\mathscr{R}_0^1$ (reading tr as $\int dx$ as introduced above) is also true and a simple property of the convolution of integrable functions. In fact, the convolution of functions may be rewritten as $(f_1*f_2)(x)=\int dy\,f_1(y)f_2(x-y)$ $=\mathrm{tr}[f_1(\alpha_x\beta_-f_2]$ or equivalently as $f_1*f_2=\int dy\,f_1(y)\alpha_yf_2$. This motivates the following:

Definition: Let $f_1f_2\in\mathscr{R}_0^1$ be integrable functions and $T_1,T_2\in\mathscr{R}_1^1$ trace class operators. Then their *convolutions* are defined by

$$f_1*f_2\in\mathscr{R}_0^1:\ (f_1*f_2)(x):=\int dy\,f_1(y)f_2(x-y)\,,$$

$$T_1*T_2\in\mathscr{R}_0^1:\ (T_1*T_2)(x):=\mathrm{tr}(T_1(\alpha_x\beta_-T_2))\,,$$

$$f_1*T_1:=T_1*f_1:=\int dy\,f_1(y)\alpha_yT_1\ \in\mathscr{R}_1^1\,.$$

By linear extension convolution becomes a composition $*:\mathscr{R}^1\times\mathscr{R}^1\to\mathscr{R}^1$ such that for $i,j\in\{0,1\}:\mathscr{R}_i^1*\mathscr{R}_j^1\subset\mathscr{R}_{i+j}^1$. Convolutions between $\mathscr{R}^1$ and its dual $\mathscr{R}^\infty$ are defined by

$$\langle S,T*A\rangle:=\langle S,A*T\rangle:=\langle(\beta_-T)*S,A\rangle$$

for $T,S\in\mathscr{R}^1$, $A\in\mathscr{R}^\infty$.

This definition implicitly asserts that the two definitions of $T*A$ for $T\in\mathscr{R}^1$ and $A\in\mathscr{R}^\infty\cap\mathscr{R}^1$ coincide or, what is the same thing, that the above formulas also define convolution on $\mathscr{R}^1\times\mathscr{R}^\infty$ [integrals understood in the $\sigma(\mathscr{R}^\infty,\mathscr{R}^1)$-sense]. This is justified by the following:

Proposition 3.2: (1) Convolution is commutative and associative. The convolution of positive elements is positive.

(2) $\alpha_x(T*A)=(\alpha_xT)*A+(\alpha_xA)$,

$$\beta_g(T*A)=(\beta_gT)*(\beta_gA)\,.$$

(3) For $T\in\mathscr{R}^1$: $T*1=(\mathrm{tr}\,T)\cdot1$.

(4) For $T\in\mathscr{R}^1$, $A\in\mathscr{R}^\infty$: $T*A\in\mathscr{C}$ is strongly continuous for α and $\langle T,A\rangle=((\beta_-T)*A)_0(0)$, where the subscript "0" refers to the first component in $\mathscr{R}^\infty=\mathscr{R}_0^\infty\oplus\mathscr{R}_1^\infty$.

(5) Young's inequality: Let $1\leqslant p,q,r\leqslant\infty$ with $p^{-1}+q^{-1}=1+r^{-1}$. Then for $T_1,T_2\in\mathscr{R}^1\cap\mathscr{R}^\infty$: $\|T_1*T_2\|_r\leqslant\|T_1\|_p\|T_2\|_q$. Consequently, convolution extends by norm continuity to $*:\mathscr{R}^p\times\mathscr{R}^q\to\mathscr{R}^r$.

Proof: For convolutions in $\mathscr{R}^1$, (1)–(4) are elementary with the exception of the associativity $T_1*(T_2*T_3)=(T_1*T_2)*T_3$ for three operators, which again uses Lemma 3.1: Let $T_0\in\mathscr{R}_1^1$; then

$$
\begin{aligned}
&\langle T_0, T_1 * (T_2 * T_3)\rangle \\
&\quad = \operatorname{tr} T_0 \int dx\, (\alpha_x T_1)\cdot \operatorname{tr}[T_2(\alpha_x \beta_- T_3)] \\
&\quad = \int dx\ \operatorname{tr}[(\alpha_x T_1) T_0]\ \operatorname{tr}[T_3(\alpha_x \beta_- T_2)] \\
&\quad = \iint dx\, dy \operatorname{tr}[(\alpha_x T_1) T_0 (\alpha_y T_3)(\alpha_{x+y} \beta_- T_2)] \\
&\quad = \iint dy\, dx \operatorname{tr}[T_0(\alpha_y T_3)\cdot \alpha_x((\alpha_y \beta_- T_2) T_1)] \\
&\quad = \int dy \operatorname{tr}[T_0(\alpha_y T_3)]\ \operatorname{tr}[T_1(\alpha_y \beta_- T_2)] \\
&\quad = \langle T_0, (T_1 * T_2) * T_3\rangle\ .
\end{aligned}
$$

These results justify the above definition and one may prove their extension to $*: \mathscr{R}^1 \times \mathscr{R}^\infty \to \mathscr{R}^\infty$ by taking adjoints. The strong continuity of $T*A$ follows from the trivial bound $\|T*A\|_\infty \leqslant \|T\|_1 \|A\|_\infty$, (2), and the strong continuity of α on $(\mathscr{R}^1, \|\cdot\|_1)$.

(5) is trivial for $p = q = r = 1$ and $p = 1, q = r = \infty$ and follows for the other values by an interpolation argument to be taken verbatim from Ref. 9, p. 28ff. Note, however, that this proof does not yield the best constants.[10] □

The map $f \mapsto T*f$ for $T \in \mathscr{R}^1_1$ assigns an operator to each function on phase space, and, conversely, $A \mapsto T*A$ transforms operators into functions. Maps of this kind are often called "correspondence rules," and the following proposition characterizes the type of correspondences that may be induced in this manner.

Proposition 3.3: Let $i = 0$ or 1. A linear operator $\Gamma: \mathscr{R}^\infty_i \to \mathscr{R}^\infty_{i+1}$ is called a *positive correspondence rule*, if it is normal [i.e., $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-continuous], $A \geqslant 0 \Rightarrow \Gamma(A) \geqslant 0$, $\Gamma(1) = 1$, and $\alpha_x \Gamma(A) = \Gamma(\alpha_x A)$. Then:

(1) The positive correspondence rules are precisely the maps of the form $\Gamma(A) = T*A$ for a (necessarily unique) trace class operator $T \in \mathscr{R}^1_1$ satisfying $T \geqslant 0$ and $\operatorname{tr} T = 1$. (The assumption of normality is essential).

(2) Γ is completely positive, so that $\Gamma(A)^* \Gamma(A) \leqslant (A^* A)$ for all $A \in \mathscr{R}^\infty_i$.

(3) Berezin–Lieb inequalities: if $A^* = A \in \mathscr{R}^\infty_i$ and φ is a positive convex continuous function on the spectrum of A, then

$$\operatorname{tr} \varphi(\Gamma(A)) \leqslant \operatorname{tr} \varphi(A)\ .$$

This result depends only on the property that Γ is "doubly stochastic" (i.e., preserves positivity, unit, and trace).

Proof: (1) Let $i = 1$. Since $\Gamma: \mathscr{R}^\infty_1 = \mathscr{B}(\mathscr{H}) \to \mathscr{L}^\infty(X) = \mathscr{R}^\infty_0$ is normal, it has a preadjoint $\Gamma_*: \mathscr{L}^1(X) \to \mathscr{T}^1(\mathscr{H})$, which has to be positive, α-covariant, and normalization-preserving. On the other hand, let $i = 0$ and consider positive elements $f \in \mathscr{R}^1_0 \cap \mathscr{R}^\infty_0$ and $S \in \mathscr{R}^1_1$. Then $\int dx \operatorname{tr}(S(\alpha_x f)) = \int dx \langle \Gamma_*(S), \alpha_x f\rangle = \langle \Gamma_*(S), 1\rangle (\int dx f(x)) < \infty$. Hence by Lemma 3.1. $\Gamma(f)$ is trace class and $\operatorname{tr} \Gamma(f) = \int dx f(x)$, so that by $\|\cdot\|_1$-continuity, Γ extends to an operator $\Gamma: \mathscr{L}_1(X) \to \mathscr{T}_1(\mathscr{H})$ which is also positive, normalized, and covariant. Thus both cases $i = 0, 1$ are reduced to the proof that this operator must be of the form $f \mapsto T*f$. Consider the $\mathscr{T}^1(\mathscr{H})$-valued measure $\sigma \mapsto \Gamma(\chi_\sigma)$ for $\sigma \subset X$ restricted to some compact subset. Then $\operatorname{tr} \Gamma(\chi_\sigma) = \|\Gamma(\chi_\sigma)\| = \mu(\sigma)$ (μ = "dx"measure). This measure is countably additive, since increasing sequences converge in $\mathscr{T}^1(\mathscr{H})$. Thus we may invoke the Radon–Nikodym theorem for $\mathscr{T}^1(\mathscr{H})$ (see Ref. 11, p. 79) to obtain a strongly measurable function $\tilde{T}: X \mapsto \mathscr{T}^1(\mathscr{H})$ such that $\Gamma(f) = \int dx f(x) \tilde{T}(x)$. Covariance implies that, for each $y \in X$, $\alpha_y \tilde{T}(x) = \tilde{T}(x + y)$ a.e. (dx). The measurable function $\tilde{T}_0(x) = \alpha_{-x} \tilde{T}(x)$ defines another map $\Gamma_0: \mathscr{L}^1(X) \to \mathscr{T}^1(\mathscr{H})$ by $\Gamma_0(f) = \int dx f(x) \tilde{T}_0(x)$, satisfying $\Gamma_0(\alpha_x f) = \Gamma_0(f)$. Hence $\Gamma_0(f) = (\int dx f(x)) \cdot T$ for a fixed $T \in \mathscr{T}^1(\mathscr{H})$. By uniqueness of Radon–Nikodym derivatives, $\tilde{T}_0(x) = T$ and $\tilde{T}(x) = \alpha_x T$ a.e. (dx). Hence $\Gamma(f) = \int dx f(x) \alpha_x T = f*T$ for a unique $T \in \mathscr{R}^1_1$, which has to be positive and normalized.

A nonnormal $\Gamma: \mathscr{R}^\infty_i \to \mathscr{R}^\infty_{i+1}$ can be constructed from any state $T \in \mathscr{R}^1_i$ and invariant mean η as $\Gamma(A) = \eta(\langle T, \alpha.A\rangle) \cdot 1$ for $A \in \mathscr{R}^\infty_i$.

(2) Positive maps from or into abelian algebras are completely positive. This and the result (2) are to be found in Ref. 12.

(3) The cone of functions φ for which the inequality is valid is closed under suprema and the operations $\tilde{\varphi}(t) = \varphi(-t)$ and $\tilde{\varphi}(t) = \varphi(t + t_0)$ [since $\Gamma(1) = 1$]. Hence we need only consider the case $\varphi(t) = t_+ =$ positive part: $\operatorname{tr} A_+ \equiv \inf\{\operatorname{tr} B \mid B \geqslant 0, B \geqslant A\} = \inf\{\operatorname{tr} \Gamma(B) \mid B \geqslant 0, B \geqslant A\} \geqslant \inf\{\operatorname{tr} \Gamma(B) \mid \Gamma(B) \geqslant 0, \Gamma(B) \geqslant (A)\} \geqslant \inf\{\operatorname{tr} B \mid B \geqslant 0, B \geqslant \Gamma(A)\} = \operatorname{tr}[\Gamma(A)_+]$. For two alternate proofs, see Ref. 1. □

For the comparison of classical and quantum harmonic analysis it is interesting to note that analogous maps $\Gamma: \mathscr{R}^\infty_i \to \mathscr{R}^\infty_i$ are not in general induced by convolution with $f \in \mathscr{R}^1_0$. (Take the identity.) Thus we have made essential use of the Radon–Nikodym property of $\mathscr{R}^1_1$, which fails for $\mathscr{R}^1_0$.

The case $i = 0$ of Proposition 3.3(1) is due to Holevo[13] and characterizes covariant observables[14,15] on phase space: Γ assigns to each $\sigma \subset X$ an operator $\Gamma(\chi_\sigma)$ describing the yes–no measurement as to whether position and momentum of the system are in σ. By far the best known example of a positive correspondence rule is the case that T is a coherent state (ground state of some oscillator). However, many properties which have been stated in this context are independent of this special form of T.

By proposition 3.2 $(\mathscr{R}^1, *)$ is a commutative Banach algebra, whose function space representation is, of course, related to the Fourier transform. As in the classical case, it will be defined as the trace (or integral) of an element with the exponentials. For operators T, the function $x \mapsto \operatorname{tr}(TE_1(x))$ is known as the Weyl transform of T. In order to distinguish between functions on X, which are transforms of operators and of functions, respectively, we shall use two copies of X, so that all transforms are in $\mathscr{B} = \mathscr{B}_0 \oplus \mathscr{B}_1$, the set of complex valued functions on $\{0,1\} \times X$. $\mathscr{B}^2 = \mathscr{B}^2_0 \oplus \mathscr{B}^2_1$ will be the subspace of square integrable functions and the product in $\mathscr{B}$ will be defined as $(fg)_i(x) = \Sigma_{k=0,1} f_k(x) g_{k+1}(x)$. Thus we arrive at the following:

Definition: The *Fourier transform* $\mathscr{F}: \mathscr{R}^1 \to \mathscr{B}$ is the map $\mathscr{F} = \mathscr{F}_0 \oplus \mathscr{F}_1, \mathscr{F}_i: \mathscr{R}^1_i \to \mathscr{B}_i$ $(i = 0,1)$ defined by

$$(\mathscr{F}_i T)(x) = \operatorname{tr}(TE_i(x))\ .$$

Note that by this definition $\mathscr{F}_i T_j = 0$ for $i \neq j$ and $T_j \in \mathscr{R}^1_j$, since the product under the trace is taken in $\mathscr{R}^\infty$. $\mathscr{F}_0$ is the ordinary Fourier transform [recall the factor $(2\pi)^{-N} = (2\pi)^{-(1/2)\dim X}$ in the definition of dx], where the

dual of X has been identified with X via the symplectic form.

Proposition 3.4: (1) For $T,S\in\mathscr{R}^1$: $\mathscr{F}(T*S)=(\mathscr{F}T)(\mathscr{F}S)$, i.e.,

$$\mathscr{F}_i(T*S)(x)=\sum_{k=0,1}(\mathscr{F}_k T)(x)\cdot(\mathscr{F}_{i+k}S)(x)\,.$$

(2) $(\mathscr{F}\alpha_x T)(y)=e^{i\{y,x\}}(\mathscr{F}T)(y)$,

$(\mathscr{F}\beta_g T)(y)=(\mathscr{F}T)(gy)$.

(3) For $f_1,f_2\in\mathscr{R}^1_0\cap\mathscr{R}^\infty_0$ and $T_1,T_2\in\mathscr{R}^1_1$,

$$(\mathscr{F}_0(f_1f_2))(x)=\int dy\,(\mathscr{F}_0 f_1)(x-y)(\mathscr{F}_0 f_2)(y)\,,$$

$$(\mathscr{F}_1(T_1T_2))(x)=\int dy(\mathscr{F}_1T_1)(x-y)(\mathscr{F}_1T_2)(y)e^{i/2\{x,y\}}\,.$$

(4) $\mathscr{F}$ extends to a unitary operator $\mathscr{F}:\mathscr{R}^2\to\mathscr{B}^2$, i.e., for $T_1,T_2\in\mathscr{R}^2$,

$$\mathrm{tr}(T_1^*T_2)=\sum_{i=0,1}\int dx\ \overline{(\mathscr{F}_i T_1)(x)}(\mathscr{F}_i T_2)(x).$$

The inverse $\mathscr{F}^{-1}=\mathscr{F}_0^{-1}\oplus\mathscr{F}_1^{-1}$ on $\mathscr{B}^2$ is given by

$$\mathscr{F}^{-1}f=\sum_{i=0,1}\int dx\,f_i(x)E_i(-x).$$

(5) Hausdorff–Young inequality: For $T\in\mathscr{R}^1\cap\mathscr{R}^2$ and $1\leqslant p\leqslant 2$, $q^{-1}=1-p^{-1}$:

$$\|\mathscr{F}T\|_p\leqslant\|T\|_p\,.$$

(6) Riemann–Lebesgue lemma: For $T\in\mathscr{R}^1$, $\mathscr{F}T$ is a continuous function vanishing at infinity. Moreover, $\mathscr{F}_1T$ is square-integrable.

(7) Bochner's theorem: A function $f:\{0,1\}\times X\to\mathbb{C}$ is the Fourier transform of some positive $T\in\mathscr{R}^1$ iff f is continuous and for all $\{x_1\cdots x_n\}\subset X$ and $\{\zeta_1\cdots\zeta_n\}\subset\mathbb{C}$:

(0) $\sum_{jk}f_0(x_j-x_k)\bar{\zeta}_j\zeta_k\geqslant 0$

and

(1) $\sum_{jk}f_1(x_j-x_k)e^{i/2\{x_j,x_k\}}\bar{\zeta}_j\zeta_k\geqslant 0$.

Proof: I shall restrict myself to samples from the less trivial parts of the proof.

(1) Let $T,S\in\mathscr{R}^1_1$. Then

$$\begin{aligned}\mathscr{F}_0(T*S)(x)&=\int dy\,e^{i\{x,y\}}\mathrm{tr}(T\alpha_y\beta_{-}S)\\&=\int dy\,\mathrm{tr}(TE_1(x)E_1(y)E_1(-x)(\beta_{-}S)E_1(-y))\\&=\int dy\,\mathrm{tr}[(TE_1(x))\alpha_y(E_1(-x)(\beta_{-}S))]\\&=(\mathrm{tr}\,TE_1(x))(\mathrm{tr}\,E_1(-x)\beta_{-}S)\\&=(\mathscr{F}_1T)(x)(\mathscr{F}_1S)(x)\,.\end{aligned}$$

(2) is trivial and the relation (3) between operator products is well known.[3]

(4) follows from (3) by setting $x=0$.

(5) is proven by interpolation[9] between $p=1$ and $p=2$. Again this proof does not give the best constants.[10]

(6) For $T\in\mathscr{R}^1_1$, $T*T\in\mathscr{R}^1_0$, hence $(\mathscr{F}_1T)^2=\mathscr{F}_0(T*T)$ goes to zero at infinity. The second part follows from (4) and $\mathscr{R}^1_1\subset\mathscr{R}^2_1$.

(7) Necessity is proven by considering $\mathrm{tr}\,T(\Sigma_j\zeta_jE_i(x_j))^*(\Sigma_k\zeta_kE_i(x_k))\geqslant 0$ for $i=0,1$. Let $f_0=0$ and f_1 continuous and "twisted positive definite." Let $S\in\mathscr{R}^1_1$, $S\geqslant 0$. Then for any choice of $\{x_1\cdots x_n\}$, the matrix

$$\begin{aligned}M_{jk}&=(\mathscr{F}_1S)(x_j-x_k)f_1(x_k-x_j)\\&=((\mathscr{F}_1S)(x_j-x_k)e^{+i/2\{x_k,x_j\}}\\&\quad\times(f_1(x_k-x_j)e^{-i/2\{x_k,x_j\}})\end{aligned}$$

is positive definite as the elementwise product of positive-definite matrices. Hence by the classical Bochner theorem there is a positive $g\in\mathscr{R}^1_0$ with $\mathscr{F}_0g(x)=(\mathscr{F}_1S)(x)f_1(-x)$. The map $S\mapsto g$ is linear, positive, and α-covariant; hence, by the proof of 3.2, there is a unique $T\in\mathscr{R}^1_1$ such that $g=S*(\beta_{-}T)$. Clearly the Fourier transform of T must be f_1. □

Since the Fourier transforms of operators and functions are both functions on X, one may ask how operators and functions with "the same" transform are related. More formally, define the map $\mathscr{J}:\mathscr{B}\to\mathscr{B}$, $(\mathscr{J}f)_i(x)=f_{i+1}(x)$ and consider $\mathscr{W}:=\mathscr{F}^{-1}\mathscr{J}\mathscr{F}$. On $\mathscr{R}^2$, $\mathscr{W}$ is unitary and it takes each density matrix $T\in\mathscr{R}^1_1$ into its "Wigner quasiprobability function" [16] (which is in general neither positive nor integrable). There is a rich literature on this subject, in particular, in connection with semiclassical limits of quantum mechanics. It is a natural extension of the above approach to study the quantum analogs of test function and distribution spaces. (The analogs of Paley–Wiener theorems would have their natural place in this setting.) For results in this direction, the reader is referred to Refs. 17, 18, and references quoted there.

The following proposition is the basis of the results of Sec. IV.

Proposition 3.5: Wiener's approximation theorem: Let $k\in\{0,1\}$ and $T\in\mathscr{R}^1_k$. Then the following five conditions are equivalent ($i=0,1$):

(1) $\forall_{x\in X}(\mathscr{F}_kT)(x)\neq 0$;

(2i) $T*\mathscr{R}^1_i$ is $\|\cdot\|_1$ dense in $\mathscr{R}^1_{i+k}$;

(3i) $A\in\mathscr{R}^\infty_i$ and $T*A=0$ imply $A=0$.

If these conditions are satisfied, T will be called *regular*.

Proof: (2i)$\Leftrightarrow$(3$i+k$): The linear space $T*\mathscr{R}^1_i$ is norm-dense in $\mathscr{R}^1_{i+k}$ iff it is $\sigma(\mathscr{R}^1_{i+k},\mathscr{R}^\infty_{i+k})$-dense iff $T*\mathscr{R}^1_i$ separates points of $\mathscr{R}^\infty_{i+k}$ iff the adjoint of $S\mapsto T*S$ for the pairings $\langle\mathscr{R}^1_i,\mathscr{R}^\infty_i\rangle$ and $\langle\mathscr{R}^1_{i+k},\mathscr{R}^\infty_{i+k}\rangle$ is injective. This adjoint is the map $A\mapsto(\beta_{-}T)*A:\mathscr{R}^\infty_{i+k}\to\mathscr{R}^\infty_i$. Thus (2$i$) is equivalent to (3$i+k$) with $\beta_{-}T$ substituted for T. By Proposition 3.2(2) this is equivalent to (3$i+k$).

(2k)$\Rightarrow$(2$i+k$): Let $S\in\mathscr{R}^1_{i+k}$. Since α is strongly continuous on $\mathscr{R}^1$, there is an $f\in\mathscr{R}^1_0$ such that $\|f*S-S\|_1\leqslant\epsilon_1$. By (2$k$) there is some $R\in\mathscr{R}^1_k$ such that $\|T*R-f\|_1\leqslant\epsilon_2$. Hence $\|T*(R*S)-S\|_1\leqslant\|(T*R-f)*S\|_1+\|f*S-S\|_1\leqslant\epsilon_1+\epsilon_2\|S\|_1$.

(3i)$\Rightarrow$(1): $T*E_i(x)=\Sigma_jE_{i+j}(x)\cdot(\mathscr{F}_jT)(-x)=E_{i+k}(x)\cdot(\mathscr{F}_kT)(-x)$. By (3$i$) the left-hand side is nonzero; hence $(\mathscr{F}_kT)(-x)\neq 0$.

(1)$\Rightarrow$(2k): $(\mathscr{F}_0T*T)(x)=(\mathscr{F}_kT)^2(x)\neq 0$. Hence by the classical theorem of Wiener,[19] $T*T*\mathscr{R}^1_0$ and hence $T*\mathscr{R}^1_k\supset T*(T*\mathscr{R}^1_0)$ is norm-dense in $\mathscr{R}^0_1$. □

The covariant observables $\sigma \to T*\chi_\sigma$ with $T \in \mathscr{R}^1_1$, $T \geqslant 0$, tr $T = 1$, and T regular have the interesting property that their measurement, i.e., the expectation values $\langle W, T*\chi_\sigma \rangle$ for $\sigma \subset X$, determine the density matrix W completely. The following example, which generalizes the coherent states, shows that such trace-class operators T exist.

A trivial consequence of the existence of a regular $T \in \mathscr{R}^1_1$ is that the set $\mathscr{R}^1_1 * \mathscr{R}^1_1 \supset T * \mathscr{R}^1_1$ is dense in $\mathscr{R}^1_0$. Note, however, that by Lemma 3.4(6) the inclusion $\mathscr{R}^1_1 * \mathscr{R}^1_1 \subset \mathscr{R}^1_0$ is strict, in contrast to the equalities $\mathscr{R}^1_0 * \mathscr{R}^1_i = \mathscr{R}^1_i$, which can be proven by means of Ref. 20, 32.50.

Example 3.6: *Gaussians*: With the notations of Sec. II, let $h(x) = h(p,q) := \frac{1}{2}\sum_{k=1}^{N}(p_k^2 + q_k^2)$. (By choice of a suitable symplectic coordinate system, every positive quadratic form $\tilde{h}(x) = \frac{1}{2}\{x, jx\}$ for some $j \in G, j^2 = -1$ can be written in this form). Let $H = \frac{1}{2}\sum_{k=1}^{N}(p_k^2 + Q_k^2)$ be the Hamiltonian operator of the corresponding quantum oscillator. For $z \in \mathbb{C}$, Re $z > 0$, define a Gaussian function $g(z) := \exp(-zh)$ and a Gaussian operator $G(z) = \exp(-zH)$. Then

$$(\mathscr{F}_0 g(z))(x) = z^{-N} \cdot \exp(-(1/z)h(x))$$

and

$$(\mathscr{F}_1 G(z))(x) = (2\sinh(z/2))^{-N} \cdot \exp(-\tfrac{1}{2}\coth(z/2) \cdot h(x)).$$

In particular, all Gaussians are regular in the sense of Proposition 3.5.

Proof: By considering tensor products, the proof is reduced to the case $N = \frac{1}{2}\dim X = 1$. Let $\psi_n(\xi) = \nu_n H_n(\xi)\exp(-\frac{1}{2}\xi^2)$ be the nth eigenfunction of H and $R_n = |\psi_n\rangle\langle\psi_n|$ the corresponding eigenprojection. If $g \in G$ leaves h invariant, then $\beta_g R_n = R_n$; hence $(\mathscr{F}_1 R_n)(x)$ depends only on $h(x)$ by Proposition 3.3(2). Thus it suffices to compute $(\mathscr{F}_1 R_n)(p,0)$. By Ref. 21, 22.13.20,

$$\begin{aligned}(\mathscr{F}_1 R_n)(p,0) &= \nu_\eta^2 \int d\xi\, e^{-\xi^2} H_n(\xi)^2 \cos(p\xi) \\ &= L_n(\tfrac{1}{2}p^2)\cdot\exp(-\tfrac{1}{4}p^2) \\ &= L_n(h(p,0))\exp(-\tfrac{1}{2}h(p,0)).\end{aligned}$$

Using the generating function of the Laguerre polynomials L_n (Ref. 21, 22.9.15) with $\zeta = e^{-z}$, we obtain

$$\begin{aligned}(\mathscr{F}_1 G(z)) &= \sum_{n=0} \exp\left(-z\left(n+\frac{1}{2}\right)\right)\mathscr{F}_1 R_n \\ &= \zeta^{1/2}\exp\left(-\frac{1}{2}h\right)\sum_{n=0}^{\infty}\zeta^n L_n(h) \\ &= \frac{\zeta^{1/2}}{1-\zeta}\cdot\exp\left(-\frac{1}{2}h - \frac{h\zeta}{1-\zeta}\right) \\ &= \left(2\sinh\left(\frac{z}{2}\right)\right)^{-1}\exp\left(-\frac{1}{2}\coth\left(\frac{z}{2}\right)h\right). \quad \square\end{aligned}$$

For small z (Gaussians of large dispersion) $\mathscr{F}_1 G(z)$ and $\mathscr{F}_0 g(z)$ are asymptotically equal. For $z \to \infty$ along the real axis, $z^N \cdot \mathscr{F}_0 g(z) \to 1$, so that the normalized Gaussian $z^N g(z)$ tends to a δ function. In the same limit $(\mathrm{tr}(G(z))^{-1} G(z)$ tends to the coherent state R_0. On on other hand, $\lim_{z \to i\pi} 2^N (\mathscr{F}_1 G(z))(x) = 1$, so that the operator $2^N G(i\pi)$ has formally the same transform as the δ function. $G(i\pi) = U_-$ is the unitary operator of phase space inversion.[22]

IV. CORRESPONDENCE OF INVARIANT SUBSPACES

In Proposition 3.2(5) we proved the relations $\mathscr{R}^1_1 * \mathscr{R}^p_0 \subset \mathscr{R}^p_1$ and $\mathscr{R}^1_1 * \mathscr{R}^p_1 \subset \mathscr{R}^p_0$. With the help of Proposition 3.5, it is easy to prove that both inclusions are dense, so that each of the spaces $\mathscr{R}^p_i$ "determines" the other. In this section this relationship will be extended to more general pairs of spaces of classical and quantum observables.

Definition: Let $1 \leqslant p \leqslant \infty$ and $\mathscr{D} \subset \mathscr{R}^p$ a linear subspace. Then $\mathscr{D}$ is called a *pair* if it is of the form $\mathscr{D} = \mathscr{D}_0 \oplus \mathscr{D}_1$ with $\mathscr{D}_i \subset \mathscr{R}^p_i$ and $\mathscr{R}^1_1 * \mathscr{D} \subset \mathscr{D}$. In this case $\mathscr{D}_0$ and $\mathscr{D}_1$ are called *corresponding spaces*. A pair $\mathscr{D} \subset \mathscr{R}^\infty$ is called *continuous* if $\mathscr{D} \subset \mathscr{C}$. (Recall that $\mathscr{C}$ was defined as the space of strongly continuous elements for α.)

Since $\mathscr{R}^1_1 * \mathscr{D}_0 \subset \mathscr{D}_1$ and $\mathscr{R}^1_1 * \mathscr{D}_1 \subset \mathscr{D}_0$, each of the spaces $\mathscr{D}_i$ is stable under convolutions with $\mathscr{R}^1_1 * \mathscr{R}^1_1$, which is a dense subset of $\mathscr{R}^1_0$. Thus both spaces come close to being α-invariant. For the closed pairs in which we will be mostly interested, α invariance and the relation $\mathscr{R}^1 * \mathscr{D} \subset \mathscr{D}$ are a consequence of the above defintion, so that for the purposes of this section we may replace the condition $\mathscr{R}^1_1 * \mathscr{D} \subset \mathscr{D}$ by these two stronger conditions.

Our main result concerning corresponding spaces is the following:

Theorem 4.1: Let $\mathscr{D} \subset \mathscr{R}^\infty$ be a pair and $\overline{\mathscr{D}} = \overline{\mathscr{D}}_0 \oplus \overline{\mathscr{D}}_1$ its $\|\cdot\|_\infty$-closure. Then

(1) $\overline{\mathscr{D}}$ is a pair.

(2) Let $i = 0,1$ and $T \in \mathscr{R}^1_1$ be regular. Then $T * \mathscr{D}_i$ is $\|\cdot\|_\infty$-dense in $\mathscr{D}_{i+1} \cap \mathscr{C}$.

(3) Let $i, j = 0,1$, $T \in \mathscr{R}^1_i$ regular, and $A \in \mathscr{C}_j$. Then $T * A \in \mathscr{D}_{i+j}$ implies $A \in \overline{\mathscr{D}}_j$.

(4) For any $\|\cdot\|_\infty$-closed α-invariant subspace $\mathscr{E}_i \subset \mathscr{C}_i$ there is a unique corresponding closed invariant subspace $\mathscr{E}_{i+1} \subset \mathscr{C}_{i+1}$.

(5) For $1 \leqslant p < \infty$, this theorem remains true, when in hypotheses and conclusions $\|\cdot\|_\infty$ is replaced by $\|\cdot\|_p$ and $\mathscr{R}^\infty$ and $\mathscr{C}$ are both replaced by $\mathscr{R}^p$.

Proof: Let $1 \leqslant p \leqslant \infty$.

(1) $\|T*A\|_p \leqslant \|T\|_1 \|A\|_p$. Hence $A_n \to A \in \overline{\mathscr{D}}_i$ implies $T*A = \lim_n T*A_n \in \overline{\mathscr{D}}_{i+1}$.

(2) $T * \mathscr{D}_i \subset \mathscr{D}_{i+1}$ by correspondence and $T * \mathscr{D}_i \subset \mathscr{C}$ by Proposition 3.2(4). Let $A \in \mathscr{D}_{i+1} \cap \mathscr{C}$. Since A is strongly continuous, there is an $f \in \mathscr{R}^1_0$ with $\|f*A - A\|_p \leqslant \epsilon_1$. Since T is regular, there is an $S \in \mathscr{R}^1_1$ such that $\|f - T*S\|_1 \leqslant \epsilon_2$. Hence $\|A - T*(S*A)\|_p \leqslant \epsilon_1 + \|A\|_p \epsilon_2$ and $S*A \in \mathscr{D}_i$ by correspondence.

(3) For $i = 1$ pick $f \in \mathscr{R}^1_0$ and $S \in \mathscr{R}^1_1$ as in (2). Then $\|A - S*(T*A)\|_p \leqslant \epsilon$ and $S*T*A \in S*\mathscr{D}_{i+j} \subset \mathscr{D}_j$. For $i = 0$ we may use the same argument, choosing S in the $\|\cdot\|_1$-dense set $\mathscr{R}^1_1 * \mathscr{R}^1_1 \subset \mathscr{R}^1_0$, so that $S*T*A \in S*\mathscr{D}_{0+j} \subset \mathscr{R}^1_1 * \mathscr{R}^1_1 * \mathscr{D}_j \subset \mathscr{R}^1_1 * \mathscr{D}_{j+1} \subset \mathscr{D}_j$.

(4) For $A \in \mathscr{C}$ and $T \in \mathscr{R}^1_0$, $T*A$ can be approximated in norm by linear combinations of translates of A. Hence $\mathscr{R}^1_0 * \mathscr{E}_i \subset \mathscr{E}_i$. Define the two spaces $\mathscr{E}^-_{i+1} := \mathrm{lin}\, \mathscr{R}^1_1 * \mathscr{E}_i$ and $\mathscr{E}^+_{i+1} := \{A \in \mathscr{C}_{i+1} \mid \mathscr{R}^1_1 * A \in \mathscr{E}_i\}$. $\mathscr{E}^+_{i+1}$ is closed as the

intersection of closed subspaces and some space $\mathscr{C}_{i+1} \subset \mathscr{C}_{i+1}$ corresponds to $\mathscr{C}_i$ iff $\mathscr{C}^-_{i+1} \subset \mathscr{C}_{i+1} \subset \mathscr{C}^+_{i+1}$. Since $\mathscr{R}^1_1 * \mathscr{C}^-_{i+1} \subset \text{lin}\, \mathscr{R}^1_1 * \mathscr{R}^1_1 * \mathscr{C}_i \subset \text{lin}\, \mathscr{R}^1_0 * \mathscr{C}_i \subset \mathscr{C}_i$, $\mathscr{C}^\pm_{i+1}$ satisfy this condition. Applying (3) to the pair $\mathscr{C}_i, \mathscr{C}^-_{i+1}$ and $A \in \mathscr{C}^+_{i+1}$, we conclude that $\mathscr{C}^-_{i+1}$ is dense in $\mathscr{C}^+_{i+1}$. □

The next corollary serves to emphasize the result Theorem 4.1(4):

Corollary 4.2: For $i = 0,1$, and $1 \leqslant p < \infty$ (resp. $p = \infty$) let $\mathscr{L}^p_i$ (resp. $\mathscr{L}^c_i$) denote the set of $\|\cdot\|_p$ -closed α-invariant subspaces of $\mathscr{R}^p_i$ (resp. $\mathscr{C}_i$) and $\mathscr{L}^p$ (resp. $\mathscr{L}^c$) the set of closed pairs $\mathscr{D} = \mathscr{D}_0 \oplus \mathscr{D}_1$ with $\mathscr{D}_i \in \mathscr{L}^p_i$ (resp. $\mathscr{L}^c_i$). Then:

(1) Correspondence is a lattice isomorphism between the complete lattices $\mathscr{L}^p_0, \mathscr{L}^p_1$, and $\mathscr{L}^p$ (resp. $\mathscr{L}^c_0, \mathscr{L}^c_1, \mathscr{L}^c$).

(2) Let $1 < p,q < \infty$ and $p^{-1} + q^{-1} = 1$. For $\mathscr{D} \subset \mathscr{L}^p$ define $\mathscr{D}^\perp := \{A \in \mathscr{R}^q \mid \forall_{B \in \mathscr{D}}\ \text{tr}\, AB = 0\}$. Then $\mathscr{D}^\perp \in \mathscr{L}^q$ and $\perp : \mathscr{L}^p \to \mathscr{L}^q$ is a lattice anti-isomorphism.

Proof: (1) by Theorem 4.1(4) correspondence is a bijective relation and $\mathscr{L}_i$ may be identified with $\mathscr{L}$. Let $\{\mathscr{D}^\iota\}_{\iota \in I}$ be a family of pairs. Then $\cap_{\iota \in I} \mathscr{D}^\iota_0$ and $\cap_{\iota \in I} \mathscr{D}^\iota_1$ are corresponding. Similarly, $\Sigma_{\iota \in I} \mathscr{D}^\iota$ is a pair, and by Theorem 4.1(1) the closure of this space, which is the lattice union of the $\mathscr{D}^\iota$, is also a pair.

(2) Since $\mathscr{R}^p$ and $\mathscr{R}^q$ are duals, $\perp$ is an anti-isomorphism for the lattices of all norm closed subspaces, so that we only have to prove that $\mathscr{D}^\perp$ is a pair. Clearly, $\mathscr{D}^\perp = \mathscr{D}^\perp_0 \oplus \mathscr{D}^\perp_1$ and $T \in \mathscr{R}^1_1$, $A \in \mathscr{D}^\perp$, $B \in \mathscr{D}$ imply $\text{tr}(T * A)B$-$= \text{tr}A(\beta_- T * B) = 0$ since $\beta_- T * \mathscr{D} \subset \mathscr{D}$. □

The lattice $\mathscr{L}^2$ will be computed in Example 5.4.

Especially interesting for applications in physics are the order unit subspaces (closed self-adjoint subspaces containing 1) of $\mathscr{R}^\infty_0$ and $\mathscr{R}^\infty_1$, since these are eligible as spaces of classical and quantum observables.[7] Theorem 4.1 shows that the spaces on whch α is strongly continuous are in one-to-one correspondence. The following corollary describes some properties which are preserved by this correspondence.

Corollary 4.3: Let $\mathscr{D}_0 \subset \mathscr{C}_0$ and $\mathscr{D}_1 \subset \mathscr{C}_1$ be closed corresponding spaces and $i = 0,1$. Suppose that $\mathscr{D}_i$ is: self-adjoint; an order unit subspace; separable; containing $E_i(x)$ for some $x \in X$; pointwise fixed under α_x for some $x \in X$; invariant under a subgroup $G_0 \subset G$; strongly continuous for a subgroup $G_0 \subset G$; generated by translates of a single element; generated by translates of elements invariant under a compact subgroup $G_0 \subset G$; $\sigma(\mathscr{R}^\infty_i, \mathscr{R}^1_i)$-dense in $\mathscr{R}^\infty_i$. Then $\mathscr{D}_{i+1}$ has the same property. (Different properties are separated by ";".)

Proof: By Theorem 4.1(2) $\mathscr{D}_{i+1}$ is the norm closure of $T * \mathscr{D}_i$, where T is some regular element of $\mathscr{R}^1_1$. Taking T to be the average of a Gaussian with positive Fourier transform over a compact subgroup G_0, we may assume T to be G_0-invariant. Recall that β is strongly continuous on $\mathscr{R}^1$ and $T * E_i(x) = E_{i+1}(x) \cdot (\mathscr{F}_1 T)(-x)$. The last statement is contained in 4.4. □

The relationship between the order properties of $\mathscr{D}_0$ and $\mathscr{D}_1$ is a much more complicated problem: we have made essential use of the fact that $T_1 * T_2$ with $T_1, T_2 \in \mathscr{R}^1_1$ can be chosen "close to a δ function." However, this is no longer true, when T_1, T_2 are required to be positive. It would be particularly interesting with what qualifications (if any) "a C^*-algebra" can be added to the above list.

The following result describes the correspondence theory for weak*-closed subspaces.

Corollary 4.4: (1) The map $\perp$ defined in Corollary 4.2 is an anti-isomorphism between the lattice $\mathscr{L}^1$ and the lattice $\mathscr{L}^\infty$ of $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-closed pairs. The $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-closure of any pair $\mathscr{D} \subset \mathscr{R}^\infty$ is $\mathscr{D}^{\perp\perp}$.

(2) Theorem 4.1 (1)–(4) remains valid when the norm topology is replaced by the $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$ topology and condition "$A \in \mathscr{C}_j$" in Theorem 4.1(3) is replaced by $A \in \mathscr{R}^\infty_j$.

(3) For any pair $\mathscr{D}$: $(\mathscr{D} \cap \mathscr{C})^{\perp\perp} = \mathscr{D}^{\perp\perp}$.

(4) For any pair $\mathscr{D}$, the four sets

$S_{ij} := \{x \in X \mid (T \in \mathscr{R}^1_i \wedge \forall_{A \in \mathscr{D}_j} (\beta_- T) * A = 0) \Rightarrow (\mathscr{F}_i T)(x) = 0\}$ coincide. This set is closed and is called the spectrum of $\mathscr{D}$, denoted by $\text{sp}(\mathscr{D})$. Moreover, $\text{sp}(\mathscr{D}) = \{x \in X \mid E_i(x) \in \mathscr{D}^{\perp\perp}_i\}$ for $i = 0,1$.

(5) If $\mathscr{D}$ is a $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-closed pair, then $\mathscr{D}_0$ is a C^*-algebra (and hence a W^*-algebra) iff $\mathscr{D}_1$ is iff $\text{sp}(\mathscr{D})$ is a closed subgroup of X. For every closed subgroup $\tilde{X} \subset X$ there is a unique σ-closed pair with spectrum $\tilde{X}$.

Proof: (1) is proven as in Corollary 4.2.

(2),(3) Theorem 4.2(1) and (4) are immediate consequences of (1). If T is regular $(T * \mathscr{D}_i)^\perp$ $= \{S \in \mathscr{R}^1_{i+1} \mid \langle \beta_- T * S, \mathscr{D}_i \rangle = 0\} = \{S \mid \beta_- T * S \in \mathscr{D}^\perp_i\}$ $= \mathscr{D}^\perp_{i+1}$ by Theorem 4.1(3) with $p = 1$. Hence $T * \mathscr{D}_i$ is σ-dense in $\mathscr{D}^{\perp\perp}_{i+1}$. On the other hand, $T * \mathscr{D}_i \subset \mathscr{D}_{i+1} \cap \mathscr{C}$. This also proves (3). Now let $T \in \mathscr{R}^1_i$ be regular. Then by Theorem 4.1(2) ($p = 1$) $\beta_- T * \mathscr{D}^\perp_{i+j}$ is $\|\cdot\|_1$-dense in $\mathscr{D}^\perp_j$. Hence $T * A \in \mathscr{D}_{i+j}$ implies $A \in (\beta_- T * \mathscr{D}^\perp_{i+j})^\perp = \mathscr{D}^{\perp\perp}_j$.

(4) Let $i,j,k,l \in \{0,1\}$ and suppose $x \in S_{ij}$, $T \in \mathscr{R}^1_k$, and, $\forall\, A \in \mathscr{D}_l, \beta_- T * A = 0$. Pick regular elements $T_1 \in \mathscr{R}^1_{i+k}$ and $T_2 \in \mathscr{R}^1_{j+l}$. Then, for $A \in \mathscr{D}_l, \beta_- T_1 * \beta_- T * T_2 * A = 0$ and, by the $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-continuity of convolutions and the density of $T_2 * \mathscr{D}_l$ in $\mathscr{D}^{\perp\perp}_j \subset \mathscr{D}_j$, $\beta_-(T_1 * T) * B = 0$ for $B \in \mathscr{D}_j$. Since $x \in S_{ij}$, $\mathscr{F}_i(T_1 * T)(x) = 0 = (\mathscr{F}_{i+k} T_1)(x)(\mathscr{F}_k T)(x)$. Hence $(\mathscr{F}_k T)(x) = 0$ by the regularity of T_1. The spectrum is closed as the intersection of zero sets of continuous functions and $S_{ii} = \{x \in X \mid T \in \mathscr{D}^\perp_i \Rightarrow \langle T, E_i(x) \rangle = 0\}$.

(5) If $\mathscr{D}_k (k = 0,1)$ is a $*$-algebra and $x,y \in \text{sp}(\mathscr{D})$, then $E_k(x + y) = \exp((ik/2)\{x,y\}) E_k(x) E_k(y) \in \mathscr{D}_k$ and $E_k(x)^* = E_k(-x) \in \mathscr{D}_k$. Hence $\text{sp}(\mathscr{D})$ is a subgroup. On the other hand, let $\tilde{X} \subset X$ be a closed subgroup. Then the linear span $\tilde{\mathscr{D}}_k$ of $\{E_k(x) \mid x \in \tilde{X}\}$ is a $*$-algebra and $\tilde{\mathscr{D}} = \tilde{\mathscr{D}}_0 \oplus \tilde{\mathscr{D}}_1$ is a pair. Hence $\tilde{\mathscr{D}}^{\perp\perp}$ is a W^*-algebra and $\text{sp}\tilde{\mathscr{D}} = \tilde{X}$. It remains to be proven that there is only one σ-closed pair with spectrum $\tilde{X}$, or equivalently there is only one closed translation invariant subspace $\mathscr{D}^\perp_0 \subset \mathscr{R}^1_0 = \mathscr{L}^1(X)$ whose Fourier transforms vanish precisely on $\tilde{X}$, or in the terminology of spectral synthesis that closed subgroups are spectral sets. This is shown in Ref. 20, 40.24. □

The investigation of the lattice $\mathscr{L}^1_0$ (or equivalently the lattices $\mathscr{L}^1$ and $\mathscr{L}^\infty$) is the main problem of "spectral synthesis."[23] The "failure of spectral synthesis in $\mathbb{R}^{2N}$" (Ref. 20, 42.19) is equivalent in our context to the statement that there are $\sigma(\mathscr{R}^\infty_1, \mathscr{R}^1_1)$-closed α-invariant subspaces in

$\mathscr{R}_1^\infty = \mathscr{B}(\mathscr{H})$ which are not generated by the Weyl operators they contain.

V. EXAMPLES OF CORRESPONDING SPACES

A. Compacts

Let $\mathscr{K}_0$ be the set of continuous functions vanishing at infinity and $\mathscr{K}_1$ the set of compact operators. Then $\mathscr{K} = \mathscr{K}_0 \oplus \mathscr{K}_1$ is a pair: Since $\mathscr{R}_0^1 \cap \mathscr{K}_0$ is $\|\cdot\|_\infty$-dense in $\mathscr{K}_0$, the corresponding space of $\mathscr{K}_0$ is the $\|\cdot\|_\infty$-closure of $\mathscr{R}_1^1 * (\mathscr{R}_0^1 \cap \mathscr{K}_0)$. Clearly this space is $\|\cdot\|_1$-dense in $\mathscr{R}_1^1$ and hence $\|\cdot\|_\infty$-dense in $\mathscr{K}_1$. Note that the uniqueness statement of Theorem 4.1(4) applies only to subspaces of $\mathscr{C}$. For example, the space of all bounded measurable functions vanishing at infinity is also a corresponding space for $\mathscr{K}_1$. For illustrative purposes it may be useful to state one of the typical corollaries of the theory in Sec. IV in standard notation:

Corollary 5.1: Let A be a bounded operator. Then the following are equivalent: (1) A is compact. (2) $\lim_{x\to 0} \|\alpha_x A - A\| = 0$, and, for some regular trace class operator T, $x \mapsto \mathrm{tr}(T(\alpha_x A))$ vanishes at infinity. (3) Condition (2) holds for all trace-class operators T. (4) A can be approximated in norm by operators of the form $B_f = \int dx\, f(x) E_1(x)$ with $f \in \mathscr{L}^1(X)$.

Proof: The equivalences (1)$\Leftrightarrow$(2)$\Leftrightarrow$(3) are a restatement of Theorem 4.1. To prove the last equivalence, note that $\mathscr{D}_i := \{ \int dx\, f(x) E_i(x) \,|\, f \in \mathscr{L}^1(X) \}$ are corresponding spaces. Since the algebra of Fourier transforms of $\mathscr{L}^1$ functions separates points, $\bar{\mathscr{D}}_0 = \mathscr{K}_0$; hence $\bar{\mathscr{D}}_1 = \mathscr{K}_1$ by uniqueness of corresponding spaces. □

By restricting the support of f in Corollary 5.1(4) to some closed subset of X, we obtain corresponding proper subspaces of $\mathscr{K}_0$ and $\mathscr{K}_1$. However, $\mathscr{K}_0$ and $\mathscr{K}_1$ are both minimal in the lattices of α-invariant C^*-subalgebras of $\mathscr{R}_0^\infty$ and $\mathscr{R}_1^\infty$.

Proof: (0) Suppose $x_1, x_2 \in X$ are not separated by translates of some function f, i.e., $\forall_x f(x_1 - x) = f(x_2 - x)$. Then $\alpha_{x_1 - x_2} f = f$, which is impossible for $x_1 \neq x_2$ and $f \in \mathscr{K}_0$.

(1) If $\mathscr{D}_1$ is an α-invariant proper C^*-subalgebra of $\mathscr{K}_1$, $\mathscr{D}_1^1 \neq \{0\}$ since $\mathscr{R}_1^1$ is the dual of $\mathscr{K}_1$. Hence the $\sigma(\mathscr{R}^\infty, \mathscr{R}^1)$-closure of $\mathscr{D}_1$ is a proper subalgebra of $\mathscr{R}_1^\infty$. Thus the commutant $\mathscr{D}_1'$ of $\mathscr{D}_1$ is a nontrivial α-invariant von Neumann algebra and by Corollary 4.4(5) contains a Weyl operator $E_1(x)$ with $x \neq 0$. Hence $\mathscr{D}_1$ is pointwise fixed under α_x contradicting $\lim_{n\to\infty} \mathrm{tr}\, T(\alpha_{n\cdot x} A) = 0$ for $A \in \mathscr{K}_1$. □

B. Differentiable elements

For $k \in \mathbb{N}$ let $\mathscr{C}^k \subset \mathscr{R}^\infty$ be the space of elements whose derivatives up to order k exist and are strongly continuous. For $k = \infty$ let $\mathscr{C}^\infty = \cap_{k\in\mathbb{N}} \mathscr{C}^k$. For any finite set $Y = \{ y_1 \cdots y_n \}$, D_Y will denote the formal differentiation operator $D_Y = \lim_{\{t_i\}\to 0} \Pi_{i=1}^n t_i^{-1}(1 - \alpha_{t_i y_i})$. Thus $A \in \mathscr{C}^k$ means that for any finite set Y with $|Y| \leqslant k$ there is some $\tilde{A} \equiv D_Y A \in \mathscr{C}$ such that $\lim_{\{t_i\}\to 0} \|(\Pi_i t_i^{-1}(1 - \alpha_{t_i y_i}))A - \tilde{A}\|_\infty = 0$. The seminorms $A \mapsto \|D_Y A\|_\infty$ define the locally convex topology of $\mathscr{C}^k$.

Since $\alpha_x(T*A) = T*(\alpha_x A)$ and $\|T*A\|_\infty \leqslant \|T\|_1 \|A\|_\infty$, it is clear that $\mathscr{R}^1 * \mathscr{C}^k \subset \mathscr{C}^k$ and $D_Y(T*A) = T*D_Y A$ for $|Y| \leqslant k$, $T \in \mathscr{R}^1$, and $A \in \mathscr{C}^k$. In particular, the maps $A \mapsto T*A$ are continuous in the $\mathscr{C}^k$-topology. Thus we arrive at the following:

Corollary 5.2: Let $1 \leqslant k \leqslant \infty$. Then Theorem 4.1 (1)–(4) remains valid for pairs $\mathscr{D} \mapsto \mathscr{C}^k$, when $\mathscr{C}$ is replaced by $\mathscr{C}^k$ and topological notions are referred to the $\mathscr{C}^k$-topology.

Proof: In the proof of Theorem 4.1 substitute $\|A\|_p$ by $\max\{ \|D_Y A\|_\infty \,|\, Y \in \mathscr{Y} \} = \|A\|_{\mathscr{Y}}$, where $\mathscr{Y}$ is an arbitrary finite collection of finite sets Y with $|Y| \leqslant k$. If $A \in \mathscr{C}^k$, $\{ D_Y A \,|\, Y \in \mathscr{Y} \}$ is a finite subset of $\mathscr{C}$, so that we may find one $f \in \mathscr{R}_0^1$ with $\|f*A - A\|_{\mathscr{Y}} \leqslant \epsilon_1$. With the obvious further substitutions the proof of Theorem 4.1 applies. □

A glance at the classical situation shows that in general the correspondence for spaces of test functions $\oplus$ test operators cannot be defined by $\mathscr{R}_1^1 * \mathscr{D} \subset \mathscr{D}$. Instead, the space $\mathscr{R}_1^1$ should be replaced by some subspace, e.g., the space of Schwartz operators $\mathscr{S}_1 = \{ A \in \mathscr{R}_1^1 \,|\, \mathscr{F}_1 A \in \mathscr{S}(X) \}$. Clearly $\mathscr{S}_1$ contains regular elements, so that the theory of Sec. IV remains unchanged. The usefulness of this correspondence depends on efficient intrinsic characterizations of $\mathscr{S}_1$.

C. Functions of position

Let $X = \hat{\Xi} \oplus \Xi$ be a decomposition of X into momentum and position space as described in Sec. II, so that $\mathscr{H} = \mathscr{L}^2(\Xi)$. Then we can consider a function $f \in \mathscr{L}^\infty(\Xi)$ either as the classical observable $\pi_0^* f \in \mathscr{R}_0^\infty$ with $(\pi_0^* f)(p,q) := f(q)$ or as the quantum observable $\pi_1^* f \in \mathscr{R}_1^\infty$ with $[(\pi_1^* f)\psi](q) := f(q)\psi(q)$. This map π_i^* takes $\mathscr{L}^\infty(\Xi)$ isomorphically onto a $\sigma(\mathscr{R}_i^\infty, \mathscr{R}_i^1)$-closed α-invariant algebra $\mathscr{D}_i \subset \mathscr{R}_i^\infty$, which can be characterized intrinsically as $\mathscr{D}_i = \{ A \in \mathscr{R}_i^\infty \,|\, x \in \hat{\Xi} \Rightarrow \alpha_x A = A \}$. Hence $\mathscr{D} = \mathscr{D}_0 \oplus \mathscr{D}_1$ is the weak*-closed pair with spectrum $\hat{\Xi}$.

Since π_i^* is normal, it has a preadjoint $\pi_i : \mathscr{R}_i^1 \to \mathscr{L}^1(\Xi)$, which associates with each state $T \in \mathscr{R}_i^1$ its distribution of position $\pi_i T \in \mathscr{L}^1(\Xi)$. [Explicitly: $(\pi_0 g)(q) = (2\pi)^{-N} \times \int dp\, g(p,q)$ and $(\pi_1 \Sigma \lambda_i |\varphi_i\rangle\langle\psi_i|)(q) = \Sigma \lambda_i\, \overline{\varphi_i(q)} \psi_i(q)$. With the help of the relation $(\pi_i \alpha_{(p,q)} T)(q') = (\pi_i T)(q' - q)$ one easily proves that $\pi = \pi_0 + \pi_1 : \mathscr{R}^1 \to \mathscr{L}(\Xi)$ is a convolution homomorphism, i.e., $\pi(T_1 * T_2) = (\pi T_1) * (\pi T_2)$. When $\mathscr{D}_0$ and $\mathscr{D}_1$ are identified with $L^\infty(\Xi)$ through the isomorphisms π_i^*, the map $A \mapsto T*A$ for $A \in \mathscr{D}_i$, $T \in \mathscr{R}_j^1$ becomes convolution with πT.

Let $h \in \mathscr{L}^1(\hat{\Xi})$ such that $(2\pi)^{-N} \int dp\, h(p) = 1$. Then the map $\sigma_0 : \mathscr{L}^1(\Xi) \to \mathscr{R}_0^1$ given by $(\sigma_0 g)(p,q) = h(p) g(q)$ is norm-continuous and satisfies $\pi_0 \sigma_0 = \mathrm{id}$. This map has no quantum mechanical analog σ_1: If this map existed, $\mathscr{L}^1(\Xi)$ would inherit the Radon–Nikodym property from $\mathscr{R}_1^1$, which holds for $\mathscr{L}^1(\Omega, \Sigma, \mu)$ only if the measure space (Ω, Σ, μ) is atomic. Equivalently, there are no normal projections from $\mathscr{R}_1^\infty$ onto $\mathscr{D}_1$. (Nonnormal projections are easily constructed from invariant means on $\hat{\Xi}$.)

D. The lattice $\mathscr{L}^2$

For any $\|\cdot\|_2$-closed pair $\mathscr{D} \subset \mathscr{R}^2$, let $p_{\mathscr{D}}$ be the orthogonal projection from the Hilbert space $\mathscr{F}_1 \mathscr{R}_1^2 \equiv \mathscr{L}^2(X)$ to

its closed subspace $\mathscr{F}_i\mathscr{D}_i$. (The choice $i=0$ or 1 is inessential.) The condition of α-invariance for $\mathscr{D}_i$ is equivalent to $[p_{\mathscr{D}}, \mathscr{F}_i\alpha_x\mathscr{F}_i^{-1}]=0$. The unitary operators $(\mathscr{F}_i\alpha_x\mathscr{F}_i^{-1}f)(y)=e^{-i\{x,y\}}f(y)$ generate a maximally abelian von Neumann algebra in $\mathscr{B}(\mathscr{L}^2(X))$ which is isomorphic to $\mathscr{L}^\infty(X)$. (Compare with Example 5.3, replacing Ξ by X.) Hence $p_{\mathscr{D}}$ can be identified with a projection in $\mathscr{L}^\infty(X)$ and the lattice $\mathscr{L}^2$ introduced in Corollary 4.2 is isomorphic to the projection lattice of $\mathscr{L}^\infty(X)$. Thus $\mathscr{L}^2$ is a Boolean algebra, whose orthocomplementation is given by the operation $\perp$ of 4.2 (apart from a twist by the antiunitary involution $A\mapsto A^*$). If only countable lattice operations are considered, $\mathscr{L}^2$ can also be considered as the Borel algebra of X modulo null sets.

E. Almost periodic functions and the CCR algebra

For $i=0,1$ let $\mathscr{A}_i$ be the norm-closed linear hull of $\{E_i(x)|x\in X\}$. Since $T*E_i(x)=(\mathscr{F}_1T)(-x)\cdot E_{i+1}(x)$ for $T\in\mathscr{R}_1^1$, $\mathscr{A}=\mathscr{A}_0\oplus\mathscr{A}_1\subset\mathscr{C}$ is a pair and, since products and adjoints of exponentials are exponentials, $\mathscr{A}_0$ and $\mathscr{A}_1$ are both C^*-algebras. $\mathscr{A}_0$ is the space of continuous almost periodic functions on X and $\mathscr{A}_1$ is the CCR algebra, one of the most thoroughly studied objects in mathematical physics. (The case $\dim X<\infty$, which is studied here is considered trivial from the point of view of field theory.) I do not know whether (2) and (3) of the following proposition already exist in the literature. In any case, their proof is a nice application of the correspondence theory of Sec. IV.

Proposition 5.5: (1) There is a unique α-invariant state τ_i on $\mathscr{A}_i$. τ_i is a trace and satisfies $\tau_i(\Sigma_x\, a(x)E_i(x))=a(0)$. For $T\in\mathscr{R}_j^1$, $\tau_i(T*A)=(\mathscr{F}_jT)(0)\tau_{i+j}(A)$.

(2) $A\in\mathscr{A}_i$ iff $A\in\mathscr{C}_i$ and $\{\alpha_x A\,|x\in X\}$ is norm-precompact.

(3) For $A\in\mathscr{A}_i$, the normclosure of the convex set $C_j(A):=\{T*A\,|T\in\mathscr{R}_{i+j}^1, T\geqslant 0, \mathrm{tr}\,T=1\}\subset\mathscr{A}_i$ contains exactly one α-invariant element, namely, $\tau_i(A)1_j$.

Proof: (1) For any invariant state $\tau_i(E_i(x))=\tau_i(\alpha_yE_i(x))=e^{i\{y,x\}}\tau_i(E_i(x))=0$ for $x\neq 0$ and $\tau_i(E_i(0))=\tau_i(1_i)=1$. This proves the formula for finite linear combinations of exponentials. On this dense subset of $\mathscr{A}_i$ the other relations are easily checked, including the positivity of τ_i. Hence τ_i extends by normcontinuity to $\mathscr{A}_i$.

(2) Let $\tilde{\mathscr{A}}_i$ be the space of elements satisfying the second condition. Since $T\mapsto T*A$ is norm-continuous, $\tilde{\mathscr{A}}=\tilde{\mathscr{A}}_0\oplus\tilde{\mathscr{A}}_1$ is a pair. It is easy to show that $\tilde{\mathscr{A}}_1$ is norm-closed. $\tilde{\mathscr{A}}_0=\mathscr{A}_0$ is a classical result; hence $\tilde{\mathscr{A}}_1=\mathscr{A}_1$ by uniqueness of corresponding spaces.

(3) The case $i=j=0$ is a classical result.[24] Now let $T_i\in\mathscr{R}_i^1$ be a state. Then $T_i*A\in\mathscr{A}_0$ and there is a state $T_\epsilon\in\mathscr{R}_0^1$ such that $\|T_\epsilon*T_i*A-\tau_0(T_i*A)1_0\|$ $=\|T_\epsilon*T_i*A-\tau_i(A)1_0\|\leqslant\epsilon$. Hence for any state $T\in\mathscr{R}_{i+j}^1$ and with $\tilde{T}_\epsilon:=T*T_i*T_\epsilon*T_i\in\mathscr{R}_{i+j}^1$: $\|\tilde{T}_\epsilon*A-\tau_i(A)1_j\|\leqslant\epsilon$ since $T*T_i*1_0=1_j$. A similar argument shows that no other constant can be approximated. □

The closed invariant subspaces of $\mathscr{A}_i$ may be characterized in a very simple way: For any subset $\Lambda\subset X$, let $\mathscr{A}_i(\Lambda)$ denote the norm-closed linear hull of $\{E_i(x)|x\in\Lambda\}$. Then $\mathscr{A}(\Lambda)=\mathscr{A}_0(\Lambda)\oplus\mathscr{A}_1(\Lambda)$ is a pair. Using Proposition 5.5(3), one can prove that every closed invariant subspace of $\mathscr{A}_i$ is of the form $\mathscr{A}_i(\Lambda)$. (Compare Ref. 25, pp. 165.) Thus the sublattice $\{\mathscr{D}\in\mathscr{L}^c\,|\,\mathscr{D}\subset\mathscr{A}\}$ is isomorphic to the power set of X. $\mathscr{A}_i(\Lambda)$ is separable iff Λ is countable, self-adjoint iff $\Lambda=-\Lambda$, and a C^*-algebra iff Λ is a group. The spectrum of $\mathscr{A}(\Lambda)$ is the closure of Λ. The commutant of $\mathscr{A}_1(\Lambda)$ in $\mathscr{A}_1$ is $\mathscr{A}_1(\Lambda)'\cap\mathscr{A}_1=\mathscr{A}(\Lambda')$, where Λ' is the closed subgroup $\Lambda':=\{y\in X\,|x\in\Lambda\Rightarrow\{x,y\}\in 2\pi\mathbb{Z}\}$.

An interesting special case occurs when Λ is a nondegenerate point lattice in X, i.e., $\Lambda=\{\Sigma_{i=1}^{2N}n_ix_i\,|n_i\in\mathbb{Z}\}$ for some linear basis $\{x_1\cdots x_2\}\subset X$. Then Λ' has the same property and $\mathscr{A}_1(\Lambda)=\{A\in\mathscr{A}_i\,|x\in\Lambda'\Rightarrow\alpha_xA=A\}$. $\mathscr{A}_0(\Lambda)$ is just the space of Λ'-periodic continuous functions, so that the algebras $\mathscr{A}_0(\Lambda)$ are all isomorphic. On the other hand, the C^*-algebra $\mathscr{A}_1(\Lambda)$ is generated by the $2N$ unitaries $E(x_i)$ with the "discrete canonical commutation relations" $E(x_i)E(x_j)=e^{i\{x_i,x_j\}}E(x_j)E(x_i)$ and depends very sensitively on the matrix $\{x_i,x_j\}=:2\pi M_{ij}$. For example, the center of $\mathscr{A}_1(\Lambda)$ is $\mathscr{A}_1(\Lambda\cap\Lambda')$ and $\Lambda\cap\Lambda'$ may be equal to Λ or to $\{0\}$, depending on the existence of rational relations between the M_{ij}. More complete invariants are provided by K theory, in particular, the image of the group $K_0(\mathscr{A}_1(\Lambda))$ by the trace τ_1. For $N=1$, $\mathscr{A}_1(\Lambda)$ is known as the algebra of rotations by $\{x_1,x_2\}$ and that group is $\mathbb{Z}+(2\pi)^{-1}\{x_1,x_2\}\cdot\mathbb{Z}$ and a complete invariant. For higher dimensions N, $K_0(\mathscr{A}_1(\Gamma))$ has been calculated by Elliot.[26]

[1] B. Simon, Commun. Math. Phys. **71**, 247 (1980).
[2] E. H. Lieb, in *Geometry of the Laplace Operator*, edited by R. Ossermann and A. Weinstein (Amer. Math. Soc., Providence, RI, 1980).
[3] A. Grossman and P. Huguenin, Helv. Phys. Acta **51**, 252 (1978).
[4] I. E. Segal, Math. Scand. **13**, 31 (1963).
[5] F. Bopp, Ann. Inst. Henri Poincaré **15**, 81 (1956).
[6] J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[7] R. Werner, Found. Phys. **13**, 859 (1983).
[8] J. von Neumann, Math. Ann. **104**, 570 (1931).
[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[10] W. Beckner, Ann. Math. **102**, 159 (1975).
[11] J. Diestel and J. J. Uhl, *Vector Measures*, AMS Math. Surveys 15 (Am. Math. Soc., Providence, RI, 1977).
[12] E. Størmer, in *Foundations of Quantum Mechanics and Ordered Linear Spaces*, edited by A. Hartkämper and H. Neumann (Springer, New York, 1974).
[13] A. S. Holevo, Rep. Math. Phys. **16**, 385 (1979).
[14] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[15] U. Cattaneo, J. Math. Phys. **23**, 659 (1982).
[16] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[17] I. Daubechies, Commun. Math. Phys. **75**, 229 (1980).
[18] A. Voros, J. Funct. Anal. **29**, 104 (1978).
[19] H. Reiter, *Classical Harmonic Analysis and Locally Compact Groups* (Clarendon, Oxford, 1968).
[20] E. Hewett and K. A. Ross, *Abstract Harmonic Analysis* (two volumes) (Springer, New York, 1963/1970).
[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[22] A. Grossmann, Commun. Math. Phys. **48**, 191 (1976).
[23] J. J. Benedetto, *Spectral Synthesis* (Tuebner, Stuttgart, 1975).
[24] L. H. Loomis, *An Introduction to Abstract Harmonic Analysis* (Nostrand, Princeton, 1953).
[25] Y. Katznelson, *An Introduction to Harmonic Analysis* (Wiley, New York, 1968).
[26] G. A. Elliott, Københavns Math. Inst. Preprint No. 14, 1981.

Functional integrals as integrals on locally noncompact groups with generalized measures

N. N. Meiman
Maxim Gorky nab. 4/22, ap.57, Moscow 113127, USSR

(Received 29 July 1983; accepted for publication 28 October 1983)

A functional integral is considered as an integral on a locally noncompact group with a generalized measure; the general method of integration on such groups is described, and conditions under which the integral exists are formulated. In particular, the action functional is determined by means of a group-theoretical method. The path integral for a quantum-mechanical motion of a nonrelativistic particle in the Euclidean space is considered from this point of view. The corresponding locally noncompact group is the set of all alternative trajectories where the group operation is the pointwise addition. The internal group theoretical action functional present in the generalized measure is coinciding with the classical action functional. Thus a group-theoretical definition of the classical action functional is obtained.

PACS numbers: 03.65. − w, 02.20. + b

INTRODUCTION

Feynman put forth an extremely important and highly stimulating general principle: the transition probability amplitude for any quantum mechanical system is the sum of functionals $\exp[i/\hbar)\, S(\gamma)]$ over all paths γ. This was the foundation underlying the method of functional integrals in theoretical physics which has become one of its main tools.

The first questions one would ask in trying to understand the method are what do we mean by definition of the integral over all paths, what are the rules for handling it, and how can we calculate it. The well-known approach proposed by Feynman is to begin with a multiple (finite) integral over a time lattice and to calculate the limit of this integral as the step of the lattice is going down to zero. Actually, this prescription, if operable, leads to correct results. This is, however, an approach essentially different from the original principle, since before the limit of vanishing time step is taken the functional $\exp[(i/\hbar)\, S(\gamma)]$ must be integrated over the polygonal trajectories the segments of which are either straight or elements of the classical trajectories, but not over all possible paths. Even in the limit $\Delta t \to 0$ the contradiction is not removed, as though any continuous trajectory can be approximated with polygonal lines, the variation of $S(\gamma)$ for an arbitrarily close vicinity of the path remains infinite.

A specific difficulty inherent in the definition of the functional integral on the basis of Feynman's formulation is that the path integral is being contributed by the paths for which Δx is proportional to $(\Delta t)^{1/2}$. For such paths $(i/\hbar)\, S(\gamma) = i\infty$ (note that it is $-\infty$ for the diffusion equation), so that the conclusion we draw is that the functional $\exp(iS/\hbar)$ is devoid of meaning (respectively, vanishes for the diffusion equation) on the paths which, in fact, do determine the value of the integral.

We would like to attribute the general Feynman principle with an actual meaning and to define the functional integral in the space of alternative paths.[2] Such a definition may be useful in such complicated matters as quantum field theory and the theory of gauge fields only if it is sufficiently general and clear. Therefore, we feel it important that the functional integral is defined in this paper almost literally as an improper integral on the infinite axis; the difference being that the role of the line element dx is performed by a generalized measure endowed with a clear physical meaning.

It is appropriate to look at the problem from a more general point of view, as is often helpful, and to consider the integration on a locally noncompact group, abelian or nonabelian. For such groups there is no measure, invariant under the group transformations, and we have no theory of the integration in general. However, under certain conditions which are quite natural for physics, it is possible to construct a general method of integration on the groups, as it is shown in Sec. I.

The main condition involved is the existence of a basis system of vicinities in the group G, which are called channels in this paper, and a possibility to determine a completely additive function on this system, μ, and we shall call it the generalized rough measure. Then the functional integral for the group G is defined in an almost standard manner, in terms of the integral sums, and it is written as

$$\int_G \varphi(g)\, d\mu(g). \tag{0.1}$$

In the cases of actual interest there exists such a subgroup G', which is dense everywhere in G, that for any $g' \in G'$ there is a limit

$$\frac{d\mu(g')}{d\mu(I)} = \lim \left\{ \frac{\mu[g'V(I)]}{\mu[V(I)]} \right\}$$

as the channel $V(I)$ is contracted to the group identity element I. The notation used for the limiting value is $\exp[(i/\hbar)\, S(g')]$, this is an internal group-theoretical definition of the action functional on the group G. Moreover, the introduction of the action functional enables one to introduce a generalized measure invariant under the group transformations,

$$d\sigma(g') = \exp[(i/\hbar)\, S(g')]\, d\mu(g'),$$

and to write down the functional integral in the form specific for the Feynman path integral,

$$\int_G \varphi(g) \exp\left[\frac{i}{\hbar} S(g)\right] d\sigma(g). \tag{0.2}$$

 0022-2488/84/051412-22$02.50

The form of Eq. (0.1) is more adequate than that of (0.2), as is argued in Sec. I. The main advantage of the latter is the formal applicability of some rules of the standard integral calculus. It is remarkable that the action functional, which is determined from the group-theoretical arguments, is the same as the classical action functional. In other words, we have got an internal group-theoretical definition of the classical action functional.

The group G, and the generalized rough measure μ, are defined in a natural way to fit the quantity that is represented by means of the functional integral. In the present work we deal with the path integral representation of the evolution kernel $K(x^f,t^f|H\mid x^i,t^i)$, $x\in\mathbb{R}^n$, for the n-dimensional Schrödinger equation

$$-\frac{\hbar}{i}\frac{\partial\psi}{\partial t}=H\psi,$$

describing the motion of a nonrelativistic particle of mass m. In this case the group G is the group of all continuous curves connecting the initial point (x^i,t^i) with the final point (x^f,t^f), $\gamma=x(t)$. The notation used for the group is $G^{x^f t^f}_{x^i t^i}$. The group operation is

$$\gamma_1\otimes\gamma_2=\bar\gamma_i^f+(\gamma_1-\bar\gamma_i^f)+(\gamma_2-\bar\gamma_i^f),$$

where $\bar\gamma_i^f$ is the segment of the straight line connecting the initial and the final points. The generalized rough measure is induced by the Schrödinger equation; it is determined in Sec. II. The approach to the functional integral treating it as the integral with the generalized measure on a locally noncompact group is revealed essentially even in this simple case.

The subject of Sec. III is the rule for the change of variables induced by group transformations in the group of paths. It is proved that the action functional derived by the internal group-theoretical method coincides with the classical action. The present approach to the path integral is illustrated with some simple problems.

Section IV presents a modification of the integral on the time lattice, basing on the proposed approach, and a general proof is given that the integral on the lattice is converging to the functional integral as $\Delta t\to 0$.

The main part of the matter presented in this paper was given in the author's reports[1] delivered to International Sessions of the Moscow Seminar on Collective Phenomena, in 1977 and 1981.

The multidimensional Schrödinger equation is concerned only in the aspects where the method exposed for the one-dimensional case needs a modification.

Comprehensive reviews of various approaches to path integrals, and extensive bibliography, are given in Refs. 2–5. In Ref. 3 it was indicated, in particular, that the concept of measure needs a modification if the path integral is considered.

I. INTEGRATION ON LOCALLY NONCOMPACT GROUPS: A GENERAL APPROACH

(1) We shall treat the functional integration as an integration on topologically noncompact groups, and so it is reasonable to start from a description of the general scheme for integration on groups of this class. Unlike compact groups and locally compact groups, in the case in view we have no invariant measure, and the theory of integration is not at hand. Nevertheless, we will propose now an outline of the integration method to work with, which is applicable under some restrictions that are fulfilled in physical problems.

Let G be a topological locally noncompact group; it can be abelian or nonabelian. To make the subsequent formulations simpler, we suppose that G is a metric space and that for any three group elements g_1,g_2,g, the distance function ρ satisfies the following condition: $\rho(gg_1,gg_2)<C\rho(g_1,g_2)$, where C is a constant independent of the group elements.

Let $\{V_\alpha\}$ be a system of vicinities in G, which includes the group G itself, is invariant under the left multiplications, and is closed with respect to the operation of intersection, i.e., for any pair of the vicinities belonging to the system their intersection also belongs to the set $\{V_\alpha\}$. Elements of such a set will be called *channels*. Any set of channels which are mutually exclusive is called as set of independent channels.

We call a set $\{V_\alpha\}$ the left basis system if the following conditions hold.

1. There is a completely additive function μ on the set $\{V_\alpha\}$ which is finite for every channel V and vanishes on boundaries of the channels, ∂V_α. In particular, $|\mu(G)|<\infty$.

2. The group G can be subdivided into a countable (or finite) set of independent channels, the diameters of which being less than an arbitrary positive quantity $\delta>0$, up to the channel boundaries,

$$G=\bigcup_{\{\alpha\}}V_\alpha+\mathfrak{N},\quad \mu(G)=\sum_{\{\alpha\}}\mu(V_\alpha). \tag{1.1}$$

(Here and below in the paper, $\mathfrak{N}$ stands for a set of group elements which belong to the channel boundaries.)

It follows from (1.1) that every channel in the system $\{V_\alpha\}$ can also be subdivided into a finite or countable set of independent channels, the diameter of which is less than a fixed quantity $\delta>0$:

$$U=\bigcup_{\{\alpha\}}V_\alpha+\mathfrak{N},\quad \mu(U)=\sum_{\{\alpha\}}\mu(V_\alpha). \tag{1.2}$$

Any subsystem $\{V_\alpha\}$ in the decomposition (1.1) for the group G will be called a complete system of independent channels, and any subsystem (1.2) will be called complete with respect to the channel U. The function μ introduced here will be called a generalized rough measure.

It is suitable to mark the channels belonging to a system of independent channels by means of arbitrarily chosen elements of the channels. Thus the decompositions in (1.1), (1.2) can be written down as follows:

$$G=\bigcup_{(g)}V(g)+\mathfrak{N},\quad \mu(G)=\sum_{(g)}\mu[V(g)], \tag{1.3}$$

$$U=\bigcup_{(g)}V(g)+\mathfrak{N},\quad \mu(U)=\sum_{(g)}\mu[V(g)]. \tag{1.4}$$

The index (g) here means that the sum is taken over the corresponding complete system of independent channels.

The quantity

$$\sum_{(\alpha)}|\mu(V_\alpha)|=\mathrm{Var}[\mu;G;(\alpha)] \tag{1.5}$$

is called the (α)-variation of the rough measure μ, taken for the (α)-decomposition (1.1). Respectively, $\mathrm{Var}[\mu; U;(\alpha)]$ means the sum of the channel measures for the decomposition (1.2).

Condition 2 holds only if $\mathrm{Var}[\mu; G;(\alpha)] < \infty$, as the sum of an absolutely convergent series is independent of the order of its terms. If $\mathrm{Var}[\mu; G;(\alpha)] = \infty$, one should be careful, as in the operations with improper integrals.

We suppose that $\mathrm{Var}[\mu; U;(\alpha)]$ is finite for any finite channel U (recall that G is a metric space). Let $\{U_n\}$ be a family exhausting the group G, i.e.,

$$U_1 \subset U_2 \subset U_3 \cdots, \tag{1.6}$$

and any group element belongs to some channel U_n. An additional requirement to be imposed upon the rough measure μ is

$$\lim_{n\to\infty} \mu(U_n) = \mu(G). \tag{1.7}$$

Thus if for an (α)-decomposition (1) of the group G one has $\mathrm{Var}[\mu; G;(\alpha)] = \infty$, one should consider instead the decompositions for the set exhausting the group G,

$$U_n = \bigcup_{(g_n)} V(g_n) + \mathfrak{N}, \quad \mu(U_n) = \sum_{(g_n)} \mu[V(g_n)]. \tag{1.8}$$

In this case we shall call the group decomposition of this type, for the sake of brevity, " a big box decomposition."

There is a quite simple, but very useful, example of a locally compact group. The real axis is a group with respect to the usual addition; the intervals (x_-, x_+) can be considered as channels in the sense explained above, and the rough measure can be defined as the integral $\mu[x_-, x_+)] = \int_{x_-}^{x_+} \exp(ix^2)\,dx$. The variation of μ taken over the decomposition of the real axis into the intervals $\mp(\sqrt{n\pi}, \sqrt{(n+1)\pi})$, $n = 0,1,2,\cdots$, is infinite. Actually, to get the correct magnitude for the measure of the group,

$$\mu(G) = \int_{-\infty}^{+\infty} \exp(ix^2)\,dx,$$

one has to sum up the channel measures $\mu[\mp(\sqrt{n\pi}, \sqrt{(n+1)\pi})]$ by means of the big box rule.

The example considered is also useful to explain why we have applied the term "generalized rough measure." One does not need a generalized measure for sophisticated sets; the sets of our present interest are those where one has an interference effect.

A difficulty specific for the integration with a strongly oscillating generalized rough measure μ is that the meaure for a subchannel may be much greater, by its absolute value, than the measure for the channel itself, and the contributions to the integral sum, $\varphi(g)\mu[V(g)]$ and $\Sigma_{(\alpha')}\varphi(g')\mu[V(g_{\alpha'})]$, where $V(g) = \cup_{(\alpha')} V(g_{\alpha'})$, are substantially different. A necessary condition for a functional to be integrable with the generalized measure μ is that the oscillation of the functional must be small as compared with the oscillation of the generalized rough measure for the channels in the subdivisions of which the oscillation of the rough measure is strong. If this is true, one has

$$\sum_{(\alpha')} \varphi(g_{\alpha'})\mu[V(g_{\alpha'})] \sim \varphi(g)\mu[V(g)].$$

(2) The decompositions of any fixed channel U, including the group G itself, into the complete systems of independent channels can be partially ordered; a decomposition (α'') follows another decomposition (α'), if for every channel in the (α')-decomposition there exists in the (α'')-decomposition a system of channels which is complete with respect to it. For any two decompositions (α') and (α'') there exists a decomposition which follows them both, namely, the $(\alpha'\alpha'')$-decomposition made up of all the intersections of channels belonging to (α') and (α''). Thus the totality of all the decompositions is a partially ordered set, and it forms the so-called directedness. The limit along the directedness is defined in the same manner as the limit for a usual sequence: a quantity $f(\alpha)$ determined on elements of the (α)-decomposition has f as its limit if for any fixed $\epsilon > 0$ there is such a decomposition $(\alpha)_\epsilon$ that $|f - f(\alpha)| < \epsilon$ for any decomposition which follows $(\alpha)_\epsilon$. The familiar Cauchy criterion indicating whether a limit does exist is applicable here as well.

Now we are in position to define the integral on any channel U (including the group G itself), for which the magnitude of $\mathrm{Var}[\mu; U;(\alpha)]$ is finite for all the (α)-decompositions.

Let us take a complex-valued function $\varphi(g)$ on the group G, and write down the integral sum for (α)-decompositions of the channel U,

$$\sum_{(\alpha)} = \sum_{(g)} \varphi(g)\mu[V(g)], \quad g \in V(g). \tag{1.9}$$

If a limit along the directedness of the (α)-decompositions exists for the integral sums, the limit is called the integral of the function $\varphi(g)$ on the channel U over the left rough measure,

$$\lim_{\to} \sum_{(g)} \varphi(g)\mu[V(g)] = \int \varphi(g)\,d\mu(g). \tag{1.10}$$

The conventional proof of the integrability is applicable to the integral over the rough measure μ for any uniformly continuous function φ on the channel U.

If for some (α)-decompositions $\mathrm{Var}[\mu; G;(\alpha)]$ is infinite, then the integral upon the group G is defined as the limit of integrals upon channels U_n which are members of a family exhausting the group G, provided that the limit does exist,

$$\int \varphi(g)\,d\mu(g) = \lim_{n\to\infty} \int_{U_n} \varphi(g)\,d\mu(g). \tag{1.11}$$

It is easily seen that if there are limits for two different exhausting families, they must coincide. It is sufficient to have a single family for which the limit in Eq. (1.11) is existing. The integral upon the whole group G will be written down with no indication of the integration domain. Sometimes we use the notation (φ, μ) for this overall integral.

The space of functions integrable upon the group G over the rough measure μ will be denoted by $O(G, \mu)$. The topological structure in the space $O(G, \mu)$ is introduced as the mean convergence with the measure μ; the sequence $\varphi_n(g)$ is converging to $\varphi(g)$ if $(|\varphi - \varphi_n|, \mu) \to 0$.

(3) Let us assume that there is a subgroup G' which is dense everywhere in G; it plays a role of importance in the following and will be called the normalizing subgroup. It is of a particular significance as for any pair of its elements, g'_1

and g_2', the ratio of the generalized rough measure, $\mu[g_2' V_\delta(I)]/\mu[g_1' V_\delta(I)]$, converges uniformly to a limit $f_\mu(g_2')/f_\mu(g_1')$ which is finite and nonzero, as the diameter δ of the channel $V_\delta(I)$, containing the group identity element, tends to zero.

There are some clear reasons, to be explained below, to write down this limiting value as $\exp[(i/\hbar) S_\mu(g_2') - (i/\hbar) S_\mu(g_1')]$, and to call the quantity $S_\mu(g')$ the action functional of the group element g'. An additional requirement to be imposed on the generalized rough measure is that the normalizing subgroup G' must exist for which the difference $S(g_2') - S(g_1')$ is determined:

$$\lim_{\delta \to 0} \frac{\mu[g_2' V_\delta(I)]}{\mu[g_1' V_\delta(I)]} = \exp\left[\frac{i}{\hbar} S_\mu(g_2') - \frac{i}{\hbar} S_\mu(g_1')\right] \quad (1.12)$$

(prime stands to indicate that the element belongs to the subgroup G').

Equation (1.12) is a proper group definition of the action functional $S_\mu(g')$ for the normalizing subgroup G'. The magnitude of $S_\mu(I)$ can be taken at will.

For the cases of real interest the normalizing subgroup G' can be chosen in such a way that the difference of the action functionals, $S_\mu(g_2) - S_\mu(g_1)$, is determined not only for any pair of the subgroup elements, but also for any two elements lying in the same coset in the quotient G/G'. That is to say, the following relation holds:

$$\lim \frac{\mu[g_2 V_\delta(I)]}{\mu[g_1 V_\delta(I)]} = \exp\left[\frac{i}{\hbar} S_\mu(g_2) - \frac{i}{\hbar} S_\mu(g_1)\right],$$

$$g_2 g_1^{-1} = g_2' g_1'^{-1}, \quad G = G' + \bigcup_\alpha G' g_\alpha. \quad (1.13)$$

For any $g \in G'$ the action functional is defined by

$$\frac{i}{\hbar} S_\mu(g) = \lim \ln \frac{\mu[g V_\delta(I)]}{\mu[V_\delta(I)]}, \quad (1.14)$$

where it is assumed that such a limit does exist and is either finite or infinite. Then Eq. (1.13) means that for all elements belonging to the same coset the action functional is either a finite quantity, or becomes finite after subtraction of a common infinite component.

The generalized rough measure can be written in terms of the action functional,

$$\mu[V(g)] = \exp[(i/\hbar) S(g') - (i/\hbar) S(g_0') + (i/\hbar)\epsilon]\, \mu[g_0' g'^{-1} V(g)], \quad (1.15)$$

where g_0' is an arbitrary fixed element, $g' \in [V(g) \cap G']$; ϵ depends on the channel $V(g)$, on the elements g' and g_0', and approaches zero, as the diameter of the channel $V(g)$ vanishes. Then it follows that

$$\int_U \varphi(g)\, d\mu(g) = \lim_{\to} \sum_{(g)} \varphi(g) \times \exp \frac{i}{\hbar} [S_\mu(g') S_\mu(g_0')] \times \mu[g_0' g'^{-1} V(g)] \quad (1.16)$$

for any channel U for which the magnitude of $\mathrm{Var}[\mu; U; (\alpha)]$ is finite on every decomposition (α).

It is suitable to introduce the notation

$$\sigma[V(g); g'; g_0'] = \exp[-(i/\hbar) S(g_0')]\, \mu[g_0' g'^{-1} V(g)] \quad (1.17)$$

so the integral over the channel U is written down as follows

$$\int_U \varphi(g)\, d\mu(g) = \lim_{\to} \sum \varphi(g) \times \exp\left[\frac{i}{\hbar} S_\mu(g')\right] \sigma[V(g); g'; g_0']. \quad (1.18)$$

In view of the condition (1.12), for infinitely small channels the quantity $\sigma[V(g); g'; g_0']$ is independent of the choice of the basis element g_0'; that is to say, the following notation is adequate:

$$\sigma[V(g); g'; g_0'] = d\sigma(g; g'), \quad \mu[V(g)] = dg, \quad (1.19)$$

so that one has

$$\int_U \varphi(g)\, d\mu(g) = \int_U \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g')\right] d\sigma(g; g'). \quad (1.20)$$

The integral over the group on the whole equals the limit of integrals over channels which are members of an exhausting family,

$$\int \varphi(g)\, d\mu(g) = \lim_{n \to \infty} \int_{U_n} \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g')\right] d\sigma(g; g'). \quad (1.21)$$

For the cases of actual interest the variation of the exponent is infinite for arbitrarily narrow channels, as is the oscillation of the ratio $d\sigma(g; g_1')/d\sigma(g; g_2')$ induced by the variation of g_1' and g_2' within the channel $V(g)$. Nevertheless, the limit in the integral sum in Eq. (1.18) does not depend on the choice of the element g' in the channel $V(g)$. This results from the correlation between $\sigma(g; g')$ and $S_\mu(g')$, and so the concept of the rough measure justifies the conventional notation for the integral in (1.18):

$$\int_U \varphi(g)\, d\mu(g) = \int_U \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g)\right] d\sigma(g). \quad (1.22)$$

It should be emphasized, however, that this expression must not be considered as a result of summing up of $\varphi(g) \exp[(i/\hbar) S_\mu(g)]$ over elements of the channel U. The exact meaning of the thing is revealed in fact, only in Eqs. (1.17) and (1.18).

A basis system of channels $\{V_\kappa\}$ generates a continuum of the systems $\sigma[V(g); g', g_0']$, since the choice of the elements g' and g_0' is not restricted, but all the systems determine the same linear operation in the space $O(G, \mu)$ because of Eqs. (1.18) and (1.19); in other words, one gets a unique generalized measure σ.

The generalized measure σ is invariant under the left multiplications. Actually, for any $g_1 \in G$ one has

$$\sigma(g_1 V(g); g_1 g'; g_0') = \exp\left[-\frac{i}{\hbar} S_\mu(g_0')\right] \mu[g_0' (g_1 g')^{-1} g_1 V(g)],$$

i.e.,

$$\sigma(g_1 V(g); g_1 g'; g_0) = \sigma(V(g); g'; g_0'). \tag{1.23}$$

So for every channel U one has

$$\int_U \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g)\right] d\sigma(g)$$
$$= \int_{g_1 U} \varphi(g_1^{-1} g) \exp\left[\frac{i}{\hbar} S_\mu(g_1^{-1} g)\right] d\sigma(g). \tag{1.24}$$

In particular, the integral over the whole group is invariant,

$$\int \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g)\right] d\sigma(g)$$
$$= \int \varphi(g_1^{-1} g) \exp\left[\frac{i}{\hbar} S_\mu(g_1^{-1} g)\right] d\sigma(g). \tag{1.25}$$

The conventional notation (1.22) is quite convenient also because the inherent formalism is simple,

$$d\sigma(g) = \exp[-(i/\hbar) S_\mu(g)]\, d\mu(g), \quad d\sigma(g_2) = d\sigma(g_1), \tag{1.26}$$

while the exact meaning of the symbols used has been elucidated above.

The integral $\int \varphi(g)\, d\mu(g)$ can be represented in terms of the integral sums with the elements belonging to any coset $G'\hat{g}_0$ of the normalizing subgroup G',

$$\int_U \varphi(g)\, d\mu(g) = \lim_{\rightarrow} \sum \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(\hat{g}) - \frac{i}{\hbar} S_\mu(\hat{g}_0)\right]$$
$$\times \mu[\hat{g}_0 \hat{g}^{-1} V(g)]\, \hat{g} \in [V(g) \cap G'\hat{g}_0], \tag{1.27}$$

as well as by means of the sums of Eq. (1.18). If $S_\mu(\hat{g}_0)$ is infinite, the difference $S_\mu(\hat{g}) - S_\mu(\hat{g}_0)$ can not be decomposed to provide us with an exponential factor present in the definition of $\sigma[V(g); g'; g_0']$; yet the integral in Eq. (27) is performed, in fact, with an invariant measure. The expression on the rhs is independent of the choice of the element $\hat{g}_0$, and $\mu[\hat{g}_0 (g_1 \hat{g})^{-1} g_1 V(g)] = \mu[\hat{g}_0 \hat{g}^{-1} V(g)]$, so that the measure defined in Eq. (22) can be used to rewrite the integral in Eq. (27):

$$\int_U \varphi(g)\, d\mu(g)$$
$$= \int \varphi(g) \exp\left[\frac{i}{\hbar} S_\mu(g) - \frac{i}{\hbar} S_\mu(g_0)\right] d\sigma_1(g),$$
$$d\sigma_1(g) = \mu[\hat{g}_0 \hat{g}^{-1} V(g)]. \tag{1.28}$$

Correspondingly, the expressions in Eq. (1.26) are replaced by

$$d\sigma_1(g) = \exp\left[-\frac{i}{\hbar} S_\mu(g) + \frac{i}{\hbar} S_\mu(g_0)\right] d\mu(g),$$
$$d\sigma_1(g_1) = d\sigma_1(g_2), \tag{1.29}$$

while the difference $[S_\mu(g) - S_\mu(g_0)]$ can be considered as a finite component of the action functional on the coset $G'g_0$.

II. CONTINUAL INTEGRALS FOR THE NONRELATIVISTIC SCHRÖDINGER EQUATION

(1) The method of integration on topological locally noncompact groups, which was presented in the preceding section, enables one to get a unified description of various quantities concerned in quantum mechanics and quantum field theory, writing them down as the path integrals. In every case, the group G where the integration is established and the appropriate rough measure on the group are defined in a way which is natural for the problem considered; and the action functional defined by Eq. (1.12), which has an internal meaning, is coinciding with the classical action. In other words, this is a way to get an internal group-theoretical definition of the conventional action functional.

We begin with a detailed consideration of the one-dimensional motion of a nonrelativistic quantum particle of mass m in an arbitrary potential field. The system is described in terms of the Schrödinger equation with the Hamiltonian operator[6] H,

$$-\frac{\hbar}{i}\frac{\partial \psi}{\partial t} = H\psi,$$
$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - \frac{\hbar}{i} A(x,t)\frac{\partial}{\partial x} + V(x,t). \tag{2.1}$$

The final state wave function of the particle at a time moment t^f is the convolution of the initial state wave function at a time moment t^i with the evolution kernel (Green's function) $K(x^f, t^f | H | x^i, t^i)$ of Eq. (2.1). The evolution kernel is the probability amplitude to find the particle at a point x^f for the moment t^f, once it was observed at a point x^i for a time moment t^i. Thus, if one has to find out the time evolution of the system, it is sufficient to get a path integral representation for the evolution kernel. It is well known that the latter satisfies the differential equation

$$\left(\frac{\partial}{\partial t} + \frac{i}{\hbar} H\right) K = \delta(t - t^i)\, \delta(x - x^i). \tag{2.2}$$

Let $G_{x^i t^i}^{x^f t^f}$ be the set of all paths starting at the point x^i at the time moment t^i and getting to x^f at the time moment t^f, and $G_{t^i}^{t^f}$ be the set of all the paths starting at t^i from *some* point and finishing at an arbitrary point at the time moment t^f. The time arguments will be omitted in all the cases where we do not expect a confusion, and the sets of paths are denoted as $G_{x^i}^{x^f}$ and $G_{t^i}^{t^f}$, respectively; elements of the sets are x or $x(t)$. Evidently, the sets in view are metric spaces, and the distance between two elements is given by

$$\rho(x_1, x_2) = \sup |x_2(t) - x_1(t)|, \quad t^i \leqslant t \leqslant t^f.$$

The set $G_{t^i}^{t^f}$ is an abelian group with respect to the pointwise addition $(x_1 + x_2)(t) = x_1(t) + x_2(t)$. As to the set $G_{x^i}^{x^f}$, it is an abelian group only for the case where both paths are closed, and $x^i = x^f = 0$, but the group structure can be introduced in the following general way.

Let G be a group and G_0 its subgroup; then

$$G = \cup\, G_0 g_\alpha$$

is a decomposition of G into cosets relevant to the subgroup G_0. The mapping $g \rightleftarrows g g_\alpha$, $g \in G_0$, is a isomorphism which determines the group operation in the coset $G_\alpha = G g_\alpha$,

$$g_1 g \circ g_2 g = (g_1 g_2) g_\alpha, \quad g_1, g_2 \in G_0.$$

If we turn back to the sets $G_{t^i}^{t^f}$ and $G_{x^i}^{x^f}$, we see that the latter is a coset with respect to the subgroup $G_{x^i = 0}^{x^f = 0}$. The simplest

choice of the element g_α fixing the coset is the path corresponding to the motion with a constant velocity, namely,

$$x_i^f(t) = x^i + (x^f - x^i)(t - t^i)/(t^f - t^i). \tag{2.3}$$

Thus the set $G_{x^i}^{x^f}$ is an abelian group where the group addition is

$$x_1(t) \oplus x_2(t) = [x_1(t) - x_i^f(t)] + [x_2(t) - x_i^f(t)] + x_i^f(t) \tag{2.4}$$

and the group identity element is the uniform trajectory $x_i^f(t)$. Of course, any path with the fixed end points may be taken as the identity element, and then this path must be implied in the definition (2.4).

Let $x_-(t)$ and $x_+(t)$ be limiting functions for continuous functions on the interval $t^i \leqslant t \leqslant t^f$ in the pointwise topology, and $x_-(t) < x_+(t)$ for any t in the interval considered. The subset of paths in G_i^f, satisfying the condition

$$x_-(t) < x(t) < x_+(t), \quad t^i \leqslant t \leqslant t^f, \tag{2.5}$$

will be called the *channel* in the group G_i^f. The channel diameter is defined by

$$d\{x_-, x_+\}_i^f = \sup|x_+(t) - x_-(t)|. \tag{2.6}$$

The channel boundary, $\partial\{x_-, x_+\}_i^f$, is the set of all paths for which one of the inequalities in (2.5) becomes the equality at some time moments. We define the channels $\{x_-, x_+\}_{x^i}^{x^f}$ as intersections of channels $\{x_-, x_+\}_i^f$ with the set $G_{x^i}^{x^f}$. The appropriate system of vicinities in the group $G_{x^i}^{x^f}$, as that mentioned in Sec. I, $\{V_\alpha\}$, is the system of all the channels $\{x_-, x_+\}_{x^i}^{x^f}$.

In analogy to $G_{x^i}^{x^f}$, one can define the groups $G_{x^i t^i}^{t^f}$ and $G_{t^i}^{x^f t^f}$. The former contains all the paths which start from a fixed point, and the latter is the set of all paths which finish at a fixed point. The channels in these groups are defined as the intersections of these groups with channels belonging to $G_{t^i}^{t^f}$. The identity elements for the groups in view are $x(t) \equiv x^i$ and $x(t) \equiv x^f$, respectively. The group operations are those translated from the vector spaces $G_{0t^i}^{t^f}$ and $G_{t^i}^{0t^f}$, respectively, for instance, in the case of $G_{x^i t^i}^{t^f}$ we have

$$x_1(t) \oplus x_2(t) = x^i + [x_1(t) - x^i] + [x_2(t) - x^i]. \tag{2.4'}$$

Suppose the functions $x_-(t)$ and $x_+(t)$ are finite (not equal to $-\infty$ and $+\infty$, respectively) only for a discrete set of time moments, $t^{(0)} = t^i, t^{(1)}, \ldots, t^{(k)} = t^f$; the channel defined by (2.5) in this case will be called a beam, and its notation is $B^{(k)}[x_-^{(j)}, x_+^{(j)}]$. The beam opening is, by definition,

$$\max[x_+^{(j)} - x_-^{(j)}], \quad j = 0,1,\ldots,k. \tag{2.7}$$

The beams in the group $G_{x^i}^{x^f}$ are called the sets defined as

$$B_{x^i}^{x^f(k)}[x_-^{(j)}, x_+^{(j)}] = B^{(k)}[x_-^{(j)}, x_+^{(j)}] \cap G_{x^i}^{x^f}.$$

It is suitable sometimes to lift the construction described into the (x,t) space, i.e., to consider the $(x(t),t)$ graphs instead of just the paths $x(t)$ on the x axis. We retain the terminology in describing such an extended picture, supplying the symbols with the time argument (t), if necessary. So, e.g., $G_i^f(t)$ is the group of all graphs having the starting point upon the line $t = t^i$, and the end point up on the line $t = t^f$; the channel $\{x_-, x_+\}_i^f(t)$ should not be confused with the band $(x_-, x_+)_i^f(t)$ in the (x,t) plane, the boundaries of the band being the curves $x_-(t)$.

For the beams $B^{(k)}[x_-^{(j)}, x_+^{(j)}](t)$, unlike arbitrary channels $\{x_-, x_+\}_i^f(t)$, there is a very simple proof that the whole path space $G_i^f(t)$ is decomposed into a complete system of independent beams, the openings of which are arbitrarily small. It is sufficient to divide the straight lines $t = t^j$, $j = 0,1,\ldots,k$, into systems of nonoverlapping intervals which are sufficiently small. All possible systems of such intervals, one interval on every line, constitute a complete system of independent beams. The intersection of the system of beams with the group $G_{x^i}^{x^f}(t)$ is a complete system of beams for this group. Replacing the graphs $(x(t),t)$ by the corresponding paths, one gets the decomposition of the group $G_{x^i}^{x^f}$ over the complete system of independent beams $B_{x^i}^{x^f(k)}[x_-^{(j)}, x_+^{(j)}]$.

Such a decomposition of the group can be taken as a basis for the following consideration; the sequence of beams $B^{(k)}$ with all values of k, while for every given k all possible beams are involved, can be used as the system of vicinities $\{V_\kappa\}$. In Sec. IV we shall be concerned with this point, in view of modification of the constructive definition of the Feynman path integral. We prefer, however, to take as the starting point the statement, which is proved in the Appendix, namely, that the groups G_i^x and $G_{x^i}^{t^f}$, as well as any fixed channel in each group, can be represented as a union of a denumerable set of arbitrarily narrow channels, up to their boundaries. In other words, the systems of vicinities $\{V_\alpha\}$ in the groups G_i^f and $G_{x^i}^{x^f}$ are the systems of channels $\{x_-, x_+\}_i^f$ and $\{x_-, x_+\}_{x^i}^{x^f}$, respectively, and the conditions (1.1) and (1.2) are supposed to hold.

(2) The main assumption is that every channel $\{x_-, x_+\}_{x^i}^{x^f}$ has its proper partial amplitude, the notation we use is $\mu_H(\{x_-, x_+\}_{x^i}^{x^f})$ for the probability of the particle transition from (x^i, t^i) to (x^f, t^f) along the paths belonging to the channel, and that the total probability amplitude $K(x^f t^f | H | x^i t^i)$ is a superposition of the partial amplitudes for any complete system of independent channels,

$$K(x^f t^f | H | x^i x^i) = \sum \mu_H(\{x_-, x_+\}_{x^i}^{x^f}). \tag{2.8}$$

The decomposition of the group $G_{x_i}^{x^f}$ over a complete system of independent channels takes place only up to the channel boundaries, so Eq. (2.8) includes implicitly the requirement that the probability amplitude for transition along the channel boundary is vanishing. The partial amplitude, as a function of x^f and t^f, satisfies the same equation, (2.2), as the total amplitude $K(x^f t^f | H | x^i t^i)$, since the differential equation tests only local properties of the function.

With Eq. (2.8) one can understand why it is the partial amplitude of the channel that is chosen as the rough measure. As the partial amplitude for the channel boundary is zero, and the differential equation (2.2) is satisfied, one has the following statement on the generalized rough measure.

Statement: A generalized rough measure $\mu_H(\{x_-, x_+\}_{x^i}^{x^f}$ of a channel $\{x_- m\, x_+\}_{x^i}^{x^f}$ is the solution of

the following mixed problem:

$$\left(\frac{\partial}{\partial t}+\frac{i}{\hbar}H\right)\psi=\delta(t-t^{i})\,\delta(x-x^{i}),$$
$$\psi[x_{\pm}(t)]=0,\quad t^{i}\leqslant t\leqslant t^{f} \tag{2.9}$$

upon the curvilinear band $(x_{-}(t), x_{+}(t))$, taken at the point (x^{f}, t^{f}).

Let the curves $x_{-}(t)$ and $x_{+}(t)$ be fixed and x^{f}, t^{f} be considered as variables (x,t). Then the rough measure for the given band (x_{-}, x_{+}) is a function $\mu(x,t)$.

The rough measure $\mu_{H}(G^{x^{f}}_{x^{i}})$ for the whole group corresponds to the infinite band with $x_{\mp}(t)=\pm\infty$, $t^{i}\leqslant t\leqslant t^{f}$, that is to say,

$$\mu_{H}(G^{x^{f}}_{x^{i}})=K(x^{f}t^{f}|H\mid x^{i}t^{i}), \tag{2.10}$$

because $K(xt\,|H\mid x^{i}t^{i})$ is vanishing, as a distribution, for $x\to\pm\infty$, and it is the solution of the problem (2.9) for the infinite band.

The zero boundary conditions in the mixed problem (2.9) result from the superposition principle for the partial amplitudes, and the value of the Green's function at the boundary curves $x_{\pm}(t)$ is determined by contributions from a denumerable set of other channels, inside which are the points.

The decomposition of the whole set of paths, $G^{x^{f}}_{x^{i}}$, over the complete systems of independent channels and the concept of the channel partial amplitudes is, seemingly, a more accurate definition of Feynman's path integral representing the kernel $K(x^{f}t^{f}|H\mid x^{i}t^{i})$ as a sum of amplitudes contributed by the paths. As to a single path amplitude, it is unreasonable to define such a thing, just as it is unreasonable to introduce a length for a single point on a line.

The rough measure of a beam, $B^{x^{f}(k)}_{x^{i}}[x^{(j)}_{-}, x^{(j)}_{+}]$, is expressed directly in terms of the $(k-1)$-fold convolution of Green's functions,

$$\mu_{H}(B^{x^{f}(k)}_{x^{i}}[x^{(j)}_{-}, x^{(j)}_{+}])$$
$$=\int_{x^{(k-1)}_{-}}^{x^{(k-1)}_{+}}\int_{x^{(1)}_{-}}^{x^{(1)}_{+}}K(x^{f}|H\mid x^{(k-1)})\,dx^{(k-1)}$$
$$\times K(x^{(k-1)}|H\mid x^{(k-2)})\cdots dx^{(1)}K(x^{(1)}|H\mid x^{i}), \tag{2.11}$$

where we have omitted the arguments $t^{(j)}$ for brevity.

There is the last point relevant to the construction exposed in Sec. I, which has not been mentioned until now, namely, the normalizing subgroup. Actually, it will be determined below (Sec. III) in such a way that the condition (1.12) is satisfied.

The group of closed paths, $G^{x^{i}t^{f}}_{x^{i}t^{i}}$, is especially important. If we have the motion in the infinite axis, or in an unbounded flat space, all the closed paths can be deformed continuously to the identity element of the group, i.e., that describing the point at rest, $x(t)=x^{i}$; that is to say, the fundamental group of $G^{x^{i}}_{x^{i}}$ is trivial. In general, the fundamental group plays a very important role in the definition of the path integral.

(3) A fundamental property of the path integral for the Schrödinger equation is that the generalized rough measures $d\mu_{H_1}$ and $d\mu_{H_2}$, corresponding to different Hamiltonians, H_1 and H_2, for a fixed particle mass m are related in a clear way. In fact, it is sufficient to express the rough measure for the general Hamiltonian μ_{H}, in terms of μ_{H_0}, where $H_0=-(\hbar^{2}/2m)\,\partial^{2}/\partial x^{2}$ is the free particle Hamiltonian. To this end, we will find an approximation for $\mu_{H}(\{x_{-}, x_{+}\}^{x}_{x^{i}})$, the solution of the mixed problem (2.9), which is related to the free particle solution,

$$\nu(x,t)=\exp\left[\frac{i}{\hbar}\int_{\gamma}\omega-\frac{i}{\hbar}\int_{(\gamma\delta^{-1}\hat{\gamma}^{-1})}d\omega\right]\mu_{H_0}(\{x_{-}, x_{+}\}^{x}_{x^{i}}), \tag{2.12}$$

where

$$\omega=f(x,t)\,dx+\varphi(x,t)\,dt \tag{2.13}$$

is an as-yet unknown differential form, $\hat{\gamma}$ is a certain fixed path within the channel $\{x_{-}, x_{+}\}_{x^{i}}$, γ is an arbitrary path within the channel $\{x_{-}, x_{+}\}^{x}_{x^{i}}$, δ is the segment connecting the points $(\hat{x}, t^{f})$ and (x, t^{f}), and $(\gamma\delta^{-1}\hat{\gamma}^{-1})$ is the two-dimensional simplex, the boundary of which is the closed curve $\gamma\delta^{-1}\hat{\gamma}^{-1}$. Let γ' and γ be some paths lying inside the channel $\{x_{-}, x_{+}\}^{x}_{x^{i}}$. In view of the Stokes formula, i.e.,

$$\int_{(\gamma'\delta^{-1}\hat{\gamma}^{-1})}d\omega-\int_{(\gamma\delta^{-1}\hat{\gamma}^{-1})}d\omega=\int_{\gamma'}\omega-\int_{\gamma}\omega, \tag{2.14}$$

the exponential factor in Eq. (2.12) depends only on the point (x,t) for fixed $\hat{\gamma}$, and is independent on the curve γ. The simplest choice of the latter is $\gamma=\hat{\gamma}+\delta$; then the two-dimensional simplex $(\gamma\delta^{-1}\hat{\gamma}^{-1})$ degenerates to a curvilinear segment. The expression in Eq. (2.12) satisfies the equation

$$\left(\frac{\partial}{\partial t}+\frac{i}{\hbar}H\right)\nu(x,t)$$
$$=\exp\left[\frac{i}{\hbar}\int_{\hat{\gamma}+\delta}\omega\right]\left\{\left(\frac{\partial}{\partial t}+\frac{i}{\hbar}H_0\right)\mu_{H_0}\right.$$
$$+\left(\frac{f}{m}-A\right)\frac{\partial\mu_{H_0}}{\partial x}+\frac{i}{\hbar}\left[\varphi(\hat{x},t)+\int_{\hat{x}}^{x}\frac{\partial f}{\partial t}dx\right.$$
$$\left.\left.-\frac{f^{2}}{2m}+\frac{\hbar}{2im}\frac{\partial f}{\partial x}+V\right]\mu_{H_0}\right\}. \tag{2.15}$$

The 1-form ω should be chosen in such a way that the coefficient at the term $\partial\mu_{H_0}/\partial x$ in Eq. (2.15) equals zero and the coefficient at μ_{H_0} vanishes at $x=\hat{x}$. The result is

$$f(x,t)=mA(x,t),$$
$$\varphi(x,t)=\tfrac{1}{2}mA^{2}(x,t)+\tfrac{1}{2}i\hbar\frac{\partial A}{\partial x}-V(x,t)-\int_{\hat{x}}^{x}\frac{\partial f}{\partial t}dx,$$
$$\omega=mA(x,t)\,dx+\left[\tfrac{1}{2}mA^{2}(x,t)+\frac{1}{2}\left(\hbar\frac{\partial A}{\partial x}\right)\right.$$
$$\left.-V(x,t)-\int_{\hat{x}}^{x}m\left(\frac{\partial A}{\partial t}\right)dx\right]dt. \tag{2.16}$$

With this choice of the form ω, the function $\nu(x,t)$ is the solution of the following mixed problem:

$$\left(\frac{\partial}{\partial t}+\frac{i}{\hbar}H+\frac{i}{\hbar}[\varphi(x,t)-\varphi(\hat{x},t)]\right)\nu(x,t)$$
$$=\delta(t-t^{i})\,\delta(x-x^{i}),\quad \nu(x_{\pm}(t))=0, \tag{2.17}$$

for the curvilinear band $[x_{-}(t), x_{+}(t)]$. So it is the solution of the same problem as that for μ_{H}, but with the difference

$\varphi(x,t) - \varphi(\hat{x},t)$ added to the potential. When the band (x_-, x_+) is contracted to the segment $\hat{\gamma}$, this correction to the potential $V(x,t)$ goes to zero; thus

$$\lim[\nu(x,t)/\mu_H(x,t)] = 1, \tag{2.18}$$

and

$$\lim\left(\frac{\mu_H(x,t)}{\mu_{H_0}(x,t)}\right) = \exp\left(\frac{i}{\hbar}\int_\gamma \omega^L\right), \tag{2.19}$$

where $\omega^L = \lim \omega$, i.e.,

$$\omega^L = mA(x,t)\,dx + \left[-V(x,t) + \tfrac{1}{2}mA^2(x,t) + \tfrac{1}{2}i\hbar\frac{\partial A(x,t)}{\partial x}\right]dt. \tag{2.20}$$

For two different Hamiltonians, H_1 and H_2, $H_2 - H_1 = -(\hbar/i)(A_2 - A_1)\partial/\partial x + V_2 - V_1$, we get from Eq. (2.19)

$$\lim\left[\frac{\mu_{H_2}(x,t)}{\mu_{H_1}(x,t)}\right] = \exp\left[\frac{i}{\hbar}\int_\gamma(\omega_2^L - \omega_1^L)\right], \tag{2.21}$$

where ω_1^L and ω_2^L are the 1-forms corresponding to H_1 and H_2.

The Hamiltonian H is Hermitian, and so the form ω^L is real, if $\operatorname{Im} V(x,t) = \frac{1}{2}\hbar\,\partial A(x,t)/\partial x$; in this case

$$\omega^L = mA(x,t)\,\partial x + [-\operatorname{Re} V(x,t) + \tfrac{1}{2}mA^2(x,t)]\,dt, \quad \operatorname{Im} V = \tfrac{1}{2}\hbar\frac{\partial A}{\partial x}. \tag{2.22}$$

In particular, when the forces are potential, one has

$$H_V = H_0 + V(x,t), \quad \omega^L = -V(x,t)\,dt. \tag{2.23}$$

In the following we shall suppose that the Hamiltonians we deal with are Hermitian.

Equations (2.19), (2.23), and (2.21) lead to the following relations between the generalized rough measures:

$$\begin{aligned} d\mu_H(\gamma) &= \exp\left[\frac{i}{\hbar}\int_\gamma \omega^L\right] d\mu_{H_0}(\gamma), \\ d\mu_{H_V}(\gamma) &= \exp\left[-\frac{i}{\hbar}\int_\gamma V(x,t)\,dt\right] d\mu_{H_0}(\gamma), \\ d\mu_{H_2}(\gamma) &= \exp\left[\frac{i}{\hbar}\int_\gamma(\omega_2^L - \omega_1^L)\right] d\mu_{H_1}(\gamma), \end{aligned} \tag{2.24}$$

where $d\mu_H(\gamma)$ is the measure for an infinitely narrow channel the limit of which is the path γ. Respectively, one has integral relations between the measures for an arbitrary channel $\{x_-, x_+\}_{x^i}^{x^f}$, for instance,

$$\mu_{H_2}(\{x_-,x_+\}_{x^i}^{x^f}) = \int_{\{x_-,x_+\}_i^f} \exp\left[\frac{i}{\hbar}\int_\gamma(\omega_2^L - \omega_1^L)\right] d\mu_{H_1}(\gamma). \tag{2.25}$$

The integral over the whole group $G_{x^i}^{x^f}$ is

$$K(x^f t^f|H_2|x^i t^i) = \int \exp\left[\frac{i}{\hbar}\int_\gamma(\omega_2^L - \omega_1^L)\right] d\mu_{H_1}(\gamma), \tag{2.26}$$

$$K(x^f t^f|H_V|x^i t^i) = \int \exp\left[-\frac{i}{\hbar}\int_\gamma V(x,t)\,dt\right] d\mu_{H_0}(\gamma). \tag{2.27}$$

The latter equality is the famous Kac–Feynman formula which is written down in terms of the generalized rough measure.

If the differential form is a total differential, $\omega_2^L - \omega_1^L = dF(x,t)$,

$$K(x^f t^f|H_2|x^i t^i) = \exp[(i/\hbar)]F(x^f,t^f) - F(x^i,t^i)\}K(x^f t^f|H_1|x^i t^i). \tag{2.28}$$

The above relations hold also in the case of non-Hermitian Hamiltonians.

Equations (2.21)–(2.28) are relations between the integrals over the rough measures corresponding to different Hamiltonians for the same particle mass m; given the operators H_1 and H_2, the differential forms ω_1^L, ω_2^L, and their difference, $\omega_{21}^L = \omega_2^L - \omega_1^L$, are determined. Inversely, suppose we have an operator H_1 and an arbitrary differential form ω_{21}^L; then the Hamiltonian H_2 can be found unambiguously from the differential form $\omega_2^L = \omega_1^L + \omega_{21}^L$. For any Hamiltonian H_2 which is describing the motion of the mass m particle, the corresponding rough measure μ_{H_2} is determined as well as the integral over this measure. So one has a proof that the functional $\exp[(i/\hbar)\int_\gamma \omega_{21}^L]$ is integrable with the rough measure μ_{H_1} for an arbitrary differential form ω_{21}^L.

(4) The above derivation of Eq. (2.19), and, consequently, the subsequent reasoning, has a defect. Namely, it is only for rectifiable curves that the integral $\int_\gamma \omega$ is meaningful for $A(x,t) \neq 0$, and the Stokes formula is proved for such curves only. Meanwhile, the paths contributing substantially to the continual integral are not rectifiable, i.e., they have infinite length. The simplest way to see that this statement is true is as follows.

The following expression is known for the integral kernel in the free particle case (see also the next section where we present a group-theoretical derivation):

$$K_0(x^f t^f|x^i t^i) = [2\pi i\hbar(t^f - t^i)/m]^{-1/2} \times \exp\left[\frac{im}{2\hbar}\frac{(x^f - x^i)^2}{t^f - t^i}\right]. \tag{2.29}$$

Therefore, the integral

$$\int K_0(x^{k+1},t^{k+1}|x^k,t^k)\,dx^{k+1} = (\pi i)^{-1/2}\int \exp(iy^2)\,dy,$$
$$y = [2\hbar(t^{k+1} - t^k)/m]^{-1/2}(x^{k+1} - x^k) \tag{2.30}$$

is contributed mainly by the region $1/M < |y| < M$, where $M \gg 1$. In terms of the physical variables, the essential domain is $(2h\Delta t^k m^{-1})^{1/2}M^{-1} < \Delta x^k < (2\hbar\Delta t^k m^{-1})^{1/2}M$, where $\Delta t^k = t^{k+1} - t^k$, and $\Delta x^k = x^{k+1} - x^k$. In other words, the path integral is contributed by the curves where $\Delta x \sim (2\hbar m^{-1})^{1/2}(\Delta t)^{1/2}$. For any fixed $\epsilon > 0$ and $2\hbar\Delta t/m \to 0$ one has for such curves

$$\lim \sum |\Delta x^k|^{2+\epsilon} = 0, \quad \lim \sum |\Delta x^k|^{2-\epsilon} = \infty, \tag{2.31}$$
$$\sum \Delta x^k = x^f - x^i, \quad \sum \Delta t^k = t^f - t^i, \quad \Delta t^k > 0.$$

In view of these relations, the curvilinear integral $\int_\gamma f(x,t)\,dx$ must be defined as the limit of integral sums, the

errors in which are no higher than of the order of $\Delta x^{2+\epsilon}$ at every step. So the trapezium method must be used to evaluate the integral,

$$\int_\gamma f(x,t)\,dx = \lim \sum_\gamma \tfrac{1}{2}[f(x^k,t^k)+f(x^{k+1},t^{k+1})]\Delta x^k, \tag{2.32}$$

or, equivalently,

$$\int_\gamma f(x,t)\,dx = \lim \left\{ \sum_\gamma f(x^{k+1},t^{k+1})\Delta x^k - \tfrac{1}{2}\sum_\gamma f'_x(x^k,t^k)(\Delta x^k)^2 \right\}. \tag{2.33}$$

Let $\{\gamma_n\}$ be a sequence of polygonal lines refined into the curve γ, and all $(\Delta t^k)_n$ go to zero uniformly, as $n\to\infty$; then Eq. (2.32) leads to

$$\int_\gamma \omega = \lim_{n\to\infty} \int_{\gamma_n} \omega, \quad \omega = f(x,t)\,dx + \varphi(x,t)\,dt. \tag{2.34}$$

Hence we conclude that the Stokes formula, Eq. (2.14), is valid for the curves of the type considered, provided that the integral $\int_\gamma f(x,t)\,dx$ is calculated by means of Eqs. (2.32) and (2.33). With this refinement, the derivation given in Eqs. (2.4)–(2.28) is correct. In the following we shall concern only curves satisfying the conditions of Eq. (2.31), unless indicated otherwise.

Discarding the terms $\sim(\Delta x^k)^3$, and higher-order terms, one has

$$\left(1+\frac{i}{\hbar}mA_{k+1}\Delta x^k\right) = \exp\left[\frac{i}{2\hbar}m(A_k+A_{k+1})\right] \times\exp\left[\left(\frac{m^2}{2\hbar^2}A^2_{k+1}+\frac{im}{2\hbar}A'_k\right)(\Delta x^k)^2\right], \tag{2.35}$$

where $A_k = A(x^k,t^k)$, $A'_k = \partial A(x^k,t^k)/\partial x$, so

$$\lim \prod_{k=0}^{N-1}\left(1+\frac{i}{\hbar}mA_{k+1}\Delta x^k\right) = \exp\left[\frac{i}{\hbar}\int_\gamma mA(x,t)\,dx\right] \times\lim\exp\left[\sum_{k=0}^{N-1}\left(\frac{m^2}{2\hbar^2}A^2_{k+1}+\frac{im}{2\hbar}A'_k\right)(\Delta x^k)^2\right]. \tag{2.36}$$

This relation proves the existence of a limit on the lhs. Note that A_{k+1} cannot be replaced by A_k in the product.

Let us consider the average of $(x^{k+1}-x^k)^2$ taken with the measure μ_{H_0},

$$\langle(x^{k+1}-x^k)^2\rangle = \int (x^{k+1}-x^k)^2 K_0 \times(x^{k+1},t^{k+1}|x^k,t^k)\,dx^k = -\frac{\hbar}{im}(t^{k+1}-t^k). \tag{2.37}$$

Substituting the average value for $(\Delta x^k)^2$ in the rhs of Eq. (2.36), we get

$$\exp\left[\frac{i}{\hbar}\int mA(x,t)\,dx + \frac{1}{2}\left(mA^2(x,t)+i\hbar\frac{\partial A(x,t)}{\partial x}\right)dt\right], \tag{2.38}$$

and Eq. (2.19) is transformed to

$$\frac{d\mu_H(\gamma)}{d\mu_{H_0}(\gamma)} = \lim \prod_{k=0}^{N-1}\left(1+\frac{1}{\hbar}mA_{k+1}\Delta x^k\right) \times\exp\left[-\frac{i}{\hbar}\int_\gamma V(x,t)\,dt\right]. \tag{2.39}$$

It has been assumed here that $(\Delta x^k)^2$ in the rhs of Eq. (2.36) is replaced by the average values $\langle(\Delta x^k)^2\rangle$, and the points (x^k,t^k) are vertices of the polygonal line rectified into the curve γ.

(5) All the definitions and results are extended naturally to the multidimensional case where the particle motion proceeds in $\mathbb{R}_n$, $x=(x_1,\ldots,x_n)\in\mathbb{R}_n$. The Hamiltonian describing the particle motion in the n-dimensional Euclidean space is

$$H = -\frac{\hbar^2}{2m}\nabla^2 - \frac{\hbar}{i}A(x,t)\nabla + V(x,t), \tag{2.40}$$

and the corresponding Green function satisfies the equation

$$\left(\frac{\partial}{\partial t}+\frac{i}{\hbar}H\right)K = \delta(t-t^i)\delta(x-x^i). \tag{2.41}$$

The group $G_{t^i}^{t^f}$, and all the other groups, considered above, consist now of paths in $\mathbb{R}_n$ and the group $G_{t^i}^{t^f}(t)$, etc., contain graphs of the paths in the extended space $(\mathbb{R}_n,t)$. The space of the vector functions is the direct produce of the spaces of functions, so the channels in the group for the n-dimensional case are defined as direct products of the one-dimensional channels. To be more exact, let $x_\mp(t) = (x^1_\mp(t),\ldots,x^n_\mp(t))$ and $x^k_-(t) < x^k_+(t)$ for any t, $t^i\leqslant t\leqslant t^f$, and $k=1,\ldots,n$; then the channel $\{x_-,x_+\}_i^f$ in the group G_i^f is called the direct product of n channels $\{x^k_-,x^k_+\}_i^f$, and the band (x_-,x_+) is the direct product of the corresponding two-dimensional bands. The channels $\{x_-,x_+\}_{x^i}$ and $\{x_-,x_+\}_{x^i}^{x^f}$ for the groups $G_{x^it^i}^{t^ft^f}$ and $G_{x^i}^{x^f}$ are intersections of these groups with the channels $\{x_-,x_+\}_i^f$. The channel diameter is defined by

$$d[x_-(t),x_+(t)] = \sup\|x_+(-t)-x_-(t)\|, \quad t^i\leqslant t\leqslant t^f. \tag{2.42}$$

Decompositions of the groups of paths in the one-dimensional spaces over the complete systems of independent channels induce the decomposition of the group of paths in the n-dimensional space over a complete system of independent channels, every channel being a direct product of n independent one-dimensional channels.

The partial amplitude contributed by a channel $\{x_-,x_+\}_{x^i}^{x^f}$ is defined just as in the one-dimensional case. Equation (2.8) holds because of the same reasons as in the one-dimensional case, and the partial amplitude is taken as the generalized rough measure; it equals the solution of the mixed problem in the n-dimensional band $(x_-,x_+)(t)$ for Eq. (2.41) with zero boundary conditions at the boundary $\partial(x_-,x_+)(t)$, taken at the point (x^f,t^f).

(6) Let H_1 and H_2 be Hamiltonians describing motions of the particle with a mass m under different forces, and

$$H_2-H_1 = -\frac{\hbar}{i}[A_2(x,t)-A_1(x,t)]\nabla + V_2(x,t)-V_1(x,t). \tag{2.43}$$

It is found that in the multidimensional case the ratio of the corresponding generalized rough measures for any channel $\{x_-,x_+\}_{x_i}^{x^f}$ as the channel is contracted to a path γ, tends to a limit which is given by the multidimensional analog of that given in Eq. (2.19),

$$\frac{d\mu_{H_2}(\gamma)}{d\mu_{H_1}(\gamma)} = \exp\left[\frac{i}{\hbar}\int_\gamma (\omega_2^L - \omega_1^L)\right], \tag{2.44}$$

$$\omega_j^L = m\mathbf{A}_j d\mathbf{x} + [\,-V_j + \tfrac{1}{2}mA_j^2 + \tfrac{1}{2}i\hbar\nabla\mathbf{A}_j\,]\,dt, \quad j = 1,2.$$

Moreover, all the formulas Eqs. (2.19)–(2.28) have their respective counterparts, to be obtained substituting f, A, $\partial A/\partial x$ by the vectors $\mathbf{f}$, $\mathbf{A}$, and ∇A. In the multidimensional case, however, the derivation of Eq. (2.44) meets an additional obstacle. To elucidate the matter, we shall consider the three-dimensional case with $H_1 = H_0$ and an arbitrary $H_2 = H$.

Let us suppose, first, that the curl of $\mathbf{A}(x,t)$ is zero. As in the one-dimensional case, the rough measure for a channel which is the solution of the mixed problem for the operator $\partial/\partial t + (i/\hbar)H$ in the band $(x_-,x_+)(t)$ is approximated with the expression

$$\nu(x,t) = \exp\left[\frac{i}{\hbar}\int_\gamma \omega - \frac{i}{\hbar}\int_{(\gamma\delta^{-1}\hat\gamma^{-1})} d\omega\right]\mu_{H_0}(\{x_-,x_+\}_{x^t}^{x}). \tag{2.45}$$

Here γ is an arbitrary path in the channel $\{x_-,x_+\}_{x^t}^{x}$, γ is a fixed curve in the channel $\{x_-,x_+\}_{x^i,t^i}^{t^f}$ with the end point at $(\hat x^f t^f)$, δ is an arbitrary curve in the cross section of the channel $\{x_-,x_+\}_{x^i,t^i}^{t^f}$ by the hyperplane $t = t^f$, which connects the points $\hat x^f$ and x, $(\gamma\delta^{-1}\hat\gamma^{-1})$ is an arbitrary two-dimensional simplex, the boundary of which is the closed loop $\gamma\delta^{-1}\hat\gamma^{-1}$, and $\omega = \mathbf{f}(x,t)\,d\mathbf{x} + \varphi(x,t)\,dt$ is a differential form, as yet unknown. The simplex $(\gamma\delta^{-1}\hat\gamma^{-1})$ is oriented in such a way that the Stokes formula is

$$\int_\gamma \omega - \int_{(\gamma\delta^{-1}\hat\gamma^{-1})} d\omega = \int_{\hat\gamma}\omega + \int_\delta \omega. \tag{2.46}$$

If the curl of $\mathbf{f}$ is zero and $\hat\gamma$ is fixed, both parts in Eq. (2.46) depend on the point (x,t^f) only, and are independent of the curves γ and δ. In this case

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \frac{i}{\hbar}H\right)\nu(x,t) &= \exp\left[\frac{i}{\hbar}\int_\gamma \omega - \frac{i}{\hbar}\int_{(\gamma\delta^{-1}\hat\gamma^{-1})} d\omega\right]\\ &\times\left\{\left(\frac{\partial}{\partial t} + \frac{i}{\hbar}H_0\right)\mu_{H_0} + (-\mathbf{A} + m^{-1}\mathbf{f})\nabla\mu_{H_0}\right.\\ &+ \frac{i}{\hbar}\left[\varphi(\hat x,t^f) + \int_\delta \frac{\partial \mathbf{f}}{\partial t}dx\right.\\ &\left.\left. - (2m)^{-1}\mathbf{f}^2 + \frac{\hbar}{(2im)\nabla\mathbf{f}} + V\right]\mu_{H_0}\right\}.\end{aligned} \tag{2.47}$$

The requirements that the coefficient at $\nabla\mu_{H_0}$ and the coefficient at μ_{H_0} for $x = x^f$ must vanish identically determine the form ω,

$$\mathbf{f}(x,t) = m\mathbf{A}(x,t),$$

$$\varphi(x,t) = \tfrac{1}{2}m\mathbf{A}^2(x,t + \tfrac{1}{2}i\hbar\nabla\mathbf{A}(x,t) - V(x,t) - \int_\delta \frac{\partial f}{\partial t}dx, \tag{2.48}$$

so that the vector field $\mathbf{f}(x,t)$ has, in fact, zero curl. Further we can repeat exactly the derivation presented for Eq. (2.19). Thus assuming that curl $\mathbf{A}(x,t) = 0$ we have proved that

$$\frac{d_{\mu_H}(\gamma)}{d_{\mu_{H_0}}(\gamma)} = \exp\left(\frac{i}{\hbar}\int_\gamma \omega^L\right), \tag{2.49}$$

$$\omega^L = m\mathbf{A}(x,t)d\mathbf{x} + [\,-V(x,t) + \tfrac{1}{2}mA^2(x,t) + \tfrac{1}{2}i\hbar\nabla\mathbf{A}]\,dt.$$

If curl $A(x,t) \neq 0$, the method we propose is as follows. Let us subdivide the group $G_{x^i,t^i}^{t^f}$ into arbitrarily narrow channels; in every channel we chose a fixed path $\hat\gamma$ and approximate the field $\mathbf{A}(x,t)$ in every curvilinear band $(x_-,x_+)(t)$ with the field $\tilde{\mathbf{A}}(x,t)$ which is linear in $\mathbf{x}$ for any t and tangential to the field $A(x,t)$ at the points belonging to the path $\hat\gamma$. The Hamiltonian with the modified field is denoted by $\tilde H$. Now curl $\tilde{\mathbf{A}} = 0$, so that Eq. (2.49) for the ratio of the rough measures corresponding to the Hamiltonians $\tilde H$ and H_0 is true. On the other hand, if the channels are infinitely narrow, one has, evidently,

$$\lim\left[\mu_{\tilde H}(\{x_-,x_+\}_{x^i}^{x})/\mu_H(\{x_-,x_+\}_{x^i}^{x})\right] = 1. \tag{2.50}$$

In the Hermitian case Im $V(x,t) = \frac{1}{2}\hbar\nabla A(x,t)$, and

$$\omega^L = mA\,dx + (-\operatorname{Re} V(x,t) + \tfrac{1}{2}mA^2)\,dt. \tag{2.51}$$

If we require that the approximation with $\tilde A$ does not spoil the Hermitian property, we should modify also the imaginary part of the potential to preserve the identity Im $V = \frac{1}{2}\hbar\nabla A$. The meaning of the integral $\int_\gamma \mathbf{A}(x,t)\,d\mathbf{x}$ was actually explained in the subsection (4); The extension to the multidimensional case is trivial.

III. PARALLEL TRANSLATION ON THE GROUP G_t', THE ACTION FUNCTIONAL AND THE GENERALIZED INVARIANT MEASURE

In operations with the functional integrals one needs some rules to perform the change of variables, as one needs such rules in the usual integral calculus. For theWiener integral the change of variables was considered in early works by Cameron and Martin.[7] We shall discuss here the change of variables induced by the parallel translations; more general cases are concerned in Sec. IV.

Suppose we have a rough measure $\mu^{(1)}$ and the functional integral, defined for a locally noncompact group $G^{(1)}$, according to the method of Sec. I,

$$J = \int_{U^{(1)}} \varphi(g)\,d\mu^{(1)}(g). \tag{3.1}$$

Let $G^{(2)}$ be another locally noncompact group with a rough measure $\mu^{(2)}$, and we assume that there is a one-to-one continuous correspondence between the group elements, given in terms of an operator T, which performs a channel-to-channel mapping, $U^{(2)} = TU^{(1)}$, where $U^{(j)}$ is a channel in the group $G^{(j)}$, $j = 1,2$. Suppose the functional is determined in the channel $U^{(2)}$,

$$\frac{d\mu^{(4)}(T^{-1}g^{(2)})}{d\mu^{(2)}(g^{(2)})} = \lim \frac{\mu^{(4)}(T^{-1}V_\alpha)}{\mu^{(2)}(V_\alpha)}, \tag{3.2}$$

as a subchannel V_α in the channel $U^{(2)}$ is contracted to a group element $g^{(2)}$. We assume that the limit exists and is uniform in $U^{(2)}$. Under the above conditions, the change of variables is

$$\int_{U^{(1)}} \varphi(g)\, d\mu^{(1)}(g) = \int_{TU^{(1)}} \varphi(T^{-1}g) \frac{d\mu^{(1)}(T^{-1}g)}{d\mu^{(2)}(g)}\, d\mu^{(2)}(g). \tag{3.3}$$

Presented in such a general form, the change of variables for the functional integrals looks exactly like that for the usual integrals. It is of interest, however, to understand the general formula for the case where besides the generalized rough measure $d\mu^{(2)}$ another measure exists in the group $G^{(2)}$, which is of the same type as $d\mu^{(1)}$, and it is possible to express $d\mu^{(2)}$ in terms of $d\mu^{(1)}$.

Let us consider a simple but very important transformation,

$$T\colon (y,t) = (x + y_0(t),t), \quad T^{-1}\colon (x,t) = (y - y_0(t),t), \tag{3.4}$$

where $y_0(t)$ is a fixed function. This is a parallel translation in G_t^f, for which any group $G_{x^i}^{x^f}$ is transformed to $G_{x^i+y_0(t^i)}^{x^f+y_0(t^f)}$, and the Hamiltonian H_0 in the x-space is transformed to

$$H_1 = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + \frac{\hbar}{i}\dot{y}_0(t)\frac{\partial}{\partial y} \tag{3.5}$$

in the y-space. We shall use the following notations: the paths $x = x(t), y = y(t), y = y_0(t)$ are denoted by $\gamma_x, \gamma_y, \gamma_{y_0}$, respectively; the corresponding integrals are $\int \dot{x}(t)\, dx = \int \dot{\gamma}_x\, d\gamma_x$, etc. The transformation we are dealing with is written as $T\colon \gamma_y = \gamma_x + \gamma_{y_0}$. The measure μ_{H_0} for any channel $\{x_-,x_+\}_{x^i}^{x^f}$ is equal to the measure μ_{H_1} of the image of this channel,

$$\begin{aligned} &\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f}) = \mu_{H_1}(T\{x_-,x_+\}_{x^i}^{x^f}), \\ &d\mu_{H_0}(\gamma_x) = d\mu_{H_1}(\gamma_y). \end{aligned} \tag{3.6}$$

In particular,

$$K_0(x^f,t^f|x^i,t^i) = K(y^f,t^f|H_1|y^i,t^i). \tag{3.7}$$

This is true because both the measures are solutions of the same mixed problem (2.9) in the variables (x,t) and (y,t). Because of Eqs. (2.19)–(2.20) we have

$$\begin{aligned} &d\mu_{H_1}(\gamma_y) = \exp\left[\frac{i}{\hbar}\int_{\gamma_y}\omega^L\right] d\mu_{H_0}(\gamma_y), \\ &\omega^L = -m\dot{y}_0\, dy + \tfrac{1}{2}m\dot{y}_0^l\, dt. \end{aligned}$$

From Eqs. (3.6) and (3.7) one gets the following rule for the change of variables corresponding to the parallel translation:

$$\begin{aligned} &\int_{\{x_-,x_+\}_{x^i}^{x^f}} \varphi(\gamma)\, d\mu_{H_0}(\gamma) \\ &= \exp\left[\frac{i}{\hbar}\int_{t^i}^{t^f}\frac{1}{2}m\dot{\gamma}_0^2\, dt\right] \int_{\{x_-,x_+\}_{x^i+\gamma_0}} \varphi(\gamma-\gamma_0) \\ &\quad\times \exp\left[-\frac{i}{\hbar}m\int_\gamma \dot{\gamma}_0\, dy\right] d_{H_0}(\gamma). \end{aligned} \tag{3.8}$$

In this case $d\mu_{H_0}$ plays the role of the rough measure in the group $G_{x^i}^{x^f}$. $d\mu_{H_1}$ is the rough measure in the group $TG_{x^i}^{x^f}$, and it is expressed via the measure $d\mu_{H_0}$ in the latter group.

This rule enables one to get the explicit expression for the evolution kernel $K_0(x^f,t^f|x^i,t^i)$ in the case of a free particle based on the group-theoretical arguments only. The parallel translation to be performed is $\gamma \to \gamma + x_i^f(t)$, where $x_i^f(t) = (x^f - x^i)(t - t^i)/(t^f - t^i) + x^i$. So one has $G_{x^it^i}^{x^ft^f} \to G_{x^i}^{x^f}$, $dy = dx + \dot{x}_i^f(t)\, dt$, and

$$\frac{1}{2}\int_\gamma (\dot{x}_i^f(t))^2\, dt - \int_\gamma \dot{x}_i^f(t)\,[dx\,\dot{x}_i^f(t)\, dt] = -\frac{(x^f - x^i)^2}{2(t^f - t^i)}, \tag{3.9}$$

so that

$$\begin{aligned} K_0(x^i,t^f|x^i,t^i) &= \mu_{H_0}(G_{x^it^i}^{x^ft^f}) \\ &= \exp\left[-\frac{im(x^f - x^i)^2}{2\hbar(t^f - t^i)}\right]\mu_{H_0}(G_{x^i}^{x^f}). \end{aligned} \tag{3.10}$$

Besides,

$$\int \mu_{H_0}(G_{x^i}^{x^f})\, dx^f = 1, \tag{3.11}$$

as the time derivative of the integral is zero because of the Schrödinger equation. From Eqs. (3.10) and (3.14) one gets

$$\begin{aligned} &\mu_{H_0}(G_{x^it^i}^{x^ft^f}) = (2\pi i\hbar T/m)^{-1/2}, \quad T = t^f - t^i, \\ &\mu_{H_0}(G_{x^i}^{x^f}) = \mu_{H_0}(G_{x^it^i}^{x^ft^f})\exp\left[\frac{im(x^f - x^i)^2}{2\hbar T}\right]. \end{aligned} \tag{3.12}$$

Equation (3.10) is valid for any channel $\{x_-,x_+\}_{x^i}^{x^f}$.

$$\begin{aligned} &\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f}) \\ &\quad= \exp\left[\frac{im(x^f - x^i)^l}{2\hbar(t^f - t^i)}\right]\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f} - x_i^f(t)). \end{aligned} \tag{3.13}$$

Note also the change of variables for the general local transformation:

$$T\colon (y,t) = (f_T(x,t),t), \quad T^{-1}\colon (x,t) = (f_T^{-1}(y,t),t),$$

which is an automorphism of the infinite band $t^i \leqslant t \leqslant t^f$. This mapping transforms the wave functions $\psi(x,t)$, satisfying the Schrödinger Eq. (2.1) into the wave functions $\psi_1(y,t) = \psi[x(y,t),t]\,(\partial x/\partial y)^{1/2}$, that satisfy the Schrödinger equation in the variables (y,t), $-(\hbar/i)(\partial/\partial t)\psi_1(y,t) = H_1\psi_1(y,t)$. Hence one gets

$$K(y^f,t^f|H_1|y^i,t^i) = \left(\frac{\partial x^f}{\partial y^f}\right)^{1/2} K(x^f,t^f|H|x^i,t^i)\left(\frac{\partial x^i}{\partial y^i}\right)^{1/2}.$$

In this particular case Eq. (3.3) acquires the form

$$\begin{aligned} \int_{\{x_-,x_+\}_{x^i}^{x^f}} \varphi(\gamma)\, d\mu_H(\gamma) &= \left[\frac{\partial f_T(x^f,t^f)}{\partial x^f}\cdot\frac{\partial f_T(x^i,t^i)}{\partial x^i}\right]^{1/2} \\ &\quad\times \int_{T\{x_-,x_+\}_{x^i}^{x^f}} \varphi(T^{-1}\gamma)\, d\delta\mu_{H_1}(\gamma). \end{aligned}$$

(2) Equation (3.6) is a symbolic form of the relations

$$\begin{aligned} &\frac{\mu_{H_1}(\{y_-,y_+\}_{y^i}^{y^f})}{\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f})} = 1, \\ &\lim \frac{\mu_{H_1}(\{y_-,y_+\}_{y^i}^{y^f})}{\mu_{H_0}(\{y_-,y_0\}_{y^i}^{y^f})} = \exp\left(\frac{i}{\hbar}\int_{\gamma_y}\omega^L\right), \end{aligned} \tag{3.14}$$

as the channel $\{x_-,x_+\}_{x^i}^{x^f}$ is contracted to the curve γ_x. The derivative $\dot{y}_0(t)$ is independent of y, so in accordance with subsection (4) of Sec. II, the integral $\int_{\gamma_y} \dot{y}_0(t)\,dy$ does exist, and the integral $\int_{\gamma_y} \omega^L$ exists if the function $\dot{y}_0^2(t)$ is integrable on the segment (t^i,t^f). So the limit $d\mu_{H_0}(\gamma_x+\gamma_0)/d\mu_{H_0}(\gamma_x)$ exists if the parallel translation is given by a curve with a square-integrable derivative. Hence we conclude that the normalizing subgroup G' in $G_{x^i}^{x^f}$ [cf. Eqs. (1.12)–(1.13)] can be taken as the subgroup of all paths with square-integrable derivatives. If $\dot{\gamma}_x^2$ is integrable, the function $\dot{\gamma}_y^2=(\dot{\gamma}_x+\dot{\gamma}_0)^2$ is also integrable, and

$$-\frac{i}{\hbar}\int_{\gamma_y}\omega^L=\frac{i}{\hbar}\int_{t^i}^{t^f}\frac{1}{2}m\dot{\gamma}_y^2\,dy-\frac{i}{\hbar}\int_{t^i}^{t^f}\frac{1}{2}m\dot{\gamma}_x^2\,dt\,. \tag{3.15}$$

The action functional $S_{H_0}(\gamma_x)$ on the curve $\gamma_x\subset G'$ is obtained from Eqs. (3.14) and (1.12),

$$S_{H_0}(\gamma_x)=\int_{t^i}^{t^f}\dot{\gamma}_x^2\,dt+F(T)\,,\quad T=t^f-t^i\,, \tag{3.16}$$

where $F(T)$ is an arbitrary term depending on the total motion time.

Since the integral $\int_{\gamma_y}\omega^L$ is converging, for any two curves γ_1 and γ_2, belonging to the same coset of the group $G_{x^i}^{x^f}$ by the normalizing subgroup G', we have

$$S_{H_0}(\gamma_2)-S_{H_0}(\gamma_1)=\int_{t^i}^{t^f}m\dot{\gamma}_{21}\,d\gamma_2-\int_{t^i}^{t^f}\tfrac{1}{2}m(\dot{\gamma}_{21})^2\,dt\,,\qquad \gamma_{21}=\gamma_2-\gamma_1\,, \tag{3.17}$$

and this difference is finite. If $\gamma_1,\gamma_2\in G'$,

$$S_{H_0}(\gamma_2)-S_{H_0}(\gamma_1)=\frac{1}{2}m\int_{t^i}^{t^f}(\dot{\gamma}_2\,d\gamma_2-\dot{\gamma}_1\,d\gamma_1)\,, \tag{3.18}$$

so it is reasonable to write down Eq. (3.17) always in the form of (3.18). If, however, γ_1 and γ_2 do not belong to G', Eq. (3.17) is more adequate.

The subgroups of paths with square-integrable derivatives within the group G_i^f, or within any group of this type we consider now, will be also denoted by G', a confusion would hardly arise. For any two paths γ_1,γ_2 belonging to the same coset of G_i^f by G', we have

$$\exp\left[-\frac{i}{\hbar}S_{H_0}(\gamma_2)\right]d\mu_{H_0}(\gamma_2)=\exp\left[-\frac{i}{\hbar}S_{H_0}(\gamma_1)\right]d\mu_{H_0}(\gamma_1)\,. \tag{3.19}$$

The action functional $S_H(\gamma)$ for a general Hamiltonian H, Eq. (2.1), is obtained from $S_{H_0}(\gamma)$ by means of the identity

$$\frac{d\mu_H(\gamma_2)}{d\mu_H(\gamma_1)}=\frac{d\mu_H(\gamma_2)/d\mu_{H_0}(\gamma_2)}{d\mu_H(\gamma_1)/d\mu_{H_0}(\gamma_1)}\cdot\frac{d\mu_{H_0}(\gamma_2)}{d\mu_{H_0}(\gamma_1)} \tag{3.20}$$

and Eqs. (2.19)–(2.20). So we get

$$S_H(\gamma_2)-S_H(\gamma_1)=\left[S_{H_0}(\gamma_2)+\int_{\gamma_2}\omega^L\right]-\left[S_{H_0}(\gamma_1)+\int_{\gamma_1}\omega^L\right],\qquad \gamma_2-\gamma_1\in G'\,, \tag{3.21}$$

and for $\gamma\in G'$,

$$S_H(\gamma)=S_{H_0}(\gamma)+\int_\gamma\omega^L\,. \tag{3.22}$$

It follows from Eqs. (3.21) and (2.19) that Eq. (3.19) is also true for the Hamiltonian H,

$$\exp\left[-\frac{i}{\hbar}S_H(\gamma_2)\right]d_{\mu_H}(\gamma_2)=\exp\left[-\frac{i}{\hbar}S_H(\gamma_1)\right]d\mu_H(\gamma_1),\ (\gamma_2-\gamma_1)\in G^1. \tag{3.23}$$

The term $F(T)$ is present in S_H due to S_{H_0}, and so it is independent of H.

(3) The action functional $S_H(\gamma)$ enables one to apply the method of Sec. I to the group $G_{x^i}^{x^f}$ and to determine a generalized measure on the group which is invariant under the parallel translations,

$$d\sigma_H=\exp\left[-\frac{i}{H}S_H(\gamma)\right]d\mu_H(\gamma),$$

$$d\sigma_H(\gamma_2)=d\sigma_H(\gamma_1),\quad (\gamma_2-\gamma_1)\in G^1. \tag{3.24}$$

The generalized invariant measure is independent of the lower-order terms in the Hamiltonian,

$$\exp\left[-\frac{i}{\hbar}S_{H_2}(\gamma)\right]d\mu_{H_2}(\gamma)=\exp\left[-\frac{i}{\hbar}S_{H_1}(\gamma)\right]d\mu_{H_1}(\gamma),\quad d\sigma_{H_2}(\gamma)=d\sigma_{H_1}(\gamma), \tag{3.25}$$

it is determined by a single parameter,

$$\kappa=[\hbar(t^f-t^i)/m]^{1/2},$$

which is present in $d\sigma_{H_0}(\gamma)=d\sigma_\kappa(\gamma)$.

The integral over the generalized rough measure μ_H, defined in the framework of the general theory of Sec. I, can be written with the generalized invariant measure,

$$\mu_H(G_i^f)=\int\exp\left[\frac{i}{\hbar}S_H(\gamma)\right]d\sigma_\kappa(\gamma),$$

$$(\varphi,\mu_H)=\int\varphi(\gamma)\exp\left[\frac{i}{\hbar}S_H(\gamma)\right]d\sigma_\kappa(\gamma), \tag{3.26}$$

$$\int_{\{x_-,x_+\}_{x^i}^{x^f}}\varphi(\gamma)\,d\mu_H(\gamma)=\int\varphi(\gamma)\exp\left[\frac{i}{\hbar}S_H(\gamma)\right]d\sigma_\kappa(\gamma).$$

The results discussed above are valid for Hermitian and non-Hermitian Hamiltonians. Besides, in the Hermitian case, two following equalities hold for the limits of the rough measures of channels, as they are contracted to a trajectory, $\{x_-,x_+\}_{x^i}^{x^f}\to\gamma$:

$$\lim\left|\frac{\mu_H(T\{x_-,x_+\}_{x^i}^{x^f})}{\mu_H(\{x_-,x_+\}_{x^i}^{x^f})}\right|=1,\quad \lim\left|\frac{\mu_H(\{x_-,x_+\}_{x^i}^{x^f})}{\mu_{H_0}(\{x_-,x_+\}_{x^i}^{x^f})}\right|, \tag{3.27}$$

where T is the parallel translation by a curve $f(t)\in G'$. In other words, the contributions from different paths to the transition amplitude are different in the phase shifts only;

Eqs. (3.27) are a more accurate formulation of this well-known statement.

(4) The generalized rough measure $\bar{\mu}_H$ for a channel $\{x_-,x_+\}_{x^i}$ belonging to the group $G_{x^i} = G^{t^f}_{x^i t^i}$ is determined in terms of the rough measure of the group $G^{x^f}_{x^i}$,

$$\bar{\mu}_H(\{x_-,x_+\}_{x^i}) = \int_{x^f_-}^{x^f_+} \mu_H(\{x_-,x_+\}^{x^f}_{x^i})\, dx^f,$$

$$\times \bar{\mu}_H(G_{x^i}) = \int_{-\infty}^{\infty} \mu_H(G^{x^f}_{x^i})\, dx^f. \tag{3.28}$$

Hence we see that the action functional for the group $G^{t^f}_{x^i t^i}$ coincides with the action $S_H(\gamma)$ for the group $G^{x^f}_{x^i}$, so it is the same as the classical action functional. The method of Sec. I is working in the group $G^{t^f}_{x^i t^i}$, as well as in the group $G^{x^f}_{x^i}$, but the generalized invariant measure $d\bar{\sigma}_\kappa$ contains an extra factor dx^f,

$$d\bar{\sigma}_\kappa(\gamma) = \exp\left[-\frac{i}{\hbar} S_H(\gamma)\right] d\bar{\mu}_H(\gamma),$$

$$d\bar{\sigma}_\kappa(\gamma) = d\sigma_\kappa(\gamma)\, dx^f. \tag{3.29}$$

Equation (3.29) results from the fact that the group G_{x^i} is a direct integral of the groups $G^{x^f}_{x^i}$ over the x^f axis.

Similarly, the generalized rough measure μ_H on the channels $\{x_-,x_+\}^{x^f}$, which are elements of the group $G^{x^f} = G^{x^f t^f}_{t^i}$, is given by the integral of the rough measure for the channels $\{x_-,x_+\}^{x^f}_{x^i}$ over dx^i,

$$\mu_H(\{x_-,x_+\}^{x^f}) = \int_{x^i}^{x^f_+} \mu_m(\{x_-,x_+\}^{x^f}_{x^i})\, dx^i,$$

$$\mu_H(G^f_i) = \int_{-\infty}^{\infty} \mu_H(G^{x^f}_{x^i})\, dx^i, \tag{3.30}$$

and the generalized invariant measure $d\sigma_H(\gamma)$ depends on the same single parameter κ; it is given by

$$d\sigma_\kappa(\gamma) = \exp\left[-\frac{i}{\hbar} S_H(\gamma)\right] d\underline{\mu}_H(\gamma),$$

$$d\underline{\sigma}_{-\kappa}(\gamma) = d\sigma_\kappa(\gamma)\, dx^i. \tag{3.31}$$

The rule (3.20) for the change of variables at the parallel translation, and Eqs. (3.26), are valid also for the groups G_{x^i} and G^{x^f}. It is sufficient to take $d\bar{\mu}$ and $d\bar{\sigma}$ (respectively, $d\underline{\mu}$ and $d\underline{\sigma}$) instead of $d\mu$ and $d\sigma$.

(5) The action functional $S_H(\gamma)$ has been found here in the internal group-theoretical manner, using Eqs. (1.12)–(1.13) and being based upon the generalized rough measure μ_H, induced by the Schrödinger equation with the Hamiltonian H. We have shown that the result coincides with the conventional action functional for the paths belonging to the subgroup G', for which the classical action is determined. It has been shown as well that the difference of two action functional for the group elements belonging to the same coset with respect to the subgroup G' is finite.

Suppose now that the group-theoretical action functional is known, and we try to get the rough measure $d\mu_H(\gamma)$, and, consequently, the generalized invariant measure, and the path integrals for the groups $G^{x^f}_{x^i}$ and G_{x^i}. It follows from Eqs. (3.23) and (3.25) that

$$\frac{d\mu_H(\gamma)}{d\mu_{H_0}(\gamma_0)} = \exp\left[\frac{i}{\hbar} S_H(\gamma) - \frac{i}{\hbar} S_{H_0}(\gamma_0)\right], \quad \gamma - \gamma_0 \in G'. \tag{3.32}$$

This relation determines the measure $d\mu_H(\gamma)$ for various Hamiltonians H in terms of the action functional, up to a constant factor. The normalization is fixed by Eq. (3.13),

$$\int_{G^{x^f}_{x^i}} d\mu_{H_0}(\gamma) = [2\pi i\hbar(t^f - t^i)/m]^{-1/2} \exp\left[\frac{im(x^f - x^i)^2}{2\hbar(t^f - t^i)}\right]. \tag{3.33}$$

Let us rewrite the general formulas (1.17) and (1.18) for the particular case of the group $G^{x^f}_{x^i}$. The appropriate choice of the element g'_0 is the segment $\bar{\gamma}_0$ connecting the points $(0,t^i)$ and $(0,t^f)$, so

$$\sigma_\kappa\left[\{x_-,x_+\}^{x^f}_{x^i};\gamma^1,\bar{\gamma}_0\right] = \exp\left[-\frac{i}{\hbar} S_{H_0}(\bar{\gamma}_0)\right] \mu_{H_0}\left[\{x_-,x_+\}^{x^f}_{x^i} - \gamma'\right], \tag{3.34}$$

where γ' is an arbitrary path belonging to $G' \cap \{x_-,x_+\}^{x^f}_{x^i}$, $S_{H_0}(\bar{\gamma}_0) = F(t^f - t^i)$ is an arbitrary function present in $S_H(\gamma)$, and

$$\int \varphi(\gamma) \exp\left[\frac{i}{\hbar} S_H(\gamma)\right] d\sigma_\kappa(\gamma) = \lim \sum \varphi(\gamma) \exp\left[\frac{i}{\hbar} S_H(\gamma')\right] \sigma_\kappa\left[\{x_-,x_+\}^{x^f}_{x^i} - \gamma'\right]. \tag{3.35}$$

In particular,

$$K(x^f,t^f|H|x^i,t^i) = \lim \sum \exp\left[\frac{i}{\hbar} S_H(\gamma')\right] \sigma_\kappa\left[\{x_-,x_+\}^{x^f}_{x^i} - \gamma'\right]. \tag{3.36}$$

(6) The representation of the evolution kernel K in terms of the path integral enables one to verify in a simple and direct way the unitarity relation,

$$\int \psi^*(x,t)\psi(x,t)\, dx = \text{const}, \tag{3.37}$$

and the relations of the type

$$\int K^*(x_2,t_2,x'_1,t_1) K(x_2,t_2,x_1,t_1)\, dx_2 = \delta(x'_1 - x_1) \tag{3.38}$$

and

$$\int K^*(x_3,t_3,x_2,t_2)\, dx_3\, K(x_3,t_3,x_1,t_1) = K(x_2,t_2,x_1,t_1). \tag{3.39}$$

It is sufficient to use the reality of the Lagrangian, the property which is equivalent to the self-conjugate structure of the Hamiltonian operator that is used in the standard derivation of Eqs. (3.37)–(3.39).

If the time is inversed, the path γ connecting the point (x_1,t_1) to (x_2,t_2) goes to the path γ^{-1} of the inverse direction, and the action functional gets inverse sign, $S_H(\gamma^{-1}) = -S_H(\gamma)$, so according to Eq. (3.32),

$$\frac{d\mu_H(\gamma^{-1})}{d\mu_{H_0}(\gamma^{-1})} = \left[\frac{d\mu_H(\gamma)}{d\mu_{H_0}(\gamma)}\right]^* . \tag{3.40}$$

The normalization condition (3.33) is also replaced by the conjugate condition,

$$\int_{G^{x_1,t_1}_{x_2,t_2}} d\mu_{H_0}(\gamma^{-1}) = \left[\frac{2\pi i\hbar(t_1-t_2)}{m}\right]^{-1/2} \exp\left[\frac{im(x_1-x_2)^2}{2\hbar(t^1-t^2)}\right]. \tag{3.41}$$

This stems from the fact that the derivation of the analytical expression for the evolution kernel $K_0(x^f,t^f|x^i,t^i)$ is valid for $t^f \geqslant t^i$. From Eqs. (3.40) and (3.41) we get, so

$$K_1(x_1,t_1|H|x_2,t_2) = K^*(x_2,t_2|H|x_1,t_1), \quad t_1 \leqslant t_2, \tag{3.42}$$

and the wave functions $\psi(x,t)$ satisfy the relations

$$\psi(x_2,t_2) = \int \psi(x_1,t_1)K(x_2,t_2|H|x_1,t_1)\,dx_1, \qquad \psi(x_1,t_1) = \int \psi(x_2,t_2)K^*(x_2,t_2|H|x_1,t_1)\,dx_2. \tag{3.43}$$

Equation (3.38) is obtained from the general relation (3.43) with $\psi(x,t) = K(x,t|H|x_1',t_1)$. Substituting Eq. (3.38) into the identity

$$\int \psi^*(x,t)\psi(x,t)dx = \int \psi^*(x_1,t_1)K^*(x,t|x_1,t_1)\,dx_1 \times \psi(x_1',t_1)K(x,t|x_1',t_1)\,dx_1'\,dx,$$

we get the unitarity relation,

$$\int \psi^*(x,t)\psi(x,t)\,dx = \int \psi^*(x_1,t_1)\psi(x_1,t_1)\,dx_1 .$$

Because of Eq. (3.42), the identity (3.39) holds not only for $t_1 < t_2 < t_3$, but for any mutual positions of t_1,t_2,t_3.

The definition of the evolution kernel $K(x_2,t_2x_1,t_1)$, given in Ref. 2, implies that it vanish for $t_2 < t_1$. This is suitable sometimes, but such a definition makes it impossible to use the fact that the motion is reversible.

(7) In the case of a quadratic action functional $S_H(\gamma)$, where the Lagrangian is a second-order polynomial in $\dot{x}$ and x, and the Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - \frac{\hbar}{i}[\alpha(t)x + \beta(t)]\frac{\partial}{\partial x} + [a(t)x^2 + b(t)x + c(t)], \tag{3.44}$$

the evolution kernel $K(x^f,t^f|H|x^i,t^i)$ is calculated explicitly, as is well known. Let us consider the calculation by means of the path integral.

Let $\gamma = \gamma_{\rm cl} + \delta$ be the mapping of the group $G^{x^f}_{x^i}$ to the group $G^{x^f=0}_{x^i=0}$. The classical trajectories $\gamma_{\rm cl}$ are stationary curves of the action functional, so that

$$S_H(\gamma_{\rm cl}+\delta) = S_H(\gamma_{\rm cl}) + S_{H_1}(\delta), \qquad H_1 = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - \frac{\hbar}{i}\alpha(t)x\frac{\partial}{\partial x} + a(t)x^2 . \tag{3.45}$$

As the generalized invariant measure is independent of the lower-order terms in the Hamiltonian, we have

$$d\mu_H(\gamma) = \exp\left[\frac{i}{\hbar}S_H(\gamma_{\rm cl})\right]d\mu_{H_1}(\delta). \tag{3.46}$$

Integrating the integral of the lhs over the group $G^{x^f}_{x^i}$ and the rhs over the group $G^{x^f=0}_{x^i=0}$, we obtain

$$K(x^f,t^f|H|x^i,t^i) = K(0,t^f|H_1|0,t^i)\exp[(i/\hbar)S_H(\gamma_{\rm cl})] . \tag{3.47}$$

To calculate the time-dependent factor $K(0,t^f|H_1|0,t^i)$, let us calculate the integral over x_f,

$$J(t^f,t^i,x_1^i) = \int_{-\infty}^{\infty} K(x^f,t^f|H_1|x_1^i,t^i)\,dx^f = K(0,t^f|H_1|0,t^i) \times \int_{-\infty}^{\infty} \exp\left[\frac{i}{\hbar}S_H(\gamma_{\rm cl},x^i=x_1^i)\right]dx^f ; \tag{3.48}$$

one can choose x_1^i here in the most suitable way. As the kernel satisfies the Schrödinger equation,

$$\frac{d}{dt^f}J(t^f,t^i,x_1^i) = \frac{\hbar}{i2m}\int \frac{\partial^2}{\partial x^2}K(x,t^f|H_1|x_1^i,t^i)\,dx + \alpha(t^f)\int x\frac{\partial}{\partial x}K(x,t^f|H_1|x_1^i,t^i)\,dx - \frac{i}{\hbar}a(t^f)\int x^2K(x,t^f|H_1|x_1^i,t^i)\,dx . \tag{3.49}$$

The first term and the second term on the r.h.s. vanish, while the third term is reduced to the form $g(t^f)J(t^f,t^i,x_1^i)$ by means of the integration by parts, where the function $g(t^f)$ is determined by $S_H(\gamma_{\rm cl})$. Thus we have obtained the differential equation

$$\frac{d}{dt^f}J(t^f,t^i,x_1^i) = g(t^f)J(t^f,t^i,x_1^i), \tag{3.50}$$

and the corresponding initial condition is $J(t^i,t^i,x_1^i) = 1$, so

$$J(t^f,t^i,x_1^i) = \exp\left[\int_{t^i}^{t^f} g(t)\,dt\right], \tag{3.51}$$

$$K(0,t^f|H_1|0,t^i) = J(t^f,t^i,x_1^i)\left(\int \exp\left[\frac{i}{\hbar}S_H(\gamma_{\rm cl})x^i=x_1^i)\right]dx^f\right)^{-1}.$$

Example: Harmonical oscillator, the Lagrangian $L = \frac{1}{2}m(x^2-\omega^2x^2)$. Here we have $H = H_1$, and with the notation $T = t^f - t^i$, the result is

$$x_{\rm cl}(t) = [x^f\sin\omega(t-t^i) + x^i\sin(t^f-t)]/\sin\omega T,$$
$$S_H(\gamma_{\rm cl}) = \frac{m\omega}{2\sin\omega T}[\cos\omega T(x^{f^2}+x^{i^2}) - 2x^ix^f],$$
$$S_H(\gamma_{\rm cl})_{x_1^i=0} = \tfrac{1}{2}m\omega\cot(\omega T)x^{f^2}, \tag{3.52}$$
$$\int \exp\left[\frac{i}{\hbar}S_H(\gamma_{\rm cl})\right]dx^f = \left[\frac{2\pi i\hbar\sin\omega T}{m\omega\cos\omega T}\right]^{1/2}.$$

The potential is $V = \frac{1}{2}m\omega^2x^2$, and we have finally

$$\frac{d}{dt^f}J$$

$$= -\frac{i}{2\hbar}m\omega^2 K(0,t^f|0,t^i)\int x^2\exp\left[\frac{i}{\hbar}S(\gamma_{\rm cl})_{x_1^i=0}\right]dx$$

$$= \frac{i}{2\hbar}m\omega^2\cdot\frac{\hbar}{im\omega}\tan\omega T\cdot J(t^f,t^i,x_1^i=0)\,. \tag{3.53}$$

So

$$g(t)=\tfrac{1}{2}\omega\tan(t-t^i)\,,\quad J(t^f,t^i,x_1^i=0)=(\cos\omega T)^{-1/2}\,,$$

$$K(0,t^f|H|0,t^i)=\left[\frac{m\omega}{2\pi i\hbar\sin\omega T}\right]^{1/2}, \tag{3.54}$$

$$K_H(x^f,t^f|H|x^i,t^i)=\left[\frac{m\omega}{2\pi i\hbar\sin\omega T}\right]^{1/2}\exp\left[\frac{i}{\hbar}S_H(\gamma_{\rm cl})\right].$$

(8) Let ϵ and M_ϵ be two positive numbers, related by the formula

$$2\left|\frac{1}{2\pi i}\int_{M_\epsilon}^{\infty}e^{iy^2}\,dy\right|=\epsilon\,. \tag{3.55}$$

The domain $U(M_\epsilon)$ in the space of paths, where the evolution kernel $K_0(x^i,t^f|x^i,t^i)$ gets the dominating contribution, with the relative accuracy of ϵ, is determined by the inequalities

$$|x-x^i|\leqslant M_\epsilon[2\hbar(t-t^i)/m]^{1/2}\,,\quad t^i\leqslant t\leqslant\tfrac{1}{2}(t^i+t^f)\,,$$
$$|x-x^i|\leqslant M_\epsilon[2\hbar(t^f-t)/m]^{1/2}\,,\quad \tfrac{1}{2}(t^f+t^i)\leqslant t\leqslant t^f. \tag{3.56}$$

The boundaries of the domain are two parabolas

$$(x-x^i)^2=M_\epsilon^2 2\hbar(t-t^i)/m\,,$$
$$(x-x^i)^2=M_\epsilon^2 2\hbar(t^f-t)/m\,.$$

The partial transition amplitude which is due to this domain will be denoted by $K_0(U(M_\epsilon))$, and the domain itself will be called the lense. For any time moment τ in the interval (t^i,t^f), the cross section of the lense by the straight line $t=\tau$ plays the role of a "big box," and

$$K_0(x^i,t^f/x^i,t^i)=(1+r)K_0(U(M_\epsilon))\,,\quad |r|<\epsilon\,. \tag{3.57}$$

In other words, the paths along which the evolution kernel gets the dominating contribution, with the relative accuracy of ϵ, deviate from the classical trajectory of the particle at rest, $x(t)=x^i$, no farther than by $M_\epsilon\kappa$, where $\kappa=[\hbar(t^f-t^i)/m]^{1/2}$.

For the general case where the free particle is moving to $x^f\neq x^i$, the contribution from the trajectory with a constant velocity must be added, so the lense $U(M_\epsilon)$ is to be replaced by the deformed lense $U(\bar\gamma_i^f,M_\epsilon)$, where the segment $\bar\gamma$, connecting the points (x^f,t^f) and (x^i,t^i), is the axis of the lense, and

$$K_0(x^f,t^f|x^i,t^i)=(1+r)K_0(U(\bar\gamma_i^f,M_\epsilon))\,,\quad |r|<\epsilon\,. \tag{3.58}$$

Formally, Eq. (3.58) is obtained from (3.57) by means of the mapping

$$\gamma=\gamma_{\rm cl}+\delta\,,\quad \gamma_{\rm cl}=x_i+(x^f-x^i)(t-t^i)/(t^f-t^i)\,,$$

which was considered in subsection (2). Because of this mapping, $d\mu_{H_0}(\gamma)$ differs from $d\mu_{H_0}(\delta)$ by the same factor $\exp[im(x-x^i)^2/2\hbar(t^f-t^i)]$ as that relating $K_0(x^f,t^f|x^i,t^i)$ to $K_0(x^i,t^f|x^i,t^i)$. Thus for $\kappa\to 0$ the quantal motion of the free particle tends to the classical limit.

This conclusion is true also if the motion is influenced by a potential $V(x,t)$, if the potential is smooth enough. Let us use once more the transformation $\gamma-\gamma_{\rm cl}=\delta$, $\gamma_{\rm cl}=\bar x(t)$, $\delta=\eta(t)$. The result is

$$S_V(\gamma)=S_V(\gamma_{\rm cl})+S_{V_1}(\delta)\,,$$
$$V_1(\eta,t)=V[\bar x(t)+\eta,t]-V[\bar x,t]-\eta V'[\bar x(t),t]\,. \tag{3.59}$$

The measure is transformed as follows:

$$d\mu_V(\gamma)=\exp\left[\frac{i}{\hbar}S_V(\gamma_{\rm cl})\right]d\mu_{V_1}(\delta)\,,$$
$$d\mu_{V_1}(\delta)=\exp\left[-\frac{i}{\hbar}\int_\delta V_1(\eta,t)\,dt\right]d\mu_{H_0}(\delta),$$
$$K_V(x^f,t^f|x^i,t^i)=K_{V_0}(0,t^f|0,t^i)\exp\left[\frac{i}{\hbar}S_V(\gamma_{\rm cl})\right]. \tag{3.60}$$

If the parameter κ is small, the amplitude $K_0(0,t^f|0,t^i)$ is contributed mainly by the paths in the lense $U(M_\epsilon)$, the diameter of which is κM_ϵ, and neglecting the terms of order ϵ, one can discard the paths which go out of the lense. If the variation of the functional $\exp[(i/t)|V_1(\eta,t)\,dt]$ is not too large, the contribution to $K_V(0,t^f|0,t^i)$ from paths lying outside the lense $U(M_\epsilon)$ is negligible for $\epsilon\ll 1$.

Let us suppose that the potential is so smooth that, within the lense $U(M_\epsilon)$, V_1 can be replaced by three higher terms,

$$V_1(\eta,t)=\tfrac{1}{2}V''[\bar x(t),t]\eta^2+\tfrac{1}{6}V''[\bar x(t),t]\eta^3$$
$$+(1/4!)f(\eta,t)\eta^4\,, \tag{3.61}$$

where $f(\eta,t)$ is a bounded function; all the x derivatives are taken at the classical trajectory. So for the paths lying in the lense we have

$$\exp\left[-\frac{i}{\hbar}\int_\delta V_1(\eta,t)\,dt\right]$$
$$=1-\frac{i}{2\hbar}V''\eta^2T-\frac{i}{6\hbar}V'''\eta^3T$$
$$-\frac{i}{4!\hbar}f\hbar^4T-\frac{1}{8\hbar^2}(V''\eta^2)^2T^2\,, \tag{3.62}$$

where $V''\eta^2$, $V'''\eta^3$, and $f\eta^4$ are values taken at some intermediate points of the curve δ. The term with the odd power of η can be discarded as $d\mu(-\delta)=d\mu(\delta)$, so that such terms cancel in the sum

$$\exp\left[-\frac{i}{\hbar}\int_{-\delta}V_1(\eta,t)\,dt\right]d\mu_{H_0}(-\delta)$$
$$+\exp\left[-\frac{i}{\hbar}\int_\delta V_1(\eta,t)\,dt\right]d\mu_{H_0}(\delta)\,.$$

Within the lense $U(M_\epsilon)$ we have $\eta\sim\sqrt{\hbar T/M}$ by the order of magnitude, so that

$$\tfrac{1}{4}V''\eta^2T\sim V''T^2/m,\quad (1/\hbar)f\eta^4T\sim\hbar fT^3/m^2,$$
$$((1/\hbar)V''\eta^2T)^2\sim V''^2(T^2/m)^2\,. \tag{3.63}$$

and the only terms to be retained in Eq. (3.61) are

$$1-(i/2\hbar)V''T\eta^2-\tfrac{1}{8}(V'')^2T^2\eta^4/\hbar^2\,.$$

Thus it is sufficient to retain only the quadratic term $\frac{1}{2}V''[\bar{x},t]\eta^2$ in Eq. (3.62). Now the calculation of $K_{V_i}(0,t^f|0,t^i)$ is reduced to that for a quadratic Hamiltonian; this is the well-known WKB approximation. The relative error in $K_{V_i}(0,t^f|0,t^i)$ is given by $V^{IV}(\hbar T/m)(T^2/m)$ by the order of magnitude. We must assume that the potential $V(x,t)$ is smooth in the neighborhood of the classical trajectory, and increases slower than x^2 for $x\to\pm\infty$.

The dominating role of the classical trajectory for $\kappa\to 0$ is equivalent to the statement that the measure $d\mu_{H_V}$ contains a δ function on the group $G_{x^i}^{x^f}$. Suppose $\varphi(\gamma)$ is a functional which is integrable with the rough measure $d\mu_H(\gamma,\kappa)$ at $\kappa=\kappa_1$, then $\varphi(\gamma)$ is integrable for all $\kappa<\kappa_1$, and

$$\lim_{\kappa\to 0}\frac{\int\varphi(\gamma)d\mu_{H_V}(\gamma,\kappa)}{\int d\mu_{H_V}(\gamma,\kappa)}=\lim_{\kappa\to 0}\frac{\int\varphi(\gamma)\exp[(i/\hbar)S(\gamma)]d\sigma}{K_V(x^f,t^f|x^i,t^i)}=\varphi(\gamma_{\rm cl})\,. \tag{3.64}$$

(9) Let us consider the system evolution for a time interval $\Delta t^k=t^{k+1}-t^k$; this interval tends to zero as $N\to\infty$, for $\Delta t=(t^f-t^i)/N$. It follows from (3.60) and (3.62) that for a finite M the lense contribution is

$$K_V(U(\bar{\gamma}_k^{k+1},M))=\exp[(i/\hbar)S_V(\gamma_{\rm cl})]\times K_0(x^{k+1},t^{k+1}|x^k,t^k)\,. \tag{3.65}$$

This formula has the accuracy of N^{-2}, but we still need an estimate for the contribution to the transition amplitude from the paths lying outside the lense, so a more accurate calculation is expedient.

The relative error r in Eq. (3.58) is of the order of $\int_M^\infty e^{iy^2}dy$, where $y=x/\kappa$, so that

$$K_0(x^{k+1},t^{k+1}|x^k,t^k)=K_0(U(\bar{\gamma}_k^{k+1},M))+O\left(\int_M^\infty e^{iy^2}dy\right)\times K_0(x^{k+1},t^{k+1}|x^k,t^k)\,.$$

Now we calculate the average of this equality over the interval $(M,M+\beta)$ where β is determined by the condition $(M+\beta)^2=M^2+2\pi$. The result is

$$K_0(x^{k+1},t^{k+1}|x^k,t^k)=K_0(U^{(1)}(\bar{\gamma}_k^{k+1},M))+O(M^{-3}e^{iM^2})\times K_0(x^{k+1},t^{k+1}|x^k,t^k)\,, \tag{3.66}$$

where $U^{(1)}$ indicates that we have averaged over the lenses. Repeating the averaging l times, we obtain

$$K_0(x^{k+1},t^{k+1}|x^k,t^k)=K_0(U^{(l)}(\bar{\gamma}_k^{k+1},M))+O(M^{-(2l+1)}e^{iM^2})K_0(x^{k+1},t^{k+1}|x^k,t^k)\,. \tag{3.67}$$

The averaging over the lenses, i.e., the transition from U to $U^{(1)}$, means that the boundary conditions on the lense go to zero by the law $M^{-(2l-1)}\exp(iM^2)$. Suppose $M=N^\alpha$; then

$$K_0(x^{k+1},t^{k+1}|x^k,t^k)=[1+O(N^{-(2l-1)\alpha})]\int d\mu_{H_0}(\gamma)\times U(\bar{\gamma}_k^{k+1},M)\,. \tag{3.68}$$

If the potential increase slower than x^2 at $x\to\pm\infty$, we get from (3.67) and (3.68) the following estimate:

$$K_V(x^{k+1},t^{k+1}|x^k,t^k)=[1+O(N^{-(2l-1)\alpha})]\exp\left[-\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}}V\,dt\right]\times U(\bar{\gamma}_k^{k+1},N^\alpha)\int\exp\left[-\frac{i}{\hbar}\int_\gamma V\,dt+\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}}V\,dt\right]d\mu_{H_0}(\gamma)\,. \tag{3.69}$$

The diameter of the deformed lense $U(\bar{\gamma}_k^{k+1},N^\alpha)$ is equal to $\kappa N^{-1/2+\alpha}$ and vanishes for $\alpha<\frac{1}{2}$. If the force is bounded within U, then

$$\exp\left[-\frac{i}{\hbar}\int_\gamma V\,dt+\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}}V\,dt\right]=1+O(\kappa N^{-(3/2-\alpha)})\,. \tag{3.70}$$

If we take $\alpha=3/21$, then

$$K_V(x^{k+1},t^{k+1}|x^k,t^k)=[1+O(N^{-3/2(1-1/l)})]\times\exp\left[-\frac{i}{\hbar}\int_{\bar{\gamma}_k^{k+1}}V\,dt\right]K_0(x^{k+1},t^{k+1}|x^k,t^k)\,. \tag{3.71}$$

The number of the averagings l must be higher than 3; it can be arbitrarily large.

The relation (3.71) for any time interval Δt^k has been derived under the condition that the derivative $\partial V/\partial x$, i.e., the classical force, is limited uniformly throughout the infinite band $t^i<t<t^f$, $-\infty<x<\infty$, but this condition can be softened.

The integral $\int_{\bar{\gamma}_k^{k+1}}V\,dt$ in Eq. (3.71) can be replaced by the integral along any curve γ, belonging to the lense $U(\bar{\gamma}_k^{k+1},N^\alpha)$.

(10) It was proved in subsection (8) that the motion in a smooth potential is quasiclassical only in the case where κ is small; that is to say, it is quasiclassical if the observation time is small as compared with $m/\hbar$; otherwise, quantum effects are substantial. This fact suggests a remark.

Poincaré[8] proved a theorem, according to which any finite classical motion, i.e., motion bounded to a region in the phase space of coordinates and velocities, is almost-periodical. One concludes, hence, for instance, that a gas that occupied a part of the volume of a closed vessel and that was freed subsequently will return after a certain time τ to an arbitrarily close vicinity of the initial position, in a clear contradiction to the second principle of thermodynamics. This paradox was discussed by a number of people, but its solution is beyond the framework of the classical mechanics. Actually, the "almost-period" τ is too large, so the quantal effects are necessarily substantial, and the Poincaré theorem is irrelevant.

IV. PATH INTEGRAL AS THE LIMIT OF MULTIPLE INTEGRAL ON A TIME LATTICE

(1) The fundamental idea to represent the probability amplitude for a system transition as an integral of $\exp[(i/\hbar)S(\gamma)]$ over alternative system histories was realized by Feynman originally in the form of a multiple integral on a time lattice, which tends to a limit as the lattice step goes to zero.

In the particular case where the system considered in the one-dimensional motion of mass-m particle in a force potential $V(x,t)$, the multiple integral representing the evolution kernel is

$$J_V^N = \frac{1}{A}\int\cdots\int \exp\left\{\frac{i}{\hbar}\sum_{k=0}^{N-1}\left[\frac{1}{2}\frac{m(x^{k+1}-x^k)^2}{\Delta t} - \frac{1}{2}(V^k+V^{k+1})\right]\right\}\frac{dx^1}{A}\cdots\frac{dx^{N-1}}{A}, \tag{4.1}$$

$$A=\left(\frac{2\pi i\hbar\Delta t}{m}\right)^{1/2}, \quad t^k = t^0 + k\Delta t, \quad \Delta t = (t^f - t^i)/N,$$

$$t^0 = t^i, \quad t^N = t^f, \quad x^0 = x^i, \quad x^N = x^f,$$

$$V^k = V(x^k,t^k), \quad N\to\infty,$$

and the sum in the exponent was interpreted as the action functional along the polygonal line with vertices at the points (x^k,t^k) (see in Ref. 2).

Accurate proofs that the limit of J_V^N for $N-\infty$ is $K_V(x^f,t^f|x^i,t^i)$ under certain conditions on the potential were given by a number of authors. The proofs are very complicated and special, and they cannot be extended to more difficult problems arising in quantum field theories.

In fact, in Feynman's definition of the path integral as the limit of the expression in Eq. (4.1), there is an essential deviation from the original idea to sum up the functional $\exp[(i/\hbar)S(\gamma)]$ over all possible paths, as only polygonal trajectories are taken into account. One would argue that both the definitions coincide in the limit of $N\to\infty$, because any continuous curve is approximated by polygonal lines, but this statement is applicable to the integral involving the potential only, since $\int_\gamma V(x,t)\,dt$ is a continuous functional, while the variation of $\int_\gamma \frac{1}{2}m\dot{x}^2(t)\,dt$ is infinite within any channel, arbitrarily narrow.

In the case of a free particle, the definition of the path integral as the limit of the integral on the time lattice lacks its meaning to an extent, as in view of the integration over the intermediate states the evolution kernel is $K_0(x^f,t^f|x^i,t^i) = J_0^N$ for any N, and the problem of finding the evolution kernel for a small time interval, that is, $K_0(x^{k+1},t^{k+1}|x^k,t^k)$, is no easier than calculating the evolution kernel for arbitrary times. The known expression for the free particle evolution kernel,

$$K_0(x^f,t^f|x^i,t^i) = \left[\frac{2\pi i\hbar(t^f-t^i)}{m}\right]^{-1/2}\exp\left[\frac{im(x^f-x^i)^2}{2\hbar(t^f-t^i)}\right], \tag{4.2}$$

was derived in Sec. III as a result of the parallel translation in the group of paths, $x(t)\to x(t)+\bar{\gamma}_i^f$. In fact, $(i/\hbar)S_{H_0}(\bar{\gamma}_i^f)$ is present in the exponential, and the derivation of Eq. (4.2) indicates that the preexponential factor $[2\pi i\hbar(t^f-t^i)/m]^{-1/2} = K_0(x^i,t^f|x^i,t^i)$ is influenced by the integration over alternative trajectories; and the reason why the integral $\int_{\bar{\gamma}}\dot{x}^2(t)\,dt$ appears in (4.2) is not at all that the contribution from $\bar{\gamma}_i^f$ to the path integral is of a particular importance. Somewhat ironically, the form of (4.2) is in contradiction to the statement that the contribution to the kernel from all smooth curves, connecting the initial and final points, is zero.

There is also the following reason to try to understand what the real meaning of the path integral is and why Feynman's approach leads to correct results. We have mentioned numerous times that the path integral is contributed by the curves along which Δx has the order of magnitude of $(\Delta t)^{1/2}$, while on such paths $S(\gamma)=\infty$, and $\exp[(i/\hbar)S(\gamma)]$ acquires no definite value. (The situation is even more paradoxical in the case of the diffusion equation where we have $\exp[(i/\hbar)S(\gamma)]=0$ on such paths.) Hence we conclude that the path integral cannot be considered as a sum of $\exp[(i/\hbar)S(\gamma)]$ over trajectories.

It was shown in Sec. II that the Schrödinger equation with a Hamiltonian H induces a generalized rough measure μ_H in the group of paths. Furthermore, in Sec. III we have shown that an action functional $S_H(\gamma)$ is determined by purely group-theoretical arguments, and it coincides with the classical action. Given the rough measure and the action functional, the method of Sec. I leads to a generalized invariant measure and enables one to write down the kernel K_V in the spectacular form,

$$K_V(x^f,t^f|x^i,t^i) = \int \exp\left[\frac{i}{\hbar}S_H(\gamma)\right]d\sigma(\gamma). \tag{4.3}$$

The exact meaning of the symbols used is revealed in Eqs. (1.17) and (1.18).

(2) The fundamental Feynman's principle is quite reasonable, but it cannot be interpreted literally; it needs some adequate mathematical concepts and formulations. The purpose of Sec. II and III was to produce such a mathematical framework. In the present section we are going to elucidate the meaning of the time lattice and present a general proof that the integral on the lattice is converging to the path integral. The proof we give here is valid for the most complicated cases.

We turn to the Dirac method[9] representing the evolution kernel by means of the multiple integral over the intermediate states,

$$K_V(x^f,t^f|x^i,t^i) = \int K_V(x^f|x^{N-1})\,dx^{N-1}\cdots dx^1 K_V(x^1|x^i),$$
$$K_V(x^{k+1}|x^k) \equiv K_V(x^{k+1},t^{k+1}|x^k,t^k). \tag{4.4}$$

The problem we meet in calculation of the kernel K_V for an x dependent potential V,

$$K_V(x^f,t^f|x^i,t^i) = \int d\mu_V(\gamma),$$
$$d\mu_V(\gamma) = \exp\left[-\frac{i}{\hbar}\int_\gamma V(x,t)\,dt\right]d\mu_{H_0}(\gamma), \tag{4.5}$$

is that only the integral of $d\mu_{H_0}(\gamma)$ over the whole group $G_{x^i}^{x^f}$ is known, namely, $\mu_{H_0}(G_{x^i}^{x^f}) = K_0(x^f,t^f|x^i,t^i)$, but not the elementary measure $d\mu_{H_0}(\gamma)$. Equation (4.4) with the time step $\Delta t = (t^f-t^i)/N$, i.e., the introduction of the time lattice, is useful just to avoid this difficulty, not to calculate approximately $S_V(\bar{\gamma}_k^{k+1})$ as one might imagine.

Actually, according to subsection (9) of Sec. III, for $N\gg 1$ the kernel $K_V(x^{k+1},t^{k+1}|x^k,t^k)$ is contributed predominantly by the paths in the lense $U(\bar{\gamma}_k^{k+1}, N^\epsilon)$, the diameter of which is $N^{-1/2+\epsilon}$, and the variation of the functional $\int_\gamma V\,dt$ in this lense is of the order of $N^{-3/2+\epsilon}$. Respective-

ly, it follows from Eq. (3.71) that

$$K_V(x^{k+1},t^{k+1}|x^k,t^k) = \left[\exp\left(-\frac{i}{\hbar}\int_{\bar\gamma_k^{k+1}} V\,dt\right) + r(x^{k+1},t^{k+1}|x^k,t^k)\right] K_0(x^{k+1}|x^k), \tag{4.6}$$

where $r(x^{k+1}|x^k)\sim N^{-3/2+\epsilon}$, and its phase varies slowly; $\epsilon>0$ is arbitrarily small. Using Eq. (4.6), we are in position to prove that the multiple integral on the time lattice,

$$K_V^N(x^f,t^f|x^i,t^i) = \int\cdots\int \exp\left(-\frac{i}{\hbar}\int_{\bar\gamma_{N-1}^N} V\,dt\right) \times K_0(x^f|x^{N-1})\,dx^{N-1}\cdots dx^1 \times \exp\left(-\frac{i}{\hbar}\int_{\bar\gamma_0^1} V\,dt\right) K_0(x^1|x^i), \tag{4.7}$$

converges to the evolution kernel $K_V(x^f,t^f|x^i,t^i)$ as $N\to\infty$, and the convergence rate is $N^{-1/2+\epsilon}$.

Let us put the expression (4.6) into the identity (4.4). At the first step we have

$$K_V(x^f|x^i) = \int K_V(x^f|x^1)\,dx^1 \exp\left(-\frac{i}{\hbar}\int_{\bar\gamma_0^1} V\,dt\right) K_0(x^1|x^i) + \int K_V(x^f|x^1)\,dx^1\, r(x^1|x^i) K_0(x^1|x^i).$$

The second term on the rhs has the order of magnitude of $N^{-3/2+\epsilon}$. After $(N-1)$ steps the result we arrive at is

$$K_V(x^f t^f|x^i t^i) = K_V^N(x^f t^f|x^i,t^i) + \sum_{l=1}^{N-1}\int K_V(x^f|x^l)dx^l\, r(x^l|x^{l-1})\,dx^{l-1} \times \exp\left[-\frac{i}{\hbar}\int_{\bar\gamma_{l-1}^l} V\,dt\right] \times K_0(x^l|x^{l-1}) K_V^{l-1}(x^{l-1}|x^i). \tag{4.8}$$

The second term on the rhs has the order of magnitude of $N\cdot N^{-3/2+\epsilon} = N^{-1/2+\epsilon}$. The above statement is proved under the assumption that the potential derivative modulus $|\partial V/\partial x|$ is bounded uniformly in the infinite band $t^i<t<t^f$, $-\infty<x<\infty$.

(3) For every segment $\bar\gamma_k^{k+1}$ we have

$$\exp\left[-\frac{i}{\hbar}\int_{\bar\gamma_k^{k+1}} V\,dt\right] K_0(x^{k+1}|x^k) = \exp\left[\frac{i}{\hbar} S_{H_V}(\bar\gamma_k^{k+1})\right] K_0(x^k t^{k+1}|x^k t^k),$$

so the Nth approximation can be written as

$$K_V^N(x^f,t^f|x^i,t^i) = \int \exp\left[\frac{i}{\hbar}\delta_{H_V}(\bar\gamma_0^N)\right] d\sigma^{(N)}(\bar\gamma_0^N), \tag{4.9}$$

where $\bar\gamma_0^N$ are polygonal lines with the vertices at the points (x^k,t^k), and

$$d\sigma^{(N)}(\bar\gamma_0^N) = K_0(0,t^f|0,t^{N-1}) \prod_{k=0}^{N-2} K_0(0,t^{k+1}|0,t^k)\,dx^{k+1}. \tag{4.10}$$

Comparison of K_V^N and J_V^N with the expression (4.4) for K_V shows that $\exp[-(i/\hbar)\int_{\bar\gamma_k^{k+1}} V\,dt\,]K_0(x^{k+1}|x^k)$ is an approximation of $K_V(x^{k+1}|x^k)$, the relative accuracy of which is $N^{-3/2+\epsilon}$ for any magnitude of the difference $x^{k+1}-x^k$, while $\exp[-(i/\hbar)\frac{1}{2}(V^k+V^{k+1})\Delta t\,]K_0(x^{k+1}|x^k)$ does not approximate $K_V(x^{k+1}|x^k)$ for finite Δx, though it has the same order of magnitude.

Thus we have seen that the integrals K_V^N have $K_V(x^f|x^i)$ as the limit, because the path integral for $K_V(x^f|x^i)$ is the integral of the functional $\exp[-(i/\hbar)\int_\gamma V\,dt\,]$ over the rough measure μ_{H_0} induced by the free-particle Schrödinger equation on the group $G_{x^i}^{x^f}$; it is not a normalized sum of $\exp[(i/\hbar)S_{H_V}(\gamma)]$ over all possible paths.

Actually, the situation here is similar to that for the usual integral of a continuous function $f(x)$ on a segment $[a,b]$, where one can choose any points ξ_i in the integral sums $\Sigma f(\xi_i)\,\Delta x_i$. Likewise, one can integrate the potential $V(x,t)$ in the expression (4.8) for K_V^N not over the straight segments $\bar\gamma_k^{k+1}$, or segments of classical trajectories, but over any curves within the lenses $U(\bar\gamma_k^{k+1}, N^\epsilon)$. The only difference as compared with the integral sums for the usual integral is that instead of the measure Δx on the axis one should take the generalized rough measure $d\mu_{H_0}(\gamma)$.

We have assumed that all the time intervals Δt^k were identical, but it was only for the sake of brevity; it was sufficient in fact to assume that the time intervals are N^{-1} by the order of magnitude.

Note that an analogous limiting relation exists for any two Hamiltonians of the general form (2.1) with the same mass m, H_1 and H_2,

$$K(x^f,t^f|H_2|x^i,t^i) = \lim_{N\to\infty}\int\cdots\int \exp\left[\frac{i}{\hbar}\int_{\bar\gamma_{N-1}} (\omega_2^L-\omega_1^L)\,K(x^f|H_1|x^{N-1})dx^{N-1}\cdots dx^1 \exp\left[\frac{i}{\hbar}\int_{\bar\gamma_0}(\omega_2^L-\omega_1^L)K(x^1|H_1|x^i)\right.\right. \tag{4.11}$$

It can be also written down in the form

$$K(x^f,t^f|H_2|x^i,t^i) = \lim_{N\to\infty}\int \exp\left[\frac{i}{\hbar} S_{H_2}(\bar\gamma_0^N)\,d\sigma(\bar\gamma_0^N)\right]. \tag{4.12}$$

Note that the equality

$$K_V(x^f,t^f|x^i,t^i) = \lim_{N\to\infty} K_V^N(x^f,t^f|x^i,t^i)$$

can be used for an immediate derivation of the perturbative series. To this end one has to replace the factors $\exp(-(i/\hbar)\int_{\gamma_k^{k+1}} V\,dt)$ on the rhs by $(1-(i/\hbar)\int_{\gamma_k^{k+1}} V\,dt)$, and to use the expansion

$$\prod_{k=0}^{N-1}\left(1-\frac{i}{\hbar}\int_{\bar\gamma_k^{k+1}} V\,dt\right) = 1+\left(-\frac{i}{\hbar}\right)\sum_{k=0}^{N-1}\int_{\gamma_k^{k+1}} V\,dt + \left(-\frac{i}{\hbar}\right)^2 \sum_{1\leqslant k_1<k_2\leqslant N-1}\int_{\bar\gamma_{k_2}^{k_2+1}} V\,dt\int_{\bar\gamma_{k_1}^{k_1+1}} V\,dt + \cdots .$$

(4) The rough measure $\mu_V(U(\bar\gamma_k^{k+1}, N^\epsilon))$ is the amplitude of the probability density for transition of the particle from the point (x^k,t^k) to another point (k^{k+1},t^{k+1}) along paths lying within the lense $U(\bar\gamma_k^{k+1}, N^\epsilon)$, so

$$\mu_V[U(\bar\gamma_0^N;\bar x_0;\Delta x^{1,\ldots,N-1};N^\epsilon)] = \int_{x_0^1-\Delta x^1}^{x_0^1+\Delta x^1}\cdots\int_{x_0^{N-1}-\Delta x^{N-1}}^{x_0^{N-1}+\Delta x^{N-1}}\prod_{k=0}^{N-2}\mu_V(U(\bar\gamma_k^{k+1},N^\epsilon))\,dx^{k+1}, \tag{4.13}$$

where $U(\bar\gamma_0^N;\bar x_0;\Delta x^{1,\ldots,N-1};N^\epsilon)$ is the domain covered by the lenses $U(\bar\gamma_k^{k+1},N^\epsilon)$ as the points x^{k+1} run on the intervals $x_0^{k+1}-\Delta x^{k+1}, x_0^{k+1}+\Delta x^{k+1})$. According to subsection (9) of Sec. III, the measure $\mu_V[U(\bar\gamma_k^{k+1},N^\epsilon)]$ can be replaced by $\exp[-(i/\hbar)\int_{\bar\gamma_k^{k+1}} V\,dt]\times K_0(x^{k+1}|x^k)$ with the relative error of $N^{-3/2+\epsilon}$, so that it follows from Eq. (4.13) that

$$\mu_V[U(\bar\gamma_0^N;\bar x_0;\Delta x^{1,\ldots,N-1};N^\epsilon)] = \int_{x_0^1-\Delta x^1}^{x_0^1+\Delta x^1}\cdots\int_{x_0^{N-1}-\Delta x^{N-1}}^{x_0^{N-1}+\Delta x^{N-1}}\exp\left[\frac{i}{\hbar}S_V(\bar\gamma_0^N)\right]d\sigma^{(N-1)}(\bar\gamma_0^N), \tag{4.14}$$

where the relative error is of the order of $N^{-1/2+\epsilon}$. The same accuracy has the equality

$$\mu_V[U(\bar\gamma_0^N;\bar x_0;\Delta x^{1,\ldots,N-1};N^\epsilon)] = \int_{x_0^1+\Delta x^1}^{x_0^1+\Delta x^1}\cdots\int_{x_0^{N-1}+\Delta x^{N-1}}^{x_0^{N-1}+\Delta x^{N-1}} K_V(x^f|x^{N-1})\,dx^{N-1}\cdots dx^1\,K_V(x^1|x^0) \tag{4.15}$$

since $\mu_V[U(\bar\gamma,N^\epsilon)] = K_V(x^{k+1}|x^k)(1+O(N^{-3/2+\epsilon}))$. The expression on the rhs is the rough measure of the beam $B_{x^i}^{x^f}(\bar x_0-\bar\Delta x,\bar x_0+\bar\Delta x)$, and with the relative accuracy of $N^{-1/2+\epsilon}$ we have

$$\mu_V[U(\bar\gamma_0^N;\bar x_0;\bar\Delta x;N^\epsilon)] = \mu_V[B_{x^i}^{x^f}(\bar x_0-\bar\Delta x;\bar x_0+\bar\Delta x)]. \tag{4.16}$$

Equations (4.14) – (4.16) are true for μ_H and S_H with the general Hamiltonian of the type (2.1).

The diameter of the channel $U(\bar\gamma_0^N;\bar x_0;\bar\Delta x;N^\epsilon)$ is different from $\max\Delta x^{k+1}$ by a quantity of the order of $N^{-1/2+\epsilon}$ only, and it goes to zero as $N\to\infty$ and $\Delta x^{k+1}\to 0$, so that Eq. (4.16) means that the measure of beams with arbitrarily narrow openings Δx can be approximated with channels having arbitrarily small diameters. This, in turn, enables one to define the integral over the generalized rough measure μ_{H_0} or μ_H, based upon the evident divisions of the groups G_i^f and $G_{x^i}^{x^f}$ into complete systems of independent beams of arbitrarily narrow openings.

(5) The equality $K_V(x^f|x^i) = \lim K_V^N(x^f|x^i)$ means that the integral of the functional $\exp[-(i/\hbar)\int_\gamma V(x,t)\,dt]$ over the generalized rough measure μ_{H_0} on the group $G_{x^i}^{x^f}$ is equal to the limit of the integral on the time lattice as the lattice spacing goes to zero, $\Delta t\to 0$. This is true for any functional $\varphi(\gamma)$ satisfying the conditions which were used in the proof, namely:

(i) The functional $\varphi(\gamma)$ is continuous and the rule of integration over the intermediate states for $\Phi(x^ft^f|x^it^i) = \int_{G_{x^i}^{x^f}}\varphi(\gamma)d\mu_{H_0}(\gamma)$ is true, that is,

$$\Phi(x^ft^f|x^it^i) = \int\Phi(x^f|x^{N-1})\,dx^{N-1}\cdots dx^1\Phi(x^1|x^i)$$

$$\Phi(x^{k+1}|x^k) = \Phi(x^{k+1}t^{k+1}|x^kt^k). \tag{4.4'}$$

(ii) For any two paths $\gamma_1,\gamma_2\in G_{x^kt^k}^{x^{k+1}t^{k+1}}$ we have

$$|\varphi(\gamma_2)-\varphi(\gamma_1)| < \text{const}\,\rho(\gamma_1,\gamma_2)\Delta t^k,$$

where the constant is independent of k.

(iii) in the infinite band $t^i\leqslant t\leqslant t^f$, $-\infty<x<\infty$, we have

$$\arg\varphi(\gamma) < \text{const}\,\rho(\gamma,\bar\gamma_{x^i=0}^{x^f=0}).$$

For a functional belonging to this class

$$\int\varphi(\gamma)\,d\mu_{H_0}(\gamma) = \lim_{N\to\infty}\int\varphi(\bar\gamma_{N-1}^N)K_0(x^N|x^{N-1})\,dx^{N-1}\cdots dx^1\,\varphi(\bar\gamma_0^1)K(x^1|x^0), \tag{4.17}$$

or

$$\int\varphi(\gamma)\,d\mu_{H_0}(\gamma) = \lim_{N\to\infty}\int\varphi(\bar\gamma_0^N)\exp\left[\frac{i}{\hbar}S_{H_0}(\bar\gamma_0^N)\right]d\sigma(\bar\gamma_0^N). \tag{4.18}$$

Hence we conclude that for such functionals the following limit exists in sense of the theory of distributions:

$$\lim_{N\to\infty} d\sigma(\bar{\gamma}_0^N;\kappa) = d\sigma(\gamma;\kappa), \quad \kappa = [\hbar(t^f - t^i)/m]^{1/2}. \tag{4.19}$$

This relation makes it possible to get the rule for the change of variables in the functional integral (4.17) in a simple and rigorous way for quite general transformations of the variables.

Let γ_x^t and γ_y^t be some curves $x = x(\tau)$ and $y = y(\tau)$, $t^i \leqslant \tau \leqslant t \leqslant t^f$, and the transformation T is defined as

$$T\colon\ \gamma_y^t = \gamma_x^t + A(\gamma_x^t;t), \quad T^{-1}\colon\ \gamma_x^t = \gamma_y^t + B(\gamma_y^t;t), \tag{4.20}$$

where the functionals A and B depend on γ_x^t and γ_y^t, and are functions of t. It is supposed that the functional derivative of $B(\gamma_y^t;t)$ exists,

$$\delta B(\gamma_y^t;t) = \int_{t^i}^{t} F(\gamma_y^t;t;\tau)\,\delta y(\tau)\,d\tau, \tag{4.21}$$

and is continuous in its functional argument γ_y^t.

The transformations T and T^{-1} induce the corresponding transformations on the time lattice. The above discussion suggests that in order to get the transformations T and T^{-1} as the limits of the induced transformations on the time lattice, it is sufficient that the relative error for every t^k, as the transformations are considered on the lattice, must be a quantity of the order of $\bar{o}(N^{-1})$. The same condition must hold also for the case where the time steps Δt^k are not identical. The only rule for the approximate calculation of the integral in Eq. (4.21), satisfying these requirements, is the trapezoid rule, so we get from Eqs. (4.20) and (4.21)

$$dx^{k+1} = \left[1 + \tfrac{1}{2} F(\gamma_y^{t^{k+1}};t^{k+1};t^{k+1})\Delta t^k\right] dy^{k+1} + \cdots. \tag{4.22}$$

So the transformation matrix is triangular. The transformation Jacobian is

$$\prod_{k=0}^{N-2} \left[1 + \tfrac{1}{2} F(\gamma_y^{t^{k+1}};t^{k+1};t^{k+1})\Delta t^k\right] \to \exp\left[\tfrac{1}{2}\int_{t^i}^{t^f} F(\gamma_y^t;t;t)\,dt\right], \tag{4.23}$$

and

$$\int \varphi(\gamma_x)\exp\left[\frac{i}{\hbar} S_{H_0}(\gamma_x)\right] d\sigma(\gamma_x) = \int \varphi(T^{-1}\gamma_y^{t^f})\exp\left[\frac{i}{\hbar} S_{H_0}(T^{-1}\gamma_y^{t^f})\right]\exp\left[\frac{1}{2}\int_{t^i}^{t^f} F(\gamma_y^t;t;t)\,dt\right] d\sigma(\gamma_y). \tag{4.24}$$

Next we have to relate the action functionals. The result is

$$S_{H_0}(\gamma_x^{t^f}) = S_{H_0}(\gamma_y^{t^f}) + m\int_{\gamma_y} \frac{d}{dt} B(\gamma_y^t,t)\,dy + \frac{1}{2} m\int \left[\frac{d}{dt} B(\gamma_y^t;t)\right]^2 dt. \tag{4.25}$$

Putting the expression for S_{H_0} into Eq. (4.24), we obtain

$$\int_{G_{x^i}^{x^f}} \varphi(\gamma_x^{t^f})\,d\mu_{H_0}(\gamma_x^{t^f}) = \int_{TG_{x^i}^{x^f}} \varphi(T^{-1}\gamma_y^{t^f})\cdot d\mu_{H_0}(\gamma_y^{t^f})$$
$$\times \exp\left\{\frac{im}{\hbar}\int_{\gamma_y}\frac{d}{dt} B(\gamma_y^t;t)\,dy + \frac{im}{2\hbar}\int\left[\frac{d}{dt} B(\gamma_y^t;t)\right]^2 dt + \frac{1}{2}\int F(\gamma_y^t;t,t)\,dt\right\}. \tag{4.26}$$

The formula is applicable not only to the integrals over the groups in the whole, $G_{x^i}^{x^f}$ and $TG_{x^i}^{x^f}$, but also to the integrals over any channels $U = \{x_-, x_+\}_{x^i}^{x^f}$ and TU.

For the Volterra transformation,

$$T^{-1}\colon\ \gamma_x^t = \gamma_y^t + \int_{t^i}^{t} C(t,\tau)\gamma_y^\tau\,d\tau, \tag{4.27}$$

the functional derivative $F = C(t,\tau)$ is independent of γ_y, and the factor $\exp\left[\frac{1}{2}\int_{t^i}^{t^f} C(t,t)\,dt\right]$ can be taken out of the integral in (4.26).

V. CONCLUSION

The specific features of the present approach to the functional integrals as integrals on a locally noncompact group with a generalized measure has been revealed here for the path integrals of quantum mechanics. The next interesting subject to be studied is a realization of this concept in quantum field theory and, in particular, in the theory of gauge fields. The fact that the functional integral on a locally noncompact group G,

$$\int \varphi(\gamma)\,d\mu(\gamma),$$

is reduced to Feynman's integral,

$$\int \varphi(\gamma)\exp\left[\frac{i}{\hbar} S(\gamma)\right] d\sigma(\gamma),$$

only under the condition that an action functional, which is determined upon a purely internal group-theoretical basis, exists for the group G, suggests an idea that a group of this

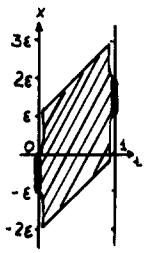


FIG. 1. Step #0.

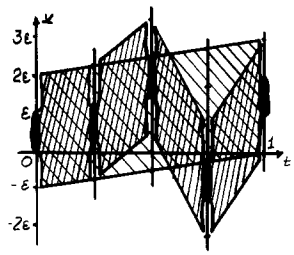


FIG. 3. Construction of the step #2 channel from the step #1 channel.

type exists and the group-theoretical functional coincides with the classical action in all the cases where the Feynman integral appears.

ACKNOWLEDGMENTS

It is a pleasure to express my gratitude to M. S. Marinov and B. V. Yusin for very useful discussions. I am also obliged to B. V. Yusin for a proof of the theorem on the subdivision of the space of paths into arbitrarily narrow channels, which is presented in the Appendix (which has been written by him).

APPENDIX: PROOF OF THE THEOREM ON CHANNEL SUBDIVISION

Theorem: The space G_i^f of all continuous paths on the line with free start and finish, as well as any channel (see Sec. II) in this space, can be subdivided into a countable family of independent channels of arbitrarily small diameters. (The subdivision is modulo the channel boundaries.)

For the sake of convenience, the paths (functions) will be represented here by their graphs, and the channels, by strips on the (x,t) plane, lying between the graphs of the boundary functions.

Note that if the entire space G_i^f is subdivided into independent channels U_α, then any other channel V is also subdivided into the channels $V \cap U_\alpha$. Thus it is sufficient to subdivide only the space G_i^f itself.

We shall put $t^i = 0$, $t^f = 1$.

Construction of the channels: We present an explicit construction of the channels of diameter 3ϵ covering the space G_i^f. The idea is to approximate the functions by their values at dyadic rational points $s/2^n$, s integer. The notation we use is $L_n = \{0,1/2^n,2/2^n,...,1\}$. At step #0 we approximate the functions with their restrictions to $L_0 = \{0,1\}$. Looking at the plane (x,t), we subdivide each of the lines $t = 0$ and $t = 1$ into segments of the length ϵ, starting from $x = 0$, and choose one of them on the first line, and one on the second line. With this at hand, we construct a channel, as it is shown in Fig. 1; its width is 3ϵ everywhere, except at $t = 0$ and $t = 1$, where it is ϵ and the boundary functions have breaks.

For any function in G_i^f we may find a channel among those containing the function graph in its strip at the points $t = 0$ and $t = 1$. However, there are functions which are not contained in these channels. At step #1 we take $L_1 = \{0,\frac{1}{2},1\}$. Given a function, its graph intersects the lines $t = 0$ and $t = 1$ at some points, and we already have a channel covering these points. If the point where the graph intersects the line $t = \frac{1}{2}$ is covered by this channel, we have nothing to do; otherwise, we construct a family of new channels. This is done by subdividing the line $t = \frac{1}{2}$ into the segments of length ϵ starting from the intersection of it with the boundary lines of the channel (see Fig. 2). Choosing any of them lying beyond the old channel, we construct a new one as is shown in Fig. 2. Its width is 3ϵ everywhere, except at $t = 0,\frac{1}{2},1$, where it is ϵ and the boundary functions have breaks.

Now for any function we can find a channel whose strip covers at least three points of its graph, namely, the intercepts with the lines $t = 0,\frac{1}{2},1$. We proceed further in a similar way, constructing at step $\#r$ the channels which cover (together with those constructed previously) the intercepts of the function graph with the union of the lines $T = s/2^n$, $s = 0,1,2,...,2^n$. To this end, we take a channel which was constructed at one of the preceding steps, say at step $\#k, 0 \leqslant k \leqslant n$, and cross it with all the lines $t = s/2^n$, $s = 0,1,2,..,2^n$, subdivide every line into segments of length ϵ, starting from the intercepts with the boundary lines of the channel, and then we choose a segment on every line in such a way that we take a segment lying *inside* the channel on the lines with even s (if $k = n - 1$ we have no choice at all; otherwise, we can choose among at most three segments on every line; see in Figs. 3 and 4). In the case of odd s we can choose among all the segments, but make sure that at least one of the chosen segments does not lie inside the old channel. The channel is constructed in the same way as before (see

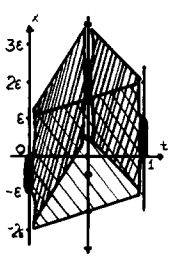


FIG. 2. Step #1.

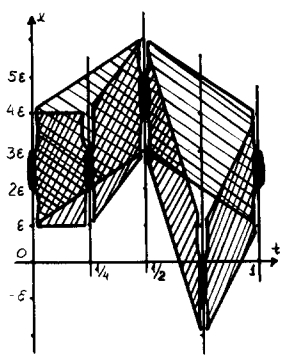


FIG. 4. Construction of the step #2 channel from the step #0 channel.

Figs. 3 and 4); its width is 3ϵ everywhere, except at $t = s/2^n$, where its width is ϵ and the boundary functions have breaks.

Details of the proof that this construction satisfies the conditions of the theorem are left up to the reader.

[1]N. N. Meiman, "Functional integrals as integrals on locally noncompact groups with a generalized measure," Report to the International Session (1977) of Moscow Seminar on Collective Phenomena (M. A. Azbel, Chairman); "Proof of convergence of integral on a time lattice to the path integral," Report to the International Session (1981) of Moscow Seminar on Collective Phenomena.

[2]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[3]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 255 (1979).

[4]M. S. Marinov, Phys. Rep. **60**, 1 (1980).

[5]F. A. Berezin, Usp. Fiz. Nauk **132**, 497 (1980).

[6]The second term in the Hamiltonian, as we write it down, is due to the electromagnetic vector potential. Respectively, U contains, besides the usual scalar potential, a term proportional to A^2, and its imaginary part is ∇A.

[7]R. H. Cameron and W. T. Martin, Trans. Am. Math. Soc. **58**, 184 (1945).

[8]H. Poincaré, Acta Math. **13**, 1 (1890).

[9]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1958).

Construction of the tensor product for the lattices of properties of physical entities

Dirk Aerts
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

(Received 10 February 1983; accepted for publication 9 September 1983)

We construct the tensor product for the property lattices of two entities. We give a physical interpretation for this tensor product. We show that the tensor product is never an orthocomplemented lattice if both entities are nontrivial. It also never satisfies the covering law. This is the reason why this tensor product does not exist in quantum mechanics and does not exist in quantum logic. We analyze the relation of the tensor product with the usual description of two entities in quantum mechanics. At the end, we give a mathematical way of constructing this tensor product.

PACS numbers: 03.65.Bz, 03.65.Fd, 02.10. + w, 02.40. + m

1. INTRODUCTION

Suppose we want to study a phenomenon by means of its properties. Then we can never study the collection of all "possible" properties of the phenomenon. What we are interested in is to consider a well-defined set of properties of the phenomenon. The "thing" described by this set of properties, and which if one wants can be considered as an idealization of the phenomenon, we have called an "entity."[1-3]

In principle, different entities can correspond to the same phenomenon, depending on what properties of the phenomenon we consider. Physical theories represent the entities and describe in this way the phenomenon. Therefore, this idealization is intrinsically connected to the choice of trying to understand reality by doing physics. Sometimes there are requirements, imposed by the physicist, which seem to imply an idealization at first sight. If one analyzes the situation more profoundly, one however sees that only when these requirements are attributed to the phenomenon, they imply perhaps an idealization. If one attributes the same requirements to an entity corresponding to the phenomenon no idealization is implied. This is the case for the requirement of "separation." Separated phenomena exist perhaps only in an idealized sense, while separated entities exist without any further idealization.[2]

We already studied, to some extent, the collection of properties and states of an entity in relation to the existing physical theories, classical mechanics, and quantum mechanics.[1-3]

This work is in a certain sense a continuation of work of Piron, that can be found in Refs. 4 and 5. In this paper we will construct the tensor product of property lattices. We will show that it is the solution of a universal problem and we will give a physical interpretation of this tensor product.

Let us consider the following situation: we have two entities S_1 and S_2 and the entity S consisting of these two entities. First of all we have to remark that this does not represent a unique situation. There are in principle many ways in which two entities S_1 and S_2 can form an entity S. These many ways correspond to different interactions that are possible between S_1 and S_2. In any of these situations we shall, however, say that the entity S is a "refinement" of both entities S_1 and S_2.

Suppose S_1 and S_2 are described by property lattices $\mathscr{L}_1$ and $\mathscr{L}_2$ and S is described by a property lattice $\mathscr{L}$. The case where $\mathscr{L}$ is the tensor product of $\mathscr{L}_1$ and $\mathscr{L}_2$ has to correspond to the situation where S is the "smallest" refinement of S_1 and S_2. Indeed if S' is an entity that is an arbitrary refinement of S_1 and S_2, then S' must also be a refinement of S. This explains why the tensor product can be used for the description of all the entities that are refinements of S_1 and S_2.

Let us consider the situation in classical mechanics. If S_1, S_2, and S are classical entities described in state spaces Γ_1, Γ_2, and Γ, then $\Gamma = \Gamma_1 \times \Gamma_2$, and this does not depend on the kind of interaction between S_1 and S_2.

In quantum mechanics the situation is not so simple. If S_1, S_2, and S are described in Hilbert spaces $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}$, then $\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2$. It is possible to show that $\mathbf{H}_1 \otimes \mathbf{H}_2$ cannot describe all possible ways in which two quantum entities S_1 and S_2 can form an entity S. It cannot describe for example the entity S if S_1 and S_2 are two separated entities.[1,2] Because in quantum mechanics there is no other way to describe two entities, this product $\mathbf{H}_1 \otimes \mathbf{H}_2$ is also used to describe two separated entities. This wrong description gives rise to the well-known paradoxes of the theory.[6] The entity S that is correctly described by $\mathbf{H}_1 \otimes \mathbf{H}_2$ consists always of two entities S_1 and S_2 that are not separated in a very specific way. Hence the entity S is in this case a very specific refinement of S_1 and S_2 and certainly not the smallest one.

A classical entity is described in a state space Γ. The relations between different classical entities are expressed by functions between the state spaces. Hence we can say that the category corresponding to classical entities is the category of which the objects are sets (the state spaces) and the morphisms are functions. In this category of sets and functions, $\Gamma_1 \times \Gamma_2$ represents the tensor product of Γ_1 and Γ_2.

A quantum entity is described in a Hilbert space $\mathbf{H}$. The relations between different quantum entities are expressed by linear or antilinear maps between their Hilbert spaces. Hence we can say that the category corresponding to quantum entities is the category of which the objects are Hilbert spaces and the morphisms are linear and antilinear maps. In the category of Hilbert spaces and linear maps, $\mathbf{H}_1 \otimes \mathbf{H}_2$ rep-

 0022-2488/84/051434-08$02.50

resents the tensor product of $\mathbf{H}_1$ and $\mathbf{H}_2$.

In the category of Hilbert spaces and linear and antilinear maps there does not exist a tensor product for two Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$. The reason is that $\mathbf{H}_1 \otimes \mathbf{H}_2$ and $\mathbf{H}_1 \otimes \mathbf{H}_2^*$ are not isomorphic.[7,8] Both categories, the one corresponding to classical entities, and the one corresponding to quantum entities, should be contained in the general category corresponding to entities in general. One of the conclusions of this paper is that while the category corresponding to classical entities is a "good" subcategory, the category corresponding to quantum entities is not. This is the reason why in this category there does not exist a tensor product. Quantum logic is a theory that has both classical mechanics and quantum mechanics as a special case. The objects of the category are orthocomplemented weakly modular posets or lattices, usually equipped with a probability measure. Sometimes the lattice is supposed to be σ-complete or complete. We shall also show here that there does not exist a tensor product. The axiom that is at the origin of this fact is the axiom that introduces an orthocomplementation of the lattice.

Hence also the category corresponding to quantum logic is not a good subcategory, and therefore, there does not exist a tensor product for quantum logics. In Refs. 1 and 2 we show that quantum mechanics and also quantum logic cannot describe separated entities, and we give a description of separated entities. The axiom that is at the origin of this fact is the axiom that makes the lattice weakly modular. The entity S that consists of the two separated entities S_1 and S_2 is certainly a refinement of both entities S_1 and S_2. For some time we thought that S would perhaps be the smallest refinement of S_1 and S_2. In this case the product constructed in Refs. 1 and 2, that we called the "separated product," would be the tensor product. It is mathematically very easy to see that this separated product is not a solution of the universal problem. This was very puzzling. Indeed, what could physically be a more simple situation than the situation where S_1 and S_2 are separated entities. But after some time it became clear that there does exist a more simple situation for two entities S_1 and S_2. Indeed if two entities S_1 and S_2 are separated, there is some interaction between them, the interaction that keeps the entities separated. This interaction will give rise to new properties for the entity S consisting of S_1 and S_2. This can be seen very well in the construction made in Refs. 1 and 2. A situation which is more simple is the situation where we do not know the interaction between S_1 and S_2. The entity S consisting of two entities S_1 and S_2, such that we do not know anything about the interaction between S_1 and S_2, will allow us to construct the tensor product. We think that this is the structure one ought to use to describe two entities.

If the two entities are classical, described in state spaces Γ_1 and Γ_2, this tensor product is the Cartesian product $\Gamma_1 \times \Gamma_2$ of the state spaces. If the two entities are nonclassical we find a completely new structure for the tensor product which is different from $\mathbf{H}_1 \otimes \mathbf{H}_2$ in the case where S_1 is described in the Hilbert space $\mathbf{H}_1$, and S_2 is described in the Hilbert space $\mathbf{H}_2$, and which is also different from the separated product constructed in Refs. 1 and 2.

The reason why this tensor product could be found neither in quantum mechanics, nor in quantum logic is that the property lattice describing S will never be orthocomplemented. The problem of the existence of a tensor product for quantum logic is an old one.[9] It has been investigated also in Refs. 7, 10, and 11.

2. CONSTRUCTION OF THE TENSOR PRODUCT

In this paper we will use the formalism as it is exposed in Ref. 2. This formalism is in a certain sense a quantum logic formalism, where the properties of the entity are not defined mathematically as some elements of a lattice, but physically as equivalence classes of questions. A question is a proposal of a test that can be performed on the phenomenon. If the test gives us the expected outcome, we will say that the answer of the test is "yes." If the test does not give us the expected outcome, we will say that the answer of the test is "no."

On the set of questions Q two operations are defined. An operation which makes correspond to every question α its inverse question $\alpha^{\sim} \cdot \alpha^{\sim}$ corresponds to the same test where yes and no are interchanged. Another operation which makes correspond to every family α_i of questions a product question $\pi_i \cdot \alpha_i \cdot \pi_i \cdot \alpha_i$ is the new test that consists of choosing an arbitrary question from the family α_i and performing the corresponding test. Q is supposed to be closed for these two operations. A question is said to be *true* iff when we should decide to perform the experiment the answer "yes" would come out with certainty. We have also a relation on the set of questions. If we have the situation that whenever a question α is true then also the question β is true, we denote $\alpha < \beta$, and we say "α is stronger than β." This defines a preorder relation on Q.

A property is represented by the equivalence class of questions testing this property. A property a is actual iff there is a question $\alpha \in a$ that is true. The preorder relation on the set of questions introduces a partial order relation on the set of properties. Namely $a < b$ iff whenever a is actual then also b is actual. It is possible to prove that the set of properties, denoted by $\mathcal{L}$, is a complete lattice (Ref. 2, Sec. 3.6).

The state is represented by the minimal element of the set of actual properties of the entity. The set of states Σ of the entity is full, which means that if a and b are properties such that for every state $p < a$ we have $p < b$ then $a < b$. Σ is also equipped with a natural orthogonality relation. Two states p and q are orthogonal iff there is a question α such that if the entity is in the state p, α is true and if the entity is in the state q, $\alpha^{\sim}$ is true. The collection of questions that are never true we will denote by O. It is easy to see that $O \in \mathcal{L}$. A trivial question is a question that is always true. The property represented by the equivalence class of all trivial questions will be denoted by I. Clearly for every property a we have $O < a < I$. A generating set of questions is a set of questions G such that $Q = \{\pi_i \alpha_i | \alpha_i \in G\}$. If $\mathcal{L}$ is the property lattice of an entity S, then we will say that $a \in \mathcal{L}$ is an atom of this lattice iff whenever $O < b < a$ we have $b = O$ or $b = a$.

1. The set of questions

We shall consider the following situation. We have two entities S_1 and S_2 and we suppose that every question of S_1 is still defined when S_2 is there and every question of S_2 is still defined when S_1 is there. The entity consisting of S_1 and S_2 in this situation, without knowing anything more about the interaction between S_1 and S_2, we will call S. If Q_1 is the set of questions of S_1 and Q_2 the set of questions of S_2, we can construct the set Q of questions of S. Clearly $Q_1 \subset Q$ and $Q_2 \subset Q$. The only new questions that are possible are the questions of the form $\alpha_1 \cdot \alpha_2$, where $\alpha_1 \in Q_1$ and $\alpha_2 \in Q_2$.

Hence,

$$Q = \{\alpha_1, \alpha_2, \alpha_1 \cdot \alpha_2 | \alpha_1 \in Q_1, \alpha_2 \in Q_2\},$$

and

$\alpha_1 \cdot \alpha_2$ is true iff α_1 is true *and* α_2 is true.

Since as explained in Ref. 2 the structure of the property lattice and the set of states are completely determined if we know the set of questions, we can now just derive these structures.

2. The property lattice

Suppose $\mathscr{L}_1$ and $\mathscr{L}_2$ are the property lattices of S_1 and S_2 and $\mathscr{L}$ is the property lattice of S. If $\alpha_1 \in Q_1$ tests $a_1 \in \mathscr{L}_1$ and $\alpha_2 \in Q_2$ tests $a_2 \in \mathscr{L}_2$, the property tested by $\alpha_1 \cdot \alpha_2$ will be denoted $a_1 \wedge a_2$.

Theorem 1: $\mathscr{L} = \{a_1 \wedge a_2 | a_1 \in \mathscr{L}_1, a_2 \in \mathscr{L}_2\}$ and $a_1 \wedge a_2$ is actual iff a_1 is actual and a_2 is actual.

For the minimal element O of $\mathscr{L}$ we have

$$O = O_1 \wedge a_1 = a_1 \wedge O_2$$

and for the maximal element I of $\mathscr{L}$ we have

$$I = I_1 \wedge I_2. \quad \blacksquare$$

3. The set of states

Suppose ϵ_1 and ϵ_2 are the collection of actual properties of S_1 and S_2 and ϵ is the collection of actual properties of S and suppose Σ_1 and Σ_2 are the set of states of S_1 and S_2 and Σ is the set of states of S.

Theorem 2: $\epsilon = \{a_1 \wedge a_2 | a_1 \in \epsilon_1, a_2 \in \epsilon_2\}$.

For every $p \in \Sigma$ there is a $p_1 \in \Sigma_1$ and $p_1 \in \Sigma_2$ such that $p = p_1 \wedge p_2$. Hence $\Sigma = \{p_1 \wedge p_2 | p_1 \in \Sigma_1 \text{ and } p_2 \in \Sigma_2\}$.

We have

$$p_1 \wedge p_2 < a_1 \wedge a_2 \quad \text{iff } p_1 < a_1 \text{ and } p_2 < a_2. \quad \blacksquare$$

4. The partial order relation

Theorem 3: If $a_1 \wedge a_2, b_1 \wedge b_2 \in \mathscr{L}$ then

$$a_1 \wedge a_2 < b_1 \wedge b_2 \quad \text{iff } a_1 < b_1 \text{ and } a_2 < b_2. \text{ or } a_1 = O_1 \text{ or } a_2 = O_2.$$

Proof: Suppose $a_1 \not< b_1$, then there is a state $p_1 < a_1$ such that $p_1 \not< b_1$. Consider a state $p_2 < a_2$. If S is in the state $p_1 \wedge p_2$, then $a_1 \wedge a_1$ is actual but $b_1 \wedge b_2$ is not actual. Hence $a_1 \wedge a_2 \not< b_1 \wedge b_2$. $\blacksquare$

5. The orthogonality relation

Theorem 4: If $p_1 \wedge p_2, q_1 \wedge q_2 \in \Sigma$, then

$$p_1 \wedge p_2 \perp q_1 \wedge q_2 \quad \text{iff } p_1 \perp q_1 \text{ or } p_2 \perp q_2.$$

Proof: If $p_1 \wedge p_2 \perp q_1 \wedge q_2$, then there exists a question α such that $p_1 \wedge p_2 < \alpha$ and $q_1 \wedge q_2 < \alpha^\sim$. Now α is of the form α_1, α_2, or $\beta_1 \cdot \beta_2$. If α is of the form α_1 then $p_1 < \alpha_1$ and $q_1 < \alpha_1^\sim$ such that $p_1 \perp q_1$. If α is of the form α_2 then $p_2 < \alpha_2$ and $q_2 < \alpha_2^\sim$ If α is of the form $\beta_1 \cdot \beta_2$ then $p_1 < \beta_1$ and $q_1 < \beta_1^\sim$ and $p_2 < \beta_2$ and $q_2 < \beta_2^\sim$. $\blacksquare$

6. The primitive questions

In Refs. 1 and 2 we defined the concept of primitive questions. The idea behind this definition is that the questions used to describe properly an entity should be only the products of such primitive questions. Hence the primitive questions should generate the property lattice of the entity. This demand is exposed in Axiom 1 of Refs. 1 and 2.

Let us repeat this definition of a primitive question.

Definition 1: If α is a question testing a property a such that $\alpha^\sim$ tests the property b, then α is a primitive question iff whenever the entity is in a state orthogonal to a, then $\alpha^\sim$ is true and whenever the entity is in a state orthogonal to b, then α is true.

We can show now that the set of primitive questions of the entity S is just the union of the set of primitive questions of S_1 and the set of primitive questions of S_2. This shows in a certain sense already that the property lattice of S will be the "smallest" one containing the property lattice of S_1 and the property lattice of S_2.

To be able to show this we have to remark on some properties of the property lattice. Suppose that we consider an entity. Any trivial question of the entity is also a primitive question. Hence the set of questions contains at least one primitive question. We can also remark that the set of states of an entity contains at least one state, since any trivial question defines the property I which is always actual.

Theorem 5: If P is the set of primitive questions of S, then $P = P_1 \cup P_2$, where P_1 is the set of primitive questions of S_1, and P_2 is the set of primitive questions of S_2.

Proof: If $\alpha_1 \in P_1$ then $\alpha_1 \in P$ and if $\alpha_2 \in P_2$ then $\alpha_2 \in P$. Suppose that $\alpha_1 \cdot \alpha_2$ is a primitive question. If $\alpha_2 \notin O_2$ we can consider a state p_2 of S_2 such that $p_2 < \alpha_2$. If $\alpha_1 \in a_1$ and $p_1 \perp a_1$ then $p_1 \wedge p_2 \perp a_1 \wedge a_2$ where $\alpha_2 \in a_2$. But then $p_1 \wedge p_2 < \alpha_1^\sim \cdot \alpha_2^\sim$. As a consequence $p_2 < \alpha_2^\sim$ which is a contradiction. If $\alpha_2 \in O_2$, then $\alpha_2^\sim \notin O_2$ if α_2 is a primitive question, so that we can make analogous reasoning for $\alpha_2^\sim$. $\blacksquare$

7. The axioms

As we remarked already in the Introduction, we will show that the axioms that are usually supposed to be satisfied in quantum logic are not satisfied if this tensor product exists. We shall use the axioms as defined in Refs. 1 and 2.

Let us state the axioms as they can be found in Refs. 1 and 2.

Axiom 1: If S is an entity, then the primitive questions of S form a generating set of questions for the property lattice.

Axiom 2: If S is an entity and p is a state of S then there exists a question that is true iff S is in a state orthogonal to p.

If Axiom 1 and Axiom 2 are satisfied for the property lattice $\mathscr{L}$ of the entity S, then $\mathscr{L}$ is orthocomplemented as we have shown in theorem 7 of Ref. 2.

Axiom 3: The states of the entity are represented by atoms of the property lattice. As a consequence of this axiom, the property lattice is an atomic lattice.

Axiom 4: (weak modularity): If $\mathscr{L}$ is the property lattice of an entity and $a, b \in \mathscr{L}$ such that $a < b$, then it is possible to find a property $c \in \mathscr{L}$ such that $c \perp a$ and $a \vee c = b$. This axiom is rather of a technical nature and it is not easy to see what it means physically. It makes the property lattice weakly modular which must be the case since all property lattices of quantum entities are weakly modular.

Axiom 5: (covering law): If $\mathscr{L}$ is the property lattice of an entity S and $a \in \mathscr{L}$ and p is a state of S such that $a \wedge p = O$, then $a \vee p$ covers a. This axiom is again of a technical nature but satisfied in quantum mechanics. Axioms 1, 2, 3, 4 and 5 are satisfied in quantum mechanics and also in classical mechanics. In a certain sense also the inverse is true. These five axioms force our formalism equivalent to quantum mechanics with superselection rules. Hence we could say that they are the axioms of quantum mechanics.

We shall show now that Axiom 2 and Axiom 5 are never satisfied in the tensor product lattice of two property lattices. And this is the reason why this tensor product does not exist in quantum mechanics.

Theorem 6: The property lattice $\mathscr{L}$ of S satisfies axiom 1 iff the property lattice $\mathscr{L}_1$ of S_1 and the property lattice $\mathscr{L}_2$ and S_2 satisfy Axiom 1.

Proof: Follows immediately from Theorem 5. So one of the axioms that leads to an orthocomplementation does not cause any problems. Also the atomicity does not cause any problems.

Theorem 7: The property lattice $\mathscr{L}$ of S satisfies Axiom 3 iff the property lattice $\mathscr{L}_1$ of S_1 and the property lattice $\mathscr{L}_2$ and S_2 satisfy Axiom 3.

Proof: Suppose $\mathscr{L}$ satisfies Axiom 3. Consider a state p_1 of S_1 and $O_1 \neq a_1 < p_1$. If p_2 is a state of S_2 then $O \neq a_1 \wedge p_2 < p_1 \wedge p_2$. Hence $p_1 \wedge p_2 < a_1 \wedge p_2$ which shows that $p_1 < a_1$. So $\mathscr{L}_1$ satisfies Axiom 3. In a similar way we show that $\mathscr{L}_2$ satisfies Axiom 3.

Suppose $\mathscr{L}_1$ and $\mathscr{L}_2$ satisfy Axiom 3. Consider $a_1 \wedge a_2 \neq O$ and $a_1 \wedge a_2 < p_1 \wedge p_2$. Then $O_1 \neq a_1 < p_1$ and $O_2 \neq a_2 < p_2$. Hence $p_1 = a_1$ and $p_2 = a_2$. ■

The second axiom that leads to an orthocomplementation will, however, only be satisfied in some trivial physically noninteresting cases.

Definition 8: We will say that an entity S is "trivial" iff for every question α of S we have $\alpha \in O$ or $\alpha^{\sim} \in O$. Hence a trivial entity is an entity that has no orthogonal states. The entities described by classical physics and the entities described by quantum physics are not trivial.

Theorem 8: If the property lattice $\mathscr{L}$ of the entity S satisfies Axiom 2, then S_1 or S_2 is a trivial entity.

Proof: Suppose that neither S_1 nor S_2 is a trivial entity. Then there exists two orthogonal states $p_1 \perp q_1$ of S_1 and two orthogonal states $p_2 \perp q_2$ of S_2. Consider the state $p_1 \wedge p_2$ of S. If Axiom 2 is satisfied, there exists a property $a_1 \wedge a_2$ such that $a_1 \wedge a_1$ is actual iff the entity S is in a state orthogonal to $p_1 \wedge p_2$. Suppose r_2 is an arbitrary state of S_2. Then $q_1 \wedge r_2 \perp p_1 \wedge p_2$. Hence $q_1 \wedge r_2 < a_1 \wedge a_2$. As a consequence $r_2 < a_2$. This shows that $a_2 = I_2$. Suppose that r_1 is an arbitrary state of S_1. Then $r_1 \wedge q_1 \perp p_1 \wedge p_2$. Hence $r_1 \wedge q_1 < a_1 \wedge a_2$. So $r_1 < a_1$. As a consequence $a_1 = I_1$. Hence $a_1 \wedge a_1 = I$. But then $p_1 \wedge p_2 \perp p_1 \wedge p_2$ which is a contradiction. ■

This theorem shows that the lattice $\mathscr{L}$ will never be orthocomplemented if it is not the product of the lattice of a trivial entity with another one. It also shows that orthocomplementation is not a good axiom for the category of lattices of properties of an entity.

Let us compute now the supremum and the infimum of elements of $\mathscr{L}$. By doing this we solve the "word problem" of lattice theory for the lattice $\mathscr{L}$.

Theorem 9: If $a_1^i \wedge a_2^i \in \mathscr{L}$ such that $a_1^i \wedge a_2^i \neq O$ for every i then

$$\bigwedge_i (a_1^i \wedge a_2^i) = (\bigwedge_i a_1^i) \wedge (\bigwedge_i a_2^i),$$

$$\bigvee_i (a_1^i \wedge a_2^i) = (\bigvee_i a_1^i) \wedge (\bigvee_i a_2^i).$$

Proof: Suppose $a_1^i \wedge a_2^i < c_1 \wedge c_2$ for every i. Then $a_1^i < c_1$ and $a_2^i < c_2$ for every i. Hence $\bigvee_i a_1^i < c_1$ and $\bigvee_i a_2^i < c_2$. As a consequence $(\bigvee_i a_1^i) \wedge (\bigvee_i a_2^i) < c_1 \wedge c_2$.

On the other hand $a_1^i \wedge a_2^i < (\bigvee_i a_1^i) \wedge (\bigvee_i a_2^i)$ for every j. ■

We can ask whether Axiom 4 that introduces the property of "weak modularity" is satisfied. Weak modularity is, however, usually only defined in an orthocomplemented lattice. In this case in the definition the orthocomplementation is needed. It is, however, possible to give a definition where the orthocomplementation is not required. This is the way we introduced weak modularity in Ref. 2 in Axiom 4. Introduced in this way weak modularity does not cause any trouble. The question, however, remains whether weak modularity without orthocomplementation is an interesting property for the lattice.

Theorem 10: The property lattice $\mathscr{L}$ of S satisfies Axiom 4 (weak modularity) iff the property lattice $\mathscr{L}_1$ of S_1 and the property lattice $\mathscr{L}_2$ of S_2 satisfy Axiom 4.

Proof: Suppose $\mathscr{L}_1$ and $\mathscr{L}_2$ satisfy Axiom 4. Consider $a_1 \wedge a_2 \neq O$ and $a_1 \wedge a_2 < b_1 \wedge b_2$. Then $a_1 < b_1$ and $a_2 < b_2$. Hence there exists $c_1 \perp a_1$ and $c_2 \perp a_2$ such that $a_1 \vee c_1 = b_1$ and $a_2 \vee c_2 = b_2$. As a consequence $(a_1 \wedge a_2) \vee (c_1 \wedge c_2) = b_1 \wedge b_2$ and $c_1 \wedge c_2 \perp a_1 \wedge a_2$. Suppose now that $\mathscr{L}$ satisfies Axiom 4. Consider $a_1 \neq O$ and $a_1 < b_1$. Then $a_1 \wedge I_1 < b_1 \wedge I_2$. So there exists $c_1 \wedge c_2 \perp a_1 \wedge I_2$ and $(a_1 \wedge I_1) \vee (c_1 \wedge c_2) = (a_1 \vee c_1) \wedge I_2 = b_1 \wedge I_2$. But then $c_1 \perp a_1$ and $a_1 \vee c_1 = b_1$. Hence $\mathscr{L}_1$ satisfies Axiom 4. In an analogous way we show that also $\mathscr{L}_2$ satisfies Axiom 4. It makes more sense to verify whether Axiom 5 (the covering law) is satisfied in $\mathscr{L}$, because the orthocomplementation is not really connected to this property. We can very easily see that Axiom 5 is in fact never satisfied in $\mathscr{L}$.

Theorem 11: If the property lattice $\mathscr{L}$ of S satisfies Axiom 5 then one of the two entities S_1 or S_2 has a property lattice containing only two elements O and I. $\mathscr{L}$ is then isomorphic to the property lattice of the other entity.

Proof: Suppose Axiom 5 to be satisfied for $\mathscr{L}$ and suppose that for example the entity S_1 has at least two different

states $p_1 \neq q_1$. Consider two states p_2 and q_2 of the entity S_2. Then $p_1 \wedge p_2$ and $q_1 \wedge q_2$ are two different states of S. Since they are atoms $(p_1 \wedge p_2) \vee (q_1 \wedge q_2)$ covers $p_1 \wedge p_2$. Now

$$p_1 \wedge p_2 < (p_1 \vee q_1) \wedge p_2 < (p_1 \vee q_1) \wedge (p_2 \vee q_2) = (p_1 \wedge p_2) \vee (q_1 \wedge q_2).$$

As a consequence

$$p_1 \vee q_1 = p_1 \quad \text{or} \quad p_2 = p_2 \vee q_2.$$

Since $p_1 \neq q_1$ we cannot have $p_1 = p_1 \vee q_1$. Hence $p_2 = p_2 \vee q_2$. But then $p_2 = q_2$. Since p_2 and q_2 were arbitrary states of S_2, it follows that S_2 has only one state. As a consequence $\mathscr{L}_2 = \{O_2, I_2\}$ and $\mathscr{L} \cong \mathscr{L}_1$. ■

From this theorem also follows that the covering law is not a good property for the category of property lattices. This covering law was also not satisfied for the property lattice of two separated entities.[1,2]

It is the covering law which allows the lattice to be represented by the set of closed subspaces of a vector space. So we cannot hope that our tensor product lattice will be representable by a vector space structure. So again, as in the case of two separated entities, the superposition principle will not be satisfied for the entity S even when S_1 and S_2 are described by quantum mechanics in Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$. Since the property lattice $\mathscr{L}$ will be a tensor product of $\mathscr{P}(\mathbf{H}_1)$ and $\mathscr{P}(\mathbf{H}_2)$ it will, however, be possible to find a morphism that maps $\mathscr{L}$ into $\mathscr{P}(\mathbf{H}_1 \otimes \mathbf{H}_2)$. We shall now show that $\mathscr{L}$ is indeed the tensor product. To do this we have to introduce the concept of morphism.

3. MORPHISMS AND A SOLUTION OF THE UNIVERSAL PROBLEM

1. Morphisms and bimorphisms

The concept that we will use to express relations between different entities is the concept of morphism. We will have the situation in mind where S and S' are entities corresponding to the same phenomenon. As explained in Ref. 12, this is in fact the general situation. In Ref. 12 it is also shown that when $Q \subset Q'$, and the entity S' is a refinement of the entity S, then the morphisms are introduced in the following way:

(1) If p represents the state of S and so is the minimum of all actual properties of S, and p' represents the state of S' and so is the minimum of all actual properties of S' we shall write $p = g(p')$. In this way we define a map

$$g: \Sigma' \to \Sigma.$$

(2) If $a \in \mathscr{L}$ and $a' \in \mathscr{L}'$ such that there is a question $\alpha \in a$ such that $\alpha \in a'$, then we shall write $a' = f(a)$.
In this way we define a map

$$f: \mathscr{L} \to \mathscr{L}'.$$

Theorem 12: If $a, b, a_i \in \mathscr{L}$ and $p', q' \in \Sigma'$ then the following holds:

(i) $a < b \Leftrightarrow f(a) < f(b)$, hence f is injective;

(ii) $f(\wedge_i a_i) = \wedge_i f(a_i)$;

(iii) $p' < f(a) \Leftrightarrow g(p') < a$;

(iv) $f(a) = \vee_{p < a} f(p)$;

(v) $p' < q' \Rightarrow g(p') < g(q')$ and g is surjective;

(vi) $a \perp b \Rightarrow f(a) \perp f(b)$.

Proof: (i) Follows immediately.

(ii) Take $\alpha_i \in a_i$ such that $\alpha_i \in f(a_i)$. Then $\pi_i \alpha_i \in \wedge_i a_i$ and $\pi_i \alpha_i \in \wedge_i f(a_i)$.

(iii) Suppose $p' < f(a)$ and S' is in state p' while S is in state $g(p')$. Then $f(a)$ is actual. Hence also a is actual. So $g(p') < a$. If $g(p') < a$. Then $p' < f(g(p')) < f(a)$.

(iv) If $p' < f(a)$ then $g(p') < a$. Since $g(p') < g(p')$ we have $p' < f(g(p'))$. So $p' < f(p)$ for some $p < a$. Hence $p' < \vee_{p<a} f(p)$. As a consequence $f(a) < \vee_{p<a} f(p)$. Clearly $\vee_{p<a} f(p) < f(a)$.

(v) We have $p' < q' < f(g(q'))$. Hence $g(p') < g(q')$.

(vi) Suppose $p' < f(a)$ and $q' < f(b)$. Then $g(p') < a$ and $g(q') < b$. So $g(p') \perp g(q')$. This means that there exists a question $\alpha \in Q$ such that $g(p') < \alpha$ and $g(q') < \alpha^{\sim}$. But since $Q \subset Q'$, we also have $\alpha \in Q'$. This shows that $f(g(p')) \perp f(g(q'))$. Since $p' < f(g(p'))$ and $q' < f(g(q'))$ we also have $p' \perp q'$. ■

To be able to define the universal problem for the tensor product we have to define the concept of bimorphisms. Suppose we consider three entities S_1, S_2, and S with sets of questions Q_1, Q_2, and Q, such that $Q_1 \subset Q$ and $Q_2 \subset Q$. Hence S is a refinement of S_1 and of S_2.

We introduce bimorphisms in the following way:

(1) If p represents the state of S, and p_1 represents the state of S_1, and p_2 represents that state of S_2, we shall write $p_1 = \nu_1(p)$ and $p_2 = \nu_2(p)$. In this way we define two morphisms

$$\nu_1: \Sigma \to \Sigma_1 \quad \text{and} \quad \nu_2: \Sigma \to \Sigma_2.$$

(2) If $a \in \mathscr{L}$, $a_1 \in \mathscr{L}_1$, and $a_2 \in \mathscr{L}_2$ such that there is a question $\alpha_1 \in a_1$ and a question $\alpha_2 \in a_2$ such that $\alpha_1 \cdot \alpha_2 \in a$ we shall write $a = \mu(a_1, a_2)$. In this way we define a bimorphism

$$\mu: \mathscr{L}_1 \times \mathscr{L}_2 \to \mathscr{L}.$$

Theorem 13: If $a_1 \in \mathscr{L}_1$, $a_2 \in \mathscr{L}_2$, and $p \in \Sigma$ we have
(i) $p < \mu(a_1, a_2) \Leftrightarrow \nu_1(p) < a_1$ and $\nu_2(p) < a_2$;
(ii) $a_1 \perp b_1$ or $a_2 \perp b_2 \Rightarrow \mu(a_1, a_2) \perp \mu(b_1, b_2)$;
(iii) $\mu(\wedge_i a_1^i, \wedge_j a_2^j) = \wedge_{ij} \mu(a_1^i, a_2^j)$;
(iv) $\mu(a_1, a_2) = \vee_{p_1 < a_1, p_2 < a_2} \mu(p_1, p_2)$.

Proof: (i) Suppose $p < \mu(a_1, a_2)$ and S is in state p while S_1 is in state $\nu_1(p)$ and S_2 is in state $\nu_2(p)$. Then $\mu(a_1, a_2)$ is actual. Hence a_1 and a_2 are actual. So $\nu_1(p) < a_1$ and $\nu_2(p) < a_2$. If $\nu_1(p) < a_1$ and $\nu_2(p) < a_2$, then $p < \mu(\nu_1(p), \nu_2(p))$. Hence $p < \mu(a_1, a_2)$.

(ii) Suppose $p < \mu(a_1, a_2)$ and $q < \mu(b_1, b_2)$. Then $\nu_1(p) < a_1$ and $\nu_2(p) < a_2$ and $\nu_1(q) < b_1$ and $\nu_2(q) < b_2$. Hence $\nu_1(p) \perp \nu_1(q)$ or $\nu_2(p) \perp \nu_2(q)$. Suppose $\nu_1(p) \perp \nu_1(q)$. Then there exists a question $\alpha_1 \in Q_1$ such that $\nu_1(p) < \alpha_1$ and $\nu_1(q) < \alpha_1^{\sim}$. Since $Q_1 \subset Q$ we have $\alpha_1 \in Q$. Then $\mu(\nu_1(p), \nu_2(p)) < \alpha_1$ and $\mu(\nu_1(q), \nu_2(q)) < \alpha_1^{\sim}$. As a consequence

$$\mu(\nu_1(p), \nu_2(p)) \perp \mu(\nu_1(q), \nu_2(q)).$$

But then $p \perp q$. ■

2. The universal problem

Let us denote the property lattice $\mathscr{L}$ constructed in 2 by $\mathscr{L}_1 \otimes \mathscr{L}_2$. The set of questions Σ we will denote by

$\Sigma_1 \otimes \Sigma_2$. The entity S considered in 2 we will denote by $S_1 \otimes S_2$.

Theorem 14: The map

$$T: \mathscr{L}_1 \times \mathscr{L}_2 \to \mathscr{L}_1 \otimes \mathscr{L}_2$$
$$(a_1, a_2) \mapsto a_1 \wedge a_2$$

is a bimorphism. The two maps

$$t_1: \Sigma_1 \otimes \Sigma_2 \to \Sigma_1$$
$$p_1 \wedge p_2 \mapsto p_1,$$
$$t_2: \Sigma_1 \otimes \Sigma_2 \to \Sigma_2$$
$$p_1 \wedge p_2 \mapsto p_2,$$

are morphisms.

Proof: Follows from the construction in 2. ■

Theorem 15: $(T, \mathscr{L}_1 \otimes \mathscr{L}_2)$ and $(t_1, t_2, \Sigma_1 \otimes \Sigma_2)$ are solutions of a universal problem defined as follows:

If we have an entity S with property lattice $\mathscr{L}$ and states space Σ and a bimorphism

$$\mu = \mathscr{L}_1 \times \mathscr{L}_2 \to \mathscr{L},$$

and a couple ν_1, ν_2 of morphisms

$$\nu_1: \Sigma \to \Sigma_1,$$
$$\nu_2: \Sigma \to \Sigma_2,$$

then there exists a morphism f,

$$f: \mathscr{L}_1 \otimes \mathscr{L}_2 \to \mathscr{L},$$

and a morphism g,

$$g: \Sigma \to \Sigma_1 \otimes \Sigma_2,$$

such that $\mu = f \circ T$, and $\nu_1 = t_1 \circ g$, and $\nu_2 = t_2 \circ g$.

Proof: Let us define f and g as follows:

$$f(a_1 \wedge a_2) = \mu(a_1, a_2),$$

and

$$g(p) = \nu_1(p) \wedge \nu_2(p).$$

Let us show that f and g are morphisms.

There exist $\alpha_1 \in a_1$ and $\alpha_2 \in a_2$ such that $\alpha_1 \cdot \alpha_2 \in \mu(a_1, a_2)$. But then $\alpha_1 \cdot \alpha_2 \in a_1 \wedge a_2$ and $\alpha_1 \cdot \alpha_2 \in f(a_1 \wedge a_2)$. Suppose the entity S is in state p. Then S_1 is in state $\nu_1(p)$ and S_2 is in state $\nu_2(p)$. As a consequence $S_1 \otimes S_2$ is in state $\nu_1(p) \wedge \nu_2(p) = g(p)$. ■

It is a general result of category theory that the solution of a universal product is unique up to an isomorphism. This theorem shows that the entity S considered in this paper in 2 really represents the tensor product of the two entities S_1 and S_2. Therefore, we can denote it by $S_1 \otimes S_2$. We can also conclude from this theorem that $S_1 \otimes S_2$ is the smallest refinement of S_1 and S_2.

3. Example of a universal situation, the separated product

Suppose that we consider an entity S composed of two separated entities S_1 and S_2. If Q_1 and Q_2 are the sets of questions, $\mathscr{L}_1$ and $\mathscr{L}_2$ the property lattices, Σ_1 and Σ_2 the state spaces of S_1 and S_2, then we constructed in Refs. 1 and 2 the set of questions $Q_1 \textcircled{\otimes} Q_2$, the property lattice $\mathscr{L}_1 \textcircled{\otimes} \mathscr{L}_2$ and the state space $\Sigma_1 \textcircled{\otimes} \Sigma_2$ of S. We called these structures the "separated" product. Let us introduce also for S the notation $S_1 \textcircled{\otimes} S_2$.

Theorem 16: If we consider the maps

$$\mu: \mathscr{L}_1 \times \mathscr{L}_2 \to \mathscr{L}_1 \textcircled{\otimes} \mathscr{L}_2$$
$$(a_1, a_2) \mapsto a_1 \wedge a_2,$$
$$\nu_1: \Sigma_1 \textcircled{\otimes} \Sigma_2 \to \Sigma_1$$
$$p_1 \wedge p_2 \mapsto p_1,$$
$$\nu_2: \Sigma_1 \textcircled{\otimes} \Sigma_2 \to \Sigma_2$$
$$p_1 \wedge p_2 \mapsto p_2,$$

then μ is a bimorphism and ν_1 and ν_2 are morphisms.

Proof: We have $\alpha_1 \in a_1$, $\alpha_2 \in a_2$, and $a_1 \cdot a_2 \in a_1 \wedge a_2$. $S_1 \textcircled{\otimes} S_2$ is in the state $p_1 \wedge p_2$ iff S_1 is in the state p_1 and S_2 is in the state p_2. ■

From Theorem 15 we can then conclude:

Theorem 17: The maps

$$f: \mathscr{L}_1 \otimes \mathscr{L}_2 \to \mathscr{L}_1 \textcircled{\otimes} \mathscr{L}_2$$
$$a_1 \wedge a_2 \mapsto a_1 \wedge a_2,$$
$$g: \Sigma_1 \textcircled{\otimes} \Sigma_2 \to \Sigma_1 \otimes \Sigma_2$$
$$p_1 \wedge p_2 \mapsto p_1 \wedge p_2$$

are morphisms.

4. QUANTUM MECHANICS AND THE TENSOR PRODUCT

We shall consider now two entities S_1 and S_2 described by quantum mechanics in Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$, and also the entity S consisting of S_1 and S_2 and described in the Hilbert space $\mathbf{H}_1 \otimes \mathbf{H}_2$. Let us try to see if S is a refinement of S_1 and S_2. Hence suppose that $\mathscr{L}(\mathbf{H}_1)$, $\mathscr{L}(\mathbf{H}_2)$, and $\mathscr{L}(\mathbf{H})$ are the property lattices and $\Sigma(\mathbf{H}_1)$, $\Sigma(\mathbf{H}_2)$, and $\Sigma(\mathbf{H})$ the state spaces of S_1, S_2, and S.

As we know for a quantum entity every property can be represented by a closed subspace of the Hilbert space describing the quantum entity. To make the notation not too complex we will make no distinction between the property and the closed subspace. Every ray of the Hilbert space represents a state of the quantum entity. Also here we will make no distinction between rays and states. We must remark that not every state must *a priori* correspond to a ray. There can be states corresponding to larger closed subspaces. We shall see that this is necessarily the case for S_1 and S_2 if we want S to be a refinement of S_1 and S_2. If S is a refinement of S_1 and S_2 there must exist a bimorphism

$$\mu: \mathscr{L}(\mathbf{H}_1) \times \mathscr{L}(\mathbf{H}_2) \to \mathscr{L}(\mathbf{H}),$$

and two morphisms

$$\nu_1: \Sigma(\mathbf{H}) \to \Sigma(\mathbf{H}_1),$$
$$\nu_2: \Sigma(\mathbf{H}) \to \Sigma(\mathbf{H}_2).$$

Theorem 18: If the quantum entity S described in the Hilbert space $\mathbf{H}_1 \otimes \mathbf{H}_2$ is a refinement of the quantum entity S_1 described in the Hilbert space $\mathbf{H}_1$ and the quantum entity S_2 described in the Hilbert space $\mathbf{H}_2$, then every property $a_1 \neq O_1$ of S_1 represents also a state of S_1 and every property $a_2 \neq O_2$ of S_2 represents also a state of S_2. In other words: $\Sigma(\mathbf{H}_1) = \mathscr{L}(\mathbf{H}_1) \setminus \{O_1\}$ and $\Sigma(\mathbf{H}_2) = \mathscr{L}(\mathbf{H}_2) \setminus \{O_2\}$.

Proof: Suppose that S is a refinement of S_1 and S_2. Consider an arbitrary property a_1 of S_1. Clearly $\mu(a_1, I_2)$

$= a_1 \otimes \mathbf{H}_2$. Suppose that $\{x_i\}$ is an orthonormal basis of a_1 and $\{y_j\}$ is an orthonormal basis of $\mathbf{H}_2$. Consider the vector

$$x = \Sigma_{ij}\delta_{ij}x_i \otimes y_j,$$

$\bar{x}$ represents a state of the entity S. If S is in the state $\bar{x}$, then $a_1 \otimes \mathbf{H}_2$ is actual. As a consequence the property a_1 is actual for the entity S_1. We shall now show that a_1 represents the state of the entity S_1. Suppose we have another property b_1 of S_1 such that $b_1 < a_1$ and $b_1 \neq a_1$. Consider an orthonormal basis $\{z_k\}$ of a_1 such that $\{z_k | k \leqslant k_1\}$ is a basis of b_1. Suppose now that $x \in b_1 \otimes \mathbf{H}_2$. Then $x = \Sigma_{kj}\lambda_{kj} z_k \otimes y_j$ such that $\lambda_{kj} = 0$ for $k > k_1$. Since $x_i \in a_1$, we can find numbers μ_{ik} such that $x_i = \Sigma_k \mu_{ik} z_k$. Since there is at least one $x_i \in a_1$ such that $x_i \notin b_1$, we must have at least one $\mu_{ik} \neq 0$ with $k > k_1$. On the other hand we have

$$\begin{aligned}\Sigma_{kj}\lambda_{kj}z_k \otimes y_j &= \Sigma_{ijk}\delta_{ij}\mu_{ik}z_k \otimes y_j \\ &= \Sigma_{kj}\mu_{jk}z_k \otimes y_j.\end{aligned}$$

From this follows that $\mu_{jk} = \lambda_{kj}$ for every j and every k. But then $\mu_{jk} = 0$ for every $k > k_1$, which is a contradiction. This shows that when S is in the state $\bar{x}$, then a_1 is the smallest property of S_1 that is actual. Hence a_1 represents the state of S_1. We show in an analogous way that every property a_2 of S_2 represents a state of S_2. ■

A consequence from this theorem is that, when S is a refinement of S_1 and of S_2, then the states of S_1 and S_2 are not atoms of the property lattices. Hence Axiom 3 of Ref. 2 is not satisfied for S_1 and S_2. We want to remark also that when we say that the state of S_1 is represented by the property a_1 which is not an atom of the property lattice, this does not mean that the state of S_1 is represented by a mixture. Indeed a mixture would mean that S_1 is in one of the vector states $\bar{x}_1$ with $x_1 \in a_1$, but we do not know which one. This is, however, not the case. If the entity S is in the state $\bar{x}$, then all the properties $\bar{x}_1$ of S_1 with $x_1 \in a_1$ are potential properties. It is only when the entity S is in a state $\overline{x_1 \otimes x_2}$, that $\bar{x}_1$ is an actual property for the entity S_1. We can also derive an interesting physical clarification for Axiom 3. We could say that when an entity S_1 has states that are not atoms, this entity is not separated from another entity S_2. When the entity S, consisting of S_1 and S_2, is in a state that is a superposition of product states, then the entity S_1 will be in a nonatom state. Hence Axiom 3, that introduces the atomicity for the states, is an axiom that will only be satisfied for entities that are more or less separated from the rest of the universe. Let us remark that for classical entities, Axiom 3 is always satisfied.[1,3]

We can now use Theorem 15 to have immediately the following result.

Theorem 19: If the quantum entity S described in the Hilbert space $\mathbf{H}_1 \otimes \mathbf{H}_2$ is a refinement of the quantum entity S_1 described in the Hilbert space $\mathbf{H}_1$ and the quantum entity S_2 described in the Hilbert space $\mathbf{H}_2$, then there exist morphisms

$$\begin{aligned}&f: \mathscr{L}(\mathbf{H}_1) \otimes \mathscr{L}(\mathbf{H}_2) \rightarrow \mathscr{L}(\mathbf{H}) \\ &\quad a_1 \wedge a_2 \mapsto a_1 \otimes a_2, \\ &g: \Sigma(\mathbf{H}) \rightarrow \Sigma(\mathbf{H}_1) \otimes \Sigma(\mathbf{H}_2) \\ &\quad \bar{x} \mapsto p_1 \wedge p_2,\end{aligned}$$

where p_1 and p_2 are constructed from the vector x as is shown in Theorem 18. From Theorem 18 follows that p_1 and p_2 are not always atoms. ■

5. MATHEMATICAL CONSTRUCTION OF THE TENSOR PRODUCT

The construction that we make in Sec. 2 depends on our formalism of questions, properties, and states. If is, however, very easy to construct the same tensor product without relying on the physical content of the formalism. Suppose we have two posets $\mathscr{L}_1$ and $\mathscr{L}_2$ with minimal elements O_1 and O_2, and maximal elements I_1 and I_2.

We construct the poset $\mathscr{L}_1 \otimes \mathscr{L}_2$ as follows:

$$\begin{aligned}\mathscr{L}_1 \otimes \mathscr{L}_2 = \{&(a_1,a_2) | a_1 \in \mathscr{L}_1, a_2 \in \mathscr{L}_2, \text{ and } a_1 \neq O_1, \\ &\text{and } a_1 \neq O_2\} \cup \{O\},\end{aligned}$$

with the following partial order relation:

$$(a_1, a_2) < (b_1, b_2) \Leftrightarrow a_1 < b_1 \text{ and } a_2 < b_2$$

$$O < (a_1, a_2) \text{ for all } (a_1, a_2).$$

If $\mathscr{L}_1$ and $\mathscr{L}_2$ are complete lattices we define least upper bounds and greatest lower bounds for $\mathscr{L}_1 \otimes \mathscr{L}_2$ as follows:

$$\begin{aligned}\bigwedge_i (a_1^i, a_2^i) &= \left(\bigwedge_i a_1^i, \bigwedge_i a_2^i\right) \\ &= O \text{ if } \bigwedge_i a_1^i = O \text{ or } \bigwedge_i a_2^i = O,\end{aligned}$$

$$\bigvee_i (a_1^i, a_2^i) = \left(\bigvee_i a_1^i, \bigvee_i a_2^i\right).$$

Then also $\mathscr{L}_1 \otimes \mathscr{L}_2$ is a complete lattice. If $\mathscr{L}_1$ and $\mathscr{L}_2$ are equipped with an orthogonality relation, we define an orthogonality relation for $\mathscr{L}_1 \otimes \mathscr{L}_2$ as follows:

$$(a_1,a_2) \perp (b_1,b_2) \Leftrightarrow a_1 \perp b_1 \text{ or } a_2 \perp b_2.$$

As shown already in the foregoing, $\mathscr{L}_1 \otimes \mathscr{L}_2$ cannot be equipped with an orthocomplementation compatible with this orthogonality relation, even when $\mathscr{L}_1$ and $\mathscr{L}_2$ are equipped with an orthocomplemantation.

6. CONCLUSION

In Refs. 1 and 2 we constructed the property lattice $\mathscr{L}_1 \textcircled{\otimes} \mathscr{L}_2$ for the entity S consisting of two separated entities S_1 and S_2 with property lattices $\mathscr{L}_1$ and $\mathscr{L}_2$. We showed that this separated product is never weakly modular when both S_1 and S_2 are not classical systems. It also never satisfies the covering law. Since in quantum mechanics the property lattices are always weakly modular and always satisfy the covering law, separated entities cannot be described in quantum mechanics.

In this paper we construct the property lattice $\mathscr{L}_1 \otimes \mathscr{L}_2$ for the entity S consisting of two entities S_1 and S_2 without knowing the interaction between S_1 and S_2. We show that $\mathscr{L}_1 \otimes \mathscr{L}_2$ is a tensor product and that it is never orthocomplemented when both S_1 and S_2 are nontrivial entities. It also never satisfies the covering law. Since in quantum mechanics the property lattices are always orthocomplemented, this situation cannot be described in quantum mechanics. For the same reason there does not exist a tensor

product in the category corresponding to quantum entities. Also in quantum logic it is not possible to construct a tensor product for the logics. This is because the logics in quantum logic are always supposed to be orthocomplemented.

[1]Dirk Aerts, " 'The one and the many'. Towards a unification of the classical and quantum description of one and many physical entities," Doctoral thesis, Vrije Universiteit Brussel, TENA, 1981.

[2]Dirk Aerts, "Description of many separated physical entities without the paradoxes encountered in quantum mechanics," Found. Phys. **12**, 1131 (1982).

[3]Dirk Aerts, "Classical theories and nonclassical theories as a special case of a more general theory," J. Math. Phys. **24**, 2441 (1983).

[4]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, MA, 1976).

[5]C. Piron, Annales de la Fondation Louis de Broglie **3**, 131 (1978).

[6]Dirk Aerts, "How do we have to change quantum mechanics in order to describe separated systems," in *The Wave–Particle Dualism*, edited by S. Diner, D. Fargue, G. Lockak, and F. Selleri (Reidel, Boston, 1984).

[7]Dirk Aerts, in *Current Issues in Quantum Logic*, edited by E. Beltrametti and B. C. Van Fraassen (Plenum, New York, 1981).

[8]Dirk Aerts, "Construction of a structure which makes it possible to describe the joint system of a classical system together with a quantum system," Rep. Math. Phys. (in press).

[9]S. P. Gudder, in *Mathematical Foundations of Quantum Theory*, edited by A. R. Marlow (Academic, New York, 1978).

[10]D. Aerts and I. Daubechies, Helv. Phys. Acta **51**, 661 (1978).

[11]D. J. Foulis and C. H. Randall in *Proceedings of the Colloquium on Interpretations and Foundations of Quantum Theories, Marburg, 1979*, edited by H. Neumann (in press).

[12]Dirk Aerts, "Morphisms and Entities," Preprint, TENA, Vrije Universiteit Brussel, 1982.

A six-dimensional oscillator basis classified by $O(6) \supset S\mathcal{O}(2) \times S\mathcal{U}(3) \supset SO(3)$

E. Chacón
Instituto de Física, UNAM, A. P. 20-364, 01000, México, D.F., Mexico

O. Castaños and A. Frank
Centro de Estudios Nucleares, UNAM, A. P. 70-543, Delegación Coyoacán, 04510 México, D.F., Mexico

(Received 16 May 1983; accepted for publication 11 August 1983)

We explicitly construct a complete set of states, useful in three-body problems, both in a boson operator realization and in terms of coordinates which are of interest to microscopic collective models. The states carry the angular momentum quantum number L and, for the classification scheme mentioned in the title, our expressions generalize to arbitrary L the results previously available only for $L = 0$ and 1.

PACS numbers: 03.65.Ge, 02.20. + b,

1. INTRODUCTION

The three-body system is a classical problem of theoretical physics. In particular we are interested in the three-nucleon system, in which the main goal is the construction of the wave function associated with the ground state of the bound system by a variational method. For this, one frequently uses sets of translationally invariant harmonic oscillator basis states. A convenient realization for these states can be given in terms of bosonic creation operators associated with two relative Jacobi vectors of the system $\mathbf{X}s$ and their canonically conjugate momenta $\mathbf{P}s$ through the relation

$$\eta_s = (1/\sqrt{2})(\mathbf{X}_s - i\mathbf{P}_s), \quad s = 1,2. \tag{1.1}$$

In turn, the corresponding annihilation operators are defined as

$$\xi_s = (1/\sqrt{2})(\mathbf{X}_s + i\mathbf{P}_s), \quad s = 1,2 \tag{1.2}$$

and give zero when applied on the vacuum state.

These operators satisfy the usual commutation relation, i.e.,

$$[\xi_{i's'}, \eta_{is}] = \delta_{s's}\delta_{i'i}, \tag{1.3}$$

where $i,i' = 1,2,3$ denote the components of the Jacobi vectors.

The basis states mentioned above are homogeneous polynomials of degree N *in* η_s: $P_N(\eta_s)|0\rangle$, which are known to carry the fully symmetric irreducible representation $[N\,0]$ of the group U(6),[1] and, in order to further classify them, one could in principle make use of any chain of subgroups starting from U(6). However, for physical considerations it is convenient that the states in the basis have the orbital angular momentum as a good quantum number, i.e., one wishes to include the three-dimensional rotation group O(3) as a subgroup of U(6). To fully characterize the states, one has then to introduce additional groups between the above two. In this respect the problem is very similar to the one related to the construction of basis states for symmetric irreducible representations of U(6) used in the interacting boson model (IBM) of the nucleus.[2,3] In both cases we have the physical rotation group O(3) as a subgroup of U(6), the main difference being that in the IBM case the six boson operators are assembled into two O(3) irreducible tensors of rank 0 and 2, while in the present analysis one has the two vectors η_1, η_2. There is one additional complication in the latter case, namely, that the states are required to have a definite behavior under permutations of the three objects.

Several basis states for three-body problems are known,[1,4–6] and their application in the determination of different physical observables, as, for example, the ground state wave function of a bound three-nucleon system, or a scattering state that would appear in the collisions of neutrons with deuterons, has been reported in the literature.[7–9]

The purpose of this paper is to construct a complete set of states classified according to the chain of groups

$$U(6) \supset O(6) \supset \underset{S\mathcal{U}(3)\supset SO(3)}{\overset{S\mathcal{O}(2)}{\times}} \tag{1.4}$$

in terms of bosonic operators and to give an additional realization for the states associated with this chain in terms of a set of coordinates introduced by Zickendraht and by Dzublik *et al.*,[10] which are currently receiving a great deal of attention in problems connected with a microscopic theory of nuclear collective motions.[11,12] By using these realizations of the basis states, it is possible to make calculations on the shape of the three-nucleon system,[13] which could improve our understanding of the microscopic collective theories mentioned above.

In the next section we derive the generators and invariant operators associated to the chain (1.4). We also introduce [cf. Eq. (2.13)] "symmetry-adapted" boson operators[4] in terms of which the permutational symmetry requirement is best dealt with. Using these bosons, we give in Sec. 3 analytic expressions for the basis states classified according to (1.4). In Sec. 4 we turn our attention to the form of the states in coordinate space, and finally in Sec. 5 we present some concluding remarks about our analysis, as well as a brief discussion of possible applications.

2. GENERATORS AND INVARIANT OPERATORS

The Jacobi vectors are identified by means of an index s having the values 1, 2 and their spherical components by an index m taking the values 1, 0, -1. In terms of the six creation boson operators η_{sm} and their Hermitian conjugate annihilation operators $\xi^{sm} = (-)^m \xi_{s-m}$ we can construct the 36 operators

$$\mathcal{C}^{s'm'}_{sm} = \eta_{sm}\xi^{s'm'}. \tag{2.1}$$

 0022-2488/84/051442-07$02.50

From (1.3) we can deduce that these operators have the following commutation relations:

$$[\mathscr{C}^{s'm'}_{sm}, \mathscr{C}^{s''m''}_{s''m''}] = \mathscr{C}^{s''m''}_{sm}\delta^{s'}_{s''}\delta^{m'}_{m''} - \mathscr{C}^{s'm'}_{s''m''}\delta^{s''}_{s}\delta^{m''}_{m}, \quad (2.2)$$

which identify them[5] as generators of a unitary group in six dimensions, U(6).

Taking the antisymmetric part of the U(6) generators, we obtain the operators

$$\begin{aligned} A^{s'm'}_{sm} &= \mathscr{C}^{s'm'}_{sm} - (-1)^{m+m'}\mathscr{C}^{s,-m}_{s',-m'} \\ &= (-1)^{m+m'}A^{s,-m}_{s',-m'}, \end{aligned} \quad (2.3)$$

of which the 15 independent ones are generators of an O(6) subgroup of U(6). They have the following commutation relations:

$$\begin{aligned} [A^{s'm'}_{sm}, A^{s''m''}_{s''m''}] = {} & A^{s''m''}_{sm}\delta^{s'}_{s''}\delta^{m'}_{m''} - A^{s'm'}_{s''m''}\delta^{s''}_{s}\delta^{m''}_{m} \\ & + (-1)^{m''+m''}A^{s'm'}_{s'',-m''}\delta_{ss''}\delta^{-m''}_{m} \\ & - (-1)^{m+m''}A^{s,-m}_{s''m''}\delta^{s's''}\delta^{-m''}_{m'}. \end{aligned} \quad (2.4)$$

This O(6) group admits as subgroup $S\mathscr{O}(2)\times S\mathscr{U}(3)$, whose generators are, respectively,

$$\mathscr{M} = -i\sum_{m=-1}^{1} A^{2m}_{1m} = -i(\boldsymbol{\eta}_1\cdot\boldsymbol{\xi}^2 - \boldsymbol{\eta}_2\cdot\boldsymbol{\xi}^1), \quad (2.5a)$$

$$\hat{L}_m,\ T^{m'}_{m} = A^{2m'}_{1m} - A^{1m'}_{2m} - \tfrac{2}{3}i\mathscr{M}\delta^{m'}_{m}, \quad (2.5b)$$

where

$$L_0 = \sum_s A^{s1}_{s1},\ L_1 = -\sum_s A^{s0}_{s1},\ L_{-1} = \sum_s A^{s1}_{s0}. \quad (2.5c)$$

Notice that, from (2.3),

$$T^{m'}_{m} = (-1)^{m+m'}T^{-m}_{-m'}, \quad (2.6a)$$

and, furthermore,

$$\sum_{m=-1}^{1} T^{m}_{m} = 0. \quad (2.6b)$$

Thus there are only five independent T's, which together with the three L's give the eight generators of $S\mathscr{U}(3)$. This is confirmed by the commutation relations

$$\begin{aligned} &[\hat{L}_m, T^{l}_{k}] \\ &\quad = \sum_{m'}\{(-1)^{l}\epsilon_{mm'-l}T^{m'}_{k} - (-1)^{m'}\epsilon_{mk-m'}T^{l}_{m'}\}, \end{aligned} \quad (2.7a)$$

$$\begin{aligned} [T^{m'}_{m}, T^{l}_{k}] = \sum_{n}\{ & (-1)^{m'}\epsilon_{k-m'n}\delta^{l}_{m} - (-1)^{l}\epsilon_{m-ln}\delta^{m'}_{k} \\ & + (-1)^{k+l+m'}\epsilon_{-l-m'n}\delta^{-k}_{m} \\ & - (-1)^{l}\epsilon_{mkn}\delta^{m'}_{-l}\}(-1)^{n}\hat{L}_n, \end{aligned} \quad (2.7b)$$

and

$$[\hat{L}_m, \hat{L}_{m'}] = \sum_{m''}(-1)^{m''}\epsilon_{mm'-m''}\hat{L}_{m''}. \quad (2.7c)$$

At the same time the latter result shows that $S\mathscr{U}(3)$ admits the subgroups $SO(3)\supset SO(2)$ whose generators are given by (2.5c).

In this way we have obtained the generators of the chain of groups

$$U(6)\supset O(6)\supset \overset{S\mathscr{O}(2)}{\underset{S\mathscr{U}(3)\supset SO(3)\supset SO(2)}{\times}} \quad (2.8)$$

which was originally introduced by Dragt.[14]

We next give the expressions for the invariant operators of these groups.

The linear invariant of U(6) is given by

$$\hat{N} = \sum_{s=1}^{2}\sum_{m=1}^{-1}\mathscr{C}^{sm}_{sm}. \quad (2.9)$$

In the realization (2.1) of the U(6) generators all the higher-order invariants are polynomials in $\hat{N}$ and therefore not independent of (2.9). Thus the irreps of U(6) occurring in our problem are of the completely symmetric type, i.e., $[N,O^5]$.

For O(6) the invariant operator is given by

$$\hat{\Lambda}^2 = \frac{1}{2}\sum_{s,s'=1}^{2}\sum_{m,m'=1}^{-1}A^{s'm'}_{sm}A^{sm}_{s'm'}. \quad (2.10)$$

Since this O(6) group is embedded in U(6) and we know that in our problem the unitary group has only symmetric irreps, it follows that the same thing happens for O(6), i.e., only irreps of type $(\lambda,0,0)$ can arise.[15]

Dragt[14] has shown the following relation between the quadratic and cubic invariant operators $C^{(2)}$, $C^{(3)}$ of $S\mathscr{U}(3)$, $\hat{\mathscr{M}}$ of $S\mathscr{O}(2)$, and the Casimir operator $\hat{\Lambda}^2$ of (2.10):

$$\hat{\Lambda}^2 = 2C^{(2)} - \tfrac{1}{3}\hat{\mathscr{M}}^2, \quad (2.11a)$$

$$C^{(3)} = \hat{\mathscr{M}}[\tfrac{1}{2}C^{(2)} - \tfrac{1}{4}\hat{\mathscr{M}}^2 + 1]. \quad (2.11b)$$

These results indicate that there exists a complementarity[16] relationship between the groups $S\mathscr{O}(2)$ and $S\mathscr{U}(3)$ within the irrep $(\lambda,0,0)$ of O(6). As a consequence, the irrep labels of $S\mathscr{U}(3)$ are determined by that of O(6), λ, and by the integer label of the associated irrep of $S\mathscr{O}(2)$, ν. The irrep of $S\mathscr{U}(3)$ is $(\lambda,(\lambda-\nu)/2)$, $\nu = \lambda, \lambda-2,\ldots,-\lambda+2, -\lambda$, where we are specifying the irreps by Young patterns of two rows.

Finally, the invariant operators of $SO(3)\supset SO(2)$ are given by

$$\hat{L}^2 = \sum_m (-1)^m\hat{L}_m\hat{L}_{-m},\ \hat{L}_0. \quad (2.12)$$

In the study of the behavior of three-particle wave functions under permutations, a great simplification is achieved by the introduction of new boson operators defined in terms of the Jacobi bosons used previously as

$$\boldsymbol{\alpha} = (1/\sqrt{2})(-i\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2),\ \boldsymbol{\beta} = (1/\sqrt{2})(i\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2), \quad (2.13a)$$

$$\boldsymbol{\alpha}^+ = (1/\sqrt{2})(i\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2),\ \boldsymbol{\beta}^+ = (1/\sqrt{2})(-i\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2), \quad (2.13b)$$

which have commutation relations of the same type as the Jacobi ones, i.e.,

$$[\alpha^+_m, \alpha_{m'}] = [\beta^+_m, \beta_{m'}] = \delta_{m,m'}, \quad (2.14)$$

with all other commutators equal to zero.

Special importance in the analysis of permutational symmetry is attached to the eigenfunctions of the operator $\hat{\mathscr{M}}$ of(2.5a), i.e.,

$$\hat{\mathscr{M}} = -i(\boldsymbol{\eta}_1\cdot\boldsymbol{\xi}^2 - \boldsymbol{\eta}_2\cdot\boldsymbol{\xi}^1) = \boldsymbol{\alpha}\cdot\boldsymbol{\alpha}^+ - \boldsymbol{\beta}\cdot\boldsymbol{\beta}^+. \quad (2.15)$$

It is important to remark that, besides the chain studied, the operator (2.15) is invariant in other chains of subgroups for the three-body problem. Of special interest for physical applications is the two oscillator basis,[4] i.e.,

$$U(6)\supset U^{(1)}(3)\times U^{(2)}(3)\supset O^{(1)}(3)\times O^{(2)}(3)\supset O(3), \quad (2.16)$$

where (2.15) is diagonal provided we adopt a realization for the generators of the groups and of the states involved in terms of the "symmetry-adapted" boson operators (2.13).

3. ANALYTIC EXPRESSIONS FOR THE BASIS STATES IN TERMS OF BOSON OPERATORS

In this section we first obtain a basis classified according to the chain of groups

$$U(6)\supset \underset{O(3)\supset O(2)}{\overset{S\mathscr{U}(2)}{\times}}. \quad (3.1)$$

This basis is more conveniently expressed in terms of the boson operators α, β, of (2.13a). States belonging to an irrep of O(6) can then be obtained from the previous basis by replacing the α, β bosons by "traceless bosons"[17] defined as

$$\mathbf{a}=\boldsymbol{\alpha}-2(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})(2N+6)^{-1}\boldsymbol{\beta}^{+},$$
$$\mathbf{b}=\boldsymbol{\beta}-2(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})(2N+6)^{-1}\boldsymbol{\alpha}^{+}. \quad (3.2)$$

The states constructed in this fashion have a classification according to (2.8), but their number of quanta is equal to λ; states with N quanta are obtained by multiplying the previous states by $(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})^{(N-\lambda)/2}$. We shall now implement the steps described above.

The basis states in the chain (3.1), to be denoted as $\mathbf{P}(\boldsymbol{\alpha},\boldsymbol{\beta})|0\rangle$ and having the highest weight in O(3), satisfy the equations

$$(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}^{+})\mathbf{P}|0\rangle=n_1\mathbf{P}|0\rangle, \quad (3.3a)$$
$$(\boldsymbol{\beta}\cdot\boldsymbol{\beta}^{+})\mathbf{P}|0\rangle=n_2\mathbf{P}|0\rangle, \quad (3.3b)$$
$$\hat{L}_0\mathbf{P}|0\rangle=(\alpha_1\alpha_1^{+}+\beta_1\beta_1^{+}-\alpha_{-1}\alpha_{-1}^{+}-\beta_{-1}\beta_{-1}^{+})\mathbf{P}|0\rangle=L\,\mathbf{P}|0\rangle, \quad (3.3c)$$
$$-\hat{L}_1\mathbf{P}|0\rangle=(\alpha_1\alpha_0^{+}+\beta_1\beta_0^{+}+\alpha_0\alpha_{-1}^{+}+\beta_0\beta_{-1}^{+})\mathbf{P}|0\rangle=0. \quad (3.3d)$$

Now, because of the commutation rules (2.14) obeyed by the boson operators, $\alpha_m^{+}\mathbf{P}(\boldsymbol{\alpha},\boldsymbol{\beta})|0\rangle$ is equivalent to $[\partial\mathbf{P}(\boldsymbol{\alpha},\boldsymbol{\beta})/\partial\alpha_m]|0\rangle$ and similarly for β_m^{+}; therefore, the polynomial $\mathbf{P}(\boldsymbol{\alpha},\boldsymbol{\beta})$ appearing in (3.3) must satisfy the partial differential equations

$$\sum_{m=1}^{-1}\alpha_m\frac{\partial\mathbf{P}}{\partial\alpha_m}=n_1\mathbf{P}, \quad (3.4a)$$
$$\sum_{m=1}^{-1}\beta_m\frac{\partial\mathbf{P}}{\partial\beta_m}=n_2\mathbf{P}, \quad (3.4b)$$
$$\sum_m m\left(\alpha_m\frac{\partial\mathbf{P}}{\partial\alpha_m}+\beta_m\frac{\partial\mathbf{P}}{\partial\beta_m}\right)=L\,\mathbf{P}, \quad (3.4c)$$
$$\sum_m\left(\alpha_m\frac{\partial\mathbf{P}}{\partial\alpha_{m-1}}+\beta_m\frac{\partial\mathbf{P}}{\partial\beta_{m-1}}\right)=0. \quad (3.4d)$$

The theory of invariants of algebraic forms[18] enables us to assert that the system of equations (3.4) possess a fundamental set of elementary solutions, i.e., an integrity basis, and that every polynomial solution of (3.4) is a monomial formed by a product of the elementary solutions raised to different powers. In fact, the fundamental set we are dealing with is that of the seminvariants[18] of two ground algebraic forms, both binary quadratic, which can be written as

$$f_1(\boldsymbol{\alpha};x,y)=\sqrt{2}\alpha_1x^2+\alpha_0xy+\sqrt{2}\alpha_{-1}y^2, \quad (3.5a)$$
$$f_2(\boldsymbol{\beta};x,y)=\sqrt{2}\beta_1x^2+\beta_0xy+\sqrt{2}\beta_{-1}y^2. \quad (3.5b)$$

The semivariants of these ground forms satisfy Eqs. (3.4), and an integrity basis for them contains the following six members[18]:

$$\alpha_1,\ \beta_1,\ (\boldsymbol{\alpha}\times\boldsymbol{\beta})_1,\ (\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}),\ (\boldsymbol{\beta}\cdot\boldsymbol{\beta}),\ (\boldsymbol{\alpha}\cdot\boldsymbol{\beta}), \quad (3.6)$$

where the dot means scalar product and the cross, vector product. Furthermore, there is one, and only one, algebraic relation (*syzygy*) connecting the six elementary solutions (3.6) namely,

$$4[(\boldsymbol{\alpha}\times\boldsymbol{\beta})_1]^2+(\alpha_1)^2(\boldsymbol{\beta}\cdot\boldsymbol{\beta})+(\beta_1)^2(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha})-2\alpha_1\beta_1(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})=0. \quad (3.7)$$

Thus a general polynomial solution of Eqs. (3.4) is

$$\mathbf{P}_{\lambda\nu kq_LM=L}(\boldsymbol{\alpha},\boldsymbol{\beta})=(\alpha_1)^{L+2q-(k-\nu)/2}(\beta_1)^{(k-\nu)/2-2q-\epsilon}\times[(\boldsymbol{\alpha}\times\boldsymbol{\beta})_1]^{\epsilon}(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha})^{(k-L-\epsilon)/2-q}\times(\boldsymbol{\beta}\cdot\boldsymbol{\beta})^{q}(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})^{(\lambda-k)/2}. \quad (3.8)$$

In (3.8) we have set $n_1=\frac{1}{2}(\lambda+\nu)$, $n_2=\frac{1}{2}(\lambda-\nu)$, with λ, ν being the eigenvalues of $\hat{N}$, $\hat{\mathscr{M}}$; and k, q are nonnegative integers which serve to distinguish among independent solutions having the same quantum numbers λ, ν, L. The identity (3.7) is taken into account by restricting ϵ in (3.8) to the values

$$\epsilon=0 \text{ if } k-L \text{ even},$$
$$\epsilon=1 \text{ if } k-L \text{ odd}. \quad (3.9)$$

We then come to the conclusion that a complete set of states classified according to the chain (3.1) and with highest weight in O(3) is

$$\mathbf{P}_{\lambda\nu kqL,M=L}(\alpha_m,\beta_m)|0\rangle \quad (3.10)$$

with $\mathbf{P}$ given by (3.8), (3.9). Note that k, q are not good quantum numbers, and thus states differing only in these labels are not orthogonal.

From the properties of the traceless bosons[17] it is known that if we make in (3.10) the replacements $\boldsymbol{\alpha}\to\mathbf{a}$, $\boldsymbol{\beta}\to\mathbf{b}$, the resulting state belongs to the irrep $(\lambda,0,0)$ of O(6) and still maintains the quantum numbers associated with the orthogonal subgroups in (3.1). From the discussion following Eq. (2.11) the state also carries the irrep label of $S\mathscr{U}(3)$; hence it is classified according to (2.8). In addition, we note that the irrep label of O(6) coincides with the number of quanta of the state.

Owing to the fact that

$$\mathbf{a}\cdot\mathbf{b}=4(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})(2N+6)^{-1}(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})(2N+8)^{-1}(\boldsymbol{\alpha}^{+}\cdot\boldsymbol{\beta}^{+}), \quad (3.11)$$

one can see that the introduction of traceless bosons in (3.10) leads to nonvanishing results only when $k=\lambda$, the O(6) basis states are thus given by

$$|\lambda\nu qLL\rangle=\mathbf{P}_{\lambda\nu k=\lambda qLL}(\mathbf{a},\mathbf{b})|0\rangle=(a_1)^{L+2q-(\lambda-\nu)/2}(b_1)^{(\lambda-\nu)/2-2q-\epsilon}[(\mathbf{a}\times\mathbf{b})_1]^{\epsilon}\times(\mathbf{a}\cdot\mathbf{a})^{(\lambda-L-\epsilon)/2-q}(\mathbf{b}\cdot\mathbf{b})^{q}|0\rangle. \quad (3.12)$$

We now express the states (3.12) that carry the irrep (λ) of O(6) in terms of the symmetry adapted bosons through Eq. (3.2). In the Appendix we show that

$$[(\mathbf{a}\times\mathbf{b})_1]^{\epsilon}(\mathbf{a}\cdot\mathbf{a})^{(\lambda-L-\epsilon)/2-q}(\mathbf{b}\cdot\mathbf{b})^{q}|0\rangle$$

$$= [(\alpha\times\beta)_1]^{\epsilon}\sum_r C_r(\alpha\cdot\alpha)^{(\lambda-L-\epsilon)/2-q-r}\times(\beta\cdot\beta)^{q-r}(\alpha\cdot\beta)^{2r}|0\rangle \tag{3.13}$$

with

$$C_r = \frac{(-1)^r((\lambda-L+\epsilon)/2-r)!}{r!((\lambda-L-\epsilon)/2-q-r)!(q-r)!}. \tag{3.14}$$

Thus the states (3.12) can be written as follows:

$$|\lambda vqLL\rangle = (a_1)^{L+2q-(\lambda-\nu)/2}(b_1)^{(\lambda-\nu)/2-2q-\epsilon}[(\alpha\times\beta)_1]^{\epsilon}\times\sum_r C_r(\alpha\cdot\alpha)^{(\lambda-L-\epsilon)/2-q-r}(\beta\cdot\beta)^{q-r}(\alpha\cdot\beta)^{2r}|0\rangle. \tag{3.15}$$

In the Appendix we make the application of the powers of the traceless bosons a_1, b_1 on the polynomial in α_m, β_m standing on the right of them in (3.15). After doing this, we get

$$|\lambda vqLL\rangle = [(\alpha\times\beta)_1]^{\epsilon}\times\sum_{\sigma,\tau} C^{\lambda vqL}_{\sigma\tau}(\alpha_1)^{L-(\lambda-\nu)/2+2q+\sigma-\tau}\times(\beta_1)^{(\lambda-\nu)/2-2q-\epsilon-\sigma+\tau}(\alpha\cdot\alpha)^{(\lambda-L-\epsilon)/2-\sigma-q}\times(\beta\cdot\beta)^{q-\tau}(\alpha\cdot\beta)^{\sigma+\tau}|0\rangle, \tag{3.16}$$

with

$$C^{\lambda vqL}_{\sigma\tau} = \frac{(-2)^{\sigma+\tau}(\lambda+1-\sigma-\tau)!}{((\lambda-L-\epsilon)/2-q-\sigma)!(q-\tau)!}\times\sum_r\frac{(-1)^r((\lambda-L+\epsilon)/2-r)!}{2^{2r}r!(\sigma-r)!(\tau-r)!}\times[(\lambda-L+\epsilon+1-2r)!((\lambda-\nu)/2-2q-\epsilon-\sigma+r)!(L-(\lambda-\nu)/2+2q-\tau+r)!]^{-1}. \tag{3.17}$$

These states can be generalized to arbitrary values of $M\neq L$ and can also be trivially extended to form a basis for the irrep $[N,O]$ of U(6) by writing them as

$$|N\lambda vqLM\rangle = \sum_{\sigma,\tau} C^{\lambda vqL}_{\sigma\tau}\times[\mathscr{Y}_{n+\epsilon+\sigma-\tau}(\alpha)\times\mathscr{Y}_{L-n-\sigma+\tau}(\beta)]^{L}_{M}\times(\alpha\cdot\alpha)^{(\lambda-L-\epsilon)/2-q-\sigma}(\beta\cdot\beta)^{q-\tau}\times(\alpha\cdot\beta)^{(N-\lambda)/2+\sigma+\tau}|0\rangle, \tag{3.18}$$

where $C^{\lambda vqL}_{\sigma\tau}$ is given in (3.17), and

$$n = L-(\lambda-\nu)/2+2q. \tag{3.19}$$

In (3.18) $\mathscr{Y}_{lm}$ is a solid harmonic, and we use the notation $[f_{l_1}\times g_{l_2}]^{L}_{M}$ for vector coupling of O(3) irreducible tensors. Setting $M=L$ and $N=\lambda$ in (3.18), we recover the expression (3.16), thus proving the claim made above.

It should be noted that (3.18) is a complete but nonorthogonal basis since states which differ only in the index q are nonorthogonal. Furthermore, the states are not normalized.

In the following section we discuss the realization of these basis states and those of the chain (2.16) in terms of the system of coordinates of Zickendraht and Dzublik *et al.*[10]

4. EXPRESSION OF BASIS STATES IN TERMS OF COORDINATES

The basis states discussed in the last section are six-dimensional harmonic oscillator functions, which in terms of coordinates can be given as a product of an hyperradial function $R_{N\lambda}(\rho)$ with $\rho^2=\mathbf{x}_1^2+\mathbf{x}_2^2$, times an angular function corresponding to a six-dimensional spherical harmonic $\mathbb{Y}_{\lambda vqLM}(\Omega)$, where Ω is a set of five angles to be defined later.

The hyperradial function $R_{N\lambda}(\rho)$ is known from the quantum mechanical analysis of the harmonic oscillator in hyperspherical coordinates[19] and is given by

$$R_{N\lambda}(\rho) = \left[\frac{2[(N-\lambda)/2]!}{\Gamma((N+L+6)/2)}\right]^{1/2}\times e^{-\rho^2/2}L^{\lambda+2}_{(N-\lambda)/2}(\rho^2)\rho^{\lambda}, \tag{4.1}$$

where L^{p}_{q} is a Laguerre polynomial of the indicated indices.

The product $\rho^{\lambda}\mathbb{Y}_{\lambda vqLM}(\Omega)$ is a homogeneous polynomial of degree λ in $\mathbf{x}_1$, $\mathbf{x}_2$ and it is the six-dimensional analog of the usual three-dimensional solid harmonic. These polynomials can be obtained immediately from our results in Sec. 3 by means of a theorem due to Dragt,[14] according to which if in the basis states with $N=\lambda$ we replace $\eta_1\to\mathbf{x}_1$, $\eta_2\to\mathbf{x}_2$ and suppress the vacuum state,[14] the resulting expression is a solid harmonic. By this procedure we deduce from (3.18)

$$\rho^{\lambda}\mathbb{Y}_{\lambda vqLM}(\Omega) = \pi^{3/2}2^{-\lambda/2}\sum_{\sigma,\tau} C^{\lambda vqL}_{\sigma\tau}(\mathbf{z}_2\cdot\mathbf{z}_2)^{q-\tau}\times(\mathbf{z}_1\cdot\mathbf{z}_1)^{(\lambda-L-\epsilon)/2-q-\sigma}(\mathbf{z}_1\cdot\mathbf{z}_2)^{\sigma+\tau}\times[\mathscr{Y}_{n+\epsilon+\sigma-\tau}(\mathbf{z}_1)\times\mathscr{Y}_{L-n-\sigma+\tau}(\mathbf{z}_2)]^{L}_{M}, \tag{4.2}$$

where

$$\mathbf{z}_1 = (1/\sqrt{2})(-i\mathbf{x}_1+\mathbf{x}_2),\quad \mathbf{z}_2 = (1/\sqrt{2})(i\mathbf{x}_1+\mathbf{x}_2), \tag{4.3}$$

and the coefficient C is given by (3.17).

We have not yet introduced any angles in (4.2), and in fact there are several ways of doing this. In the following we discuss the introduction of a set of angular coordinates which, as mentioned in the first section, are currently of interest in relation to a microscopic theory of nuclear collective motions.[10–13] These coordinates, ϑ_1, ϑ_2, ϑ_3, α, and γ, are defined by

$$x^{s}_{i} = \sum_{k=1}^{2}\rho_k D^{1_3}_{ki}(\vartheta_j)D^{1_2}_{k,s}(\alpha),\quad s=1,2,\ i=1,2,3, \tag{4.4a}$$

$$\rho_1=\rho\cos\gamma,\quad \rho_2=\rho\sin\gamma,\quad 0\leqslant\gamma\leqslant\pi/2, \tag{4.4b}$$

where $D^{1_3}(\vartheta_1,\vartheta_2,\vartheta_3)$ is a 3×3 rotation matrix, specifying, through the Euler angles ϑ_1, ϑ_2, ϑ_3, the relative orientation of an intrinsic frame with respect to the Lab frame and $D^{1_2}(\alpha)$ is the matrix

$$D^{1_2}(\alpha) = \begin{pmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{pmatrix} \tag{4.5}$$

From the orthogonality of the rotation matrices D^{1_3} and D^{1_2} we can deduce from (4.4a) that

$$\hat{x}^{s}_{i}\equiv\sum_{s'=1}^{2}\sum_{i'=1}^{3}D^{1_2}_{ss'}D^{1_3}_{ii'}x^{s'}_{i'} = \sum_{k=1}^{2}\rho_k\delta_{ik}\delta_{ks}. \tag{4.6}$$

As the functions (4.2) are bases for the irreps of SO(3) and $S\mathcal{O}(2)$, from angular momentum theory we can write

$$\mathbb{Y}_{\lambda vqLM}(\mathbf{z}_1,\mathbf{z}_2) = e^{iv\alpha}\sum_K D^L_{MK}{}^*(\vartheta)\mathbb{Y}_{\lambda vqLK}(\hat{z}_1,\hat{z}_2), \tag{4.7}$$

where $\hat{z}_1$, $\hat{z}_2$ are the vectors defined in (4.3) but in terms of $\hat{x}_s$, $s=1,2$.

From (4.6) and (4.3) it is easy to see that the spherical components of the vectors $\hat{z}_s$, $s=1,2$, are

$$\begin{aligned}
\hat{z}_{1,\pm 1} &= \tfrac{1}{2}i\rho(\pm\cos\gamma - \sin\gamma),\\
\hat{z}_{2,\pm 1} &= \tfrac{1}{2}i\rho(\mp\cos\gamma - \sin\gamma),\\
\hat{z}_{s,0} &= 0, \quad s=1,2,
\end{aligned} \tag{4.8}$$

while the scalar factors appearing in (4.2) can be expressed as

$$\begin{aligned}
\hat{\mathbf{z}}_1\cdot\hat{\mathbf{z}}_1 &= -\tfrac{1}{2}\rho^2\cos 2\gamma,\\
\hat{\mathbf{z}}_2\cdot\hat{\mathbf{z}}_2 &= -\tfrac{1}{2}\rho^2\cos 2\gamma,\\
\hat{\mathbf{z}}_1\cdot\hat{\mathbf{z}}_2 &= \tfrac{1}{2}\rho^2.
\end{aligned} \tag{4.9}$$

On the other hand, for the solid harmonics we have

$$\mathcal{Y}_{lm}(\hat{z}_s) = \left[\frac{(2l+1)(l+m)!(l-m)!}{2^l 4\pi}\right]^{1/2} \times \frac{(\hat{z}_{s,1})^{(l+m)/2}(\hat{z}_{s,-1})^{(l-m)/2}}{((l+m)/2)!((l-m)/2)!}, \tag{4.10}$$

where $l-m$ must be even as the polynomial is zero otherwise.

Expressing all factors on the right-hand side of (4.7) in terms of the vectors $\hat{z}_s$ by means of (4.9) and (4.10), we finally arrive at

$$\mathbb{Y}_{\lambda vqLM}(\vartheta_1,\vartheta_2,\vartheta_3,\alpha,\gamma) = e^{iv\alpha}\sum_K{}' D^L_{MK}{}^*(\vartheta_1\vartheta_2\vartheta_3)\sum_{l_1} G^{\lambda vqL}_{l_1K}(\gamma) f^{\lambda vqL}_{l_1}(\cos 2\gamma), \tag{4.11}$$

where the prime means that the summation over K goes through two-unit steps in the interval $[-L+\epsilon, L-\epsilon]$ and

$$G^{\lambda vqL}_{l_1K}(\gamma) = b\sum_{m_1,m_2}\langle l_1m_1,l_2m_2|LK\rangle S_{l_1m_1}(\gamma)S_{l_2m_2}(-\gamma) \tag{4.12a}$$

with $l_2 = L+\epsilon-l_1$,

$$b = [(-1)^{l_1}/4\pi]\sqrt{(2l_1+1)(2l_2+1)/2^{l_1+l_2}},$$

$$S_{lm}(\gamma) = (-1)^{(l+m)/2}\frac{\sqrt{(l+m)!(l-m)!}}{2^l((l+m)/2)!((l-m)/2)!} \times(\cos\gamma-\sin\gamma)^{(l+m)/2}(\cos\gamma+\sin\gamma)^{(l-m)/2}, \tag{4.12b}$$

$$f^{\lambda vqL}_{l_1}(\cos 2\gamma) = \sum_\sigma B_{\sigma l_1}(\cos 2\gamma)^{(\lambda-L-\epsilon)/2+l_1'-2\sigma}, \tag{4.13}$$

with $l_1' = l_1 - n - \epsilon$, where n was given in (3.19), and

$$B_{\sigma l_1} = \frac{(2)^{2\sigma-l_1'}(\lambda+1+l_1'-2\sigma)!}{(q+l_1'-\sigma)!((\lambda-L-\epsilon)/2-q-\sigma)!} \times\sum_t\frac{((-\lambda+L-\epsilon-1)/2)_t}{t!(\sigma-l_1'-t)!(n+l_1'-\sigma+t)!(\sigma-t)!(L-n-\epsilon-\sigma+t)!}. \tag{4.14}$$

Notice that we have omitted an irrelevant multiplicative factor in (4.11).

Thus, putting together (4.1) and (4.11), we get a complete set of basis states for three particles in terms of the Zickendraht–Dzublik coordinates.

Now we turn our attention to the chain (2.16), whose basis states are a vector coupled product of harmonic oscillator states,[1] i.e.,

$$F_{n_1l_1n_2l_2LM}(\mathbf{x}_1,\mathbf{x}_2) = \mathcal{R}_{n_1l_1}(r_1)\mathcal{R}_{n_2l_2}(r_2) \times r_1^{-l_1}r_2^{-l_2}[\mathcal{Y}_{l_1}(\mathbf{x}_1)\times\mathcal{Y}_{l_2}(\mathbf{x}_2)]^L_M. \tag{4.15}$$

Taking linear combinations of the states (4.15), it is possible to obtain eigenfunctions of the Casimir operator $\hat{\mathcal{M}}$ of $S\mathcal{O}(2)$, namely,[4]

$$\mathcal{M}\Phi = v\Phi, \quad v = 2n_1 + l_1 - 2n_2 - l_2, \tag{4.16a}$$

the explicit form of Φ being

$$\Phi_{n_1l_1n_2l_2LM}(\mathbf{x}_1,\mathbf{x}_2) = \sum_{\substack{n_1'l_1'\\ n_2'l_2'}} i^{l_1'}(-1)^{n_1'+l_1'} \times F_{n_1'l_1'n_2'l_2'LM}(\mathbf{x}_1,\mathbf{x}_2) \times\langle n_1'l_1'n_2'l_2'L\,|n_1l_1n_2l_2L\rangle, \tag{4.16b}$$

where the symbol $\langle\ |\ \rangle$ represents a standard transformation bracket[4] for harmonic oscillator functions.

In this case instead of (4.4a) it is more convenient to adopt the following alternative as definition of "collective" coordinates:

$$x_i^s = \sum_{k=2}^3 \rho_{k-1}D^{1_3}_{ki}(\vartheta_j)D^{1_2}_{k-1,s}(\alpha). \tag{4.17}$$

Using again the orthogonality of both rotation matrices D^{1_3} and D^{1_2}, and the definition (4.6), we obtain from (4.17)

$$\hat{x}_i^s = \sum_{k=2}^3 \rho_{k-1}\delta_{ik}\delta_{k-1,s},$$

i.e.,

$$\hat{\mathbf{x}}_1 = (0,\rho_1,0), \quad \hat{\mathbf{x}}_2 = (0,0,\rho_2). \tag{4.18}$$

The states (4.16) are bases for irreps of SO(3), SO(2), and $S\mathcal{O}(2)$; therefore, we have the relation

$$\Phi_{n_1l_1n_2l_2LM}(\mathbf{x}_1,\mathbf{x}_2) = \sum_K \Phi_{n_1l_1n_2l_2LK}(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2)D^L_{MK}{}^*(\vartheta_j)e^{iv\alpha}. \tag{4.19}$$

Using (4.10) and

$$Y_{l_2m_2}(0,0) = [(2l_2+1)/4\pi]^{1/2}\delta_{m_2,0}$$

in (4.15), we can write

$$F_{n_1'l_1'n_2'l_2'LK}(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_2) = (-i)^{l_1'}\mathcal{R}_{n_1'l_1'}(\rho_1)\mathcal{R}_{n_2'l_2'}(\rho_2)A^{LK}_{l_1'l_2'}, \tag{4.20}$$

with

$$A^{LK}_{l_1'l_2'} = \frac{\langle l_1' l_2' K 0 | LK \rangle \tfrac{1}{2}[1+(-)^{l_1'+K}]}{4\pi 2^{l_1'}((l_1'+K)/2)!((l_1'-K)/2)!} \times [(2l_1'+1)(2l_2'+1)(l_1'+K)!(l_1'-K)!]^{1/2}. \tag{4.21}$$

Introducing (4.20) in (4.16b) and then this in (4.19), we obtain the two-oscillator wave functions Φ expressed in terms of Z–D coordinates, namely,

$$\Phi_{n_1 l_1 n_2 l_2 LM}(\mathbf{x}_1,\mathbf{x}_2) = e^{i\nu\alpha} \sum_{\substack{n_1' l_1' \\ n_2' l_2'}} (-1)^{n_1'+l_1'} \langle n_1' l_1' n_2' l_2' L \,|\, n_1 l_1 n_2 l_2 L \rangle \times \sum_K A^{LK}_{l_1' l_2'} \mathscr{R}_{n_1' l_1'}(\rho_1) \mathscr{R}_{n_2' l_2'}(\rho_2) D^{L}_{MK}{}^*(\vartheta_i), \tag{4.22}$$

where ν is given in (4.16a) and A in (4.21).

5. DISCUSSION

We have constructed explicit realizations of the basis states classified according to the chain of groups (1.4) and (2.16) in terms of a set of coordinates which is especially suited to extract collective features from the dynamics of a system of particles.[11] Our result (4.11) for the states classified by O(6) generalize the partial results known previously[14] for the cases $L=0$ and $L=1$.

In regard to the analysis of Sec. 3 that led to the states (3.16) we want to stress the fact that the presence of the groups O(6) and $S\mathscr{O}(2)$ in the classification scheme brings forth automatically the group $S\mathscr{U}(3)$ owing to the known[14,16] complementarity of $S\mathscr{O}(2)\times S\mathscr{U}(3)$ within the symmetric irrep $(\lambda,0,0)$ of O(6). One can avail of this property to construct basis states classified by a subgroup G of $S\mathscr{U}(3)$, other than SO(3) discussed here, in the chain

$$O(6) \supset \underset{S\mathscr{U}(3)}{\overset{S\mathscr{O}(2)}{\times}} \supset \underset{G}{\overset{S\mathscr{O}(2)}{\times}} \tag{5.1}$$

by following the same procedure of the present paper to introduce O(6), i.e., use of the traceless bosons (3.2). In fact, such a construction has been carried out in Ref. 20 for the case when $G = SA(3)$, the group of diagonal $S\mathscr{U}(3)$ matrices, giving the nonorthogonal Weyl basis of $S\mathscr{U}(3)$ expressed in terms of coordinates.

A similar analysis could also be done for $G = SU(2)\times U(1)$, which eventually should give an orthonormal basis in terms of coordinates classified according to

$$O(6)\supset S\mathscr{O}(2)\times S\mathscr{U}(3)\supset U(1)\times SU(2)\supset SO(2). \tag{5.2}$$

In this case, however, there is no need to actually do the analysis, since the corresponding results were given long time ago by Bég and Ruegg,[21] who obtained six-dimensional solid harmonics, i.e., bases for symmetric irreps of O(6), with the classification (5.2). It is a consequence of Dragt's theorem that the results of Ref. 21 should be the same as those obtained by an analysis along the sequence $U(6)\supset S\mathscr{O}(2)\times G \rightarrow$traceless bosons$\rightarrow$ordinary bosons$\rightarrow$coordinates, similar to that of our Secs. 3 and 4. We want to remark that, while for the canonical scheme (5.2) an analytical approach turned out to be feasible, for the noncanonical chains discussed both in this paper and in Ref. 20 the method of traceless bosons is indispensable in order to reach at general closed analytic expressions for the O(6) basis states.

The present results can find its main aplication in the theoretical analysis of three-body problems. Although several other bases were previously known,[4–6] we would like to point out that often a particular basis can offer an advantage over the others for the study of a specific aspect of the three-body problem. For instance, it is frequently claimed[22] that a calculation of the binding energy of the three-nucleon system in a hyperspherical basis, consisting of a hyperradial function times a six-dimensional spherical harmonic, converges faster than a similar calculation in a two-oscillator basis. The present analysis provides us with a set of hyperspherical harmonics, which after multiplication by a suitable hyperradial part could be used as basis functions for a variational calculation of the energy of the ground state of the three-nucleon system. These functions provide us with the means to study in a simple way the shape of the three-body system. In particular, they determine the collective probability density and all the relevant quantities associated with the quadrupole tensor of the mass distribution of the system.

ACKNOWLEDGMENTS

One of the authors (E. C.) wants to thank V. C. Aguilera-Navarro and D. A. Agrello of Instituto de Física Teorica, Sao Paulo, Brazil, for several conversations that led to the elucidation of some of the topics presented at the beginning of Sec. 3.

APPENDIX: DERIVATION OF EQS. (3.13) AND (3.16)

In the first part of this appendix we present some details of the analysis leading to Eq. (3.13). The state on the left-hand side of (3.13) has degree $\lambda - L + \epsilon$ in boson operators, has angular momentum ϵ, is an eigenstate of $\hat{\mathscr{M}}$ with eigenvalue $\lambda - L - \epsilon - 4q$, and, besides, is a part of a basis for an irrep of O(6); thus it can be written as

$$\mathbf{P}_\epsilon|0\rangle = ([\boldsymbol{\alpha}\times\boldsymbol{\beta}]_1)^\epsilon \sum C_{n_1 n_2 r}(\boldsymbol{\alpha}\cdot\boldsymbol{\alpha})^{n_1}(\boldsymbol{\beta}\cdot\boldsymbol{\beta})^{n_2}(\boldsymbol{\alpha}\cdot\boldsymbol{\beta})^{2r}|0\rangle, \tag{A1}$$

which has the required angular momentum but still has to satisfy

$$(\boldsymbol{\alpha}\cdot\boldsymbol{\beta}^+)\mathbf{P}_\epsilon|0\rangle = (\lambda - L + \epsilon)\mathbf{P}_\epsilon|0\rangle, \tag{A2}$$

$$\hat{\mathscr{M}}\mathbf{P}_\epsilon|0\rangle = (\lambda - L - \epsilon - 4q)\mathbf{P}_\epsilon|0\rangle, \tag{A3}$$

$$(\boldsymbol{\alpha}^+\cdot\boldsymbol{\beta}^+)\mathbf{P}_\epsilon|0\rangle = 0. \tag{A4}$$

Equations (A2) and (A3) determine n_1, n_2, namely,

$$n_1 = \tfrac{1}{2}(\lambda - L - \epsilon) - q - r, \quad n_2 = q - r, \tag{A5}$$

and we are left in (A1) with a single summation over r with coefficients C_r. Equation (A4) then gives a recurrence relation for C_r which enables one to determine these coefficients, apart from a multiplicative factor. When dealing with Eq. (A4), we recall that $\alpha_m^+ \mathbf{P}_\epsilon|0\rangle = (-1)^m(\partial\mathbf{P}_\epsilon/\partial\alpha_{-m})|0\rangle$, and similarly for β_m^+; thus it is a matter of applying on $\mathbf{P}_\epsilon$ the Laplacian operator

$$\nabla^2 = \sum_m (-1)^m \frac{\partial^2}{\partial\alpha_m \partial\beta_{-m}}. \tag{A6}$$

As $\nabla^2[\alpha\times\beta]_1 = 0$, we deduce from (A1) that

$$\nabla^2 \mathbf{P}_\epsilon = ([\alpha\times\beta]_1)^\epsilon \times \sum_r C_r \{4r\epsilon M_{n_1,n_2,2r-1} + \nabla^2 M_{n_1,n_2,2r}\}, \tag{A7}$$

where we have introduced the notation

$$M_{n_1,n_2,n_3} \equiv (\alpha\cdot\alpha)^{n_1}(\beta\cdot\beta)^{n_2}(\alpha\cdot\beta)^{n_3}. \tag{A8}$$

Since

$$\nabla^2 M_{n_1,n_2,2r} = 4n_1n_2 M_{n_1-1,n_2-1,2r+1} + 4r(n_1+n_2+1+r)M_{n_1,n_2,2r-1}, \tag{A9}$$

we have, using (A5) and making an appropriate shift of dummy index,

$$\nabla^2 \mathbf{P}_\epsilon = 4([\alpha\times\beta]_1)^\epsilon \times \sum_r \left\{ \left(\frac{\lambda - L - \epsilon}{2} - q - r \right)(q-r)C_r + (r+1) \times \left(\frac{\lambda - L + \epsilon}{2} - r \right) C_{r+1} \right\} M_{n_1-1,n_2-1,2r+1}. \tag{A10}$$

Equation (A4) will be satisfied provided we choose C_r such that the curly bracket in (A10) is equal to zero. The solution to this recurrence relation is the C_r given in (3.14). Having thus proved Eq. (3.13), we turn now to the derivation of Eq. (3.16).

In this case we have to apply on $\mathbf{P}_\epsilon$ powers of the differentials operators

$$\left[\alpha_1 + 2(\alpha\cdot\beta)(2N+6)^{-1}\frac{\partial}{\partial\beta_{-1}}\right], \tag{A11}$$

$$\left[\beta_1 + 2(\alpha\cdot\beta)(2N+6)^{-1}\frac{\partial}{\partial\alpha_{-1}}\right]. \tag{A12}$$

By the method of induction we can prove that

$$\left[\beta_1 + 2(\alpha\cdot\beta)(2N+6)^{-1}\frac{\partial}{\partial\alpha_{-1}}\right]^l ([\alpha\times\beta]_1)^\epsilon M_{n_1,n_2,2r} = \frac{l!n_1!(2n_1+2n_2+4r+1+2\epsilon)!}{(2n_1+2n_2+2r+1+2\epsilon)!(2n_1+2n_2+4r+1+2\epsilon+l)!} \times \sum_{s=0}^{l} \left\{ \frac{(-2)^s(2n_1+2n_2+2r+1+2\epsilon+l-s)!}{s!(l-s)!(n_1-s)!} ([\alpha\times\beta]_1)^\epsilon(\alpha_1)^s(\beta_1)^{l-s} M_{n_1-s,n_2,2r+s} \right\}. \tag{A13}$$

By a similar procedure, applying powers of (A12) on the previous polynomial, we have

$$\left[\alpha_1 + 2(\alpha\cdot\beta)(2N+6)^{-1}\frac{\partial}{\partial\beta_{-1}}\right]^{l'} ([\alpha\times\beta]_1)^\epsilon(\alpha_1)^s(\beta_1)^{l-s} M_{n_1-s,n_2,2r+s} = \frac{l'!n_2!(2n_1+2n_2+4r+2\epsilon+1+l)!}{(2n_1+2n_2+2r-s+2\epsilon+1+l)!(2n_1+2n_2+4r+2\epsilon+1+l+l')!} \sum_{t=0}^{l'} \frac{(-2)^t}{t!(l'-t)!(n_2-t)!} \times (2n_1+2n_2+2r-s+2\epsilon+1+l+l'-t)!([\alpha\times\beta]_1)^\epsilon(\alpha_1)^{l'+s-t}(\beta_1)^{l-s+t} M_{n_1-s,n_2-t,2r+s+t}. \tag{A14}$$

Equation (3.16) is obtained by introducing the results given in (A13), (A14) into Eq. (3.15), then making a change to new dummy indexes $\sigma \equiv r+s$, $\tau = r+t$, and giving to n_1, n_2 the values indicated in (A5) and to l, l' the values $l = (\lambda-\nu)/2 - 2q - \epsilon$, $l' = L + 2q - (\lambda-\nu)/2$.

[1] P. Kramer and M. Moshinsky, "Group Theory of Harmonic Oscillators and Nuclear Structure," in *Group Theory and Its Applications, Vol. I*, edited by E. M. Loebl (Academic, New York, 1968).
[2] A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[3] O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[4] M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms to Quarks* (Gordon and Breach, New York, 1968).
[5] E. Chacón, D. Levi, and M. Moshinsky, Rev. Mex. Fís. **22**, 291 (1973).
[6] P. Kramer and M. Moshinsky, Nucl. Phys. **82**, 241 (1966).
[7] M. R. Strayer and P. U. Sauer, Nucl. Phys. A **231**, 1 (1974).
[8] F. T. Smith, J. Math. Phys. **3**, 735 (1962).
[9] G. Erens, J. L. Vischers, and R. V. Wageningen, Ann. Phys. **67**, 461 (1971).
[10] W. Zickendraht, J. Math. Phys. **12**, 1663 (1971); A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[11] V. Vanagas, *The Microscopic Nuclear Theory, Lecture Notes in Physics* (University of Toronto, Toronto, 1977); V. Vanagas, "The Microscopic Theory of the Collective Motion in Nuclei," in *Group Theory and Its Applications in Physics—1980*, edited by J. H. Seligman (AIP, New York, 1980), p. 220.
[12] G. F. Filippov, V. I. Ovcharenko, and Yu. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (in Russian) (Navkova Dumka, Kiev, 1981).
[13] Yu. F. Smirnov and G. F. Filippov, Yad. Fiz. **27**, 73 (1978) [Sov. J. Nucl. Phys. **27**, 39 (1978)].
[14] A. J. Dragt, J. Math. Phys. **6**, 533 (1965).
[15] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, MA, 1962), Chap. 7.
[16] M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[17] M. A. Lohe, "The Development of the Boson Calculus for the Orthogonal and Symplectic Groups," thesis, University of Adelaide, 1974.
[18] H. W. Turnbull, *The Theory of Determinants, Matrices and Invariants* (Dover, New York, 1960); G. Salmon, *Modern Higher Algebra* (Chelsea, New York, 1964).
[19] M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[20] E. Chacón and G. Germán, Physica A **114**, 301 (1982).
[21] M. A. B. Bég and H. Ruegg, J. Math. Phys. **6**, 677 (1965).
[22] J. S. Levinger, *The Two and Three Body Problem*, Springer Tracts in Modern Physics, Vol. 71 (Springer-Verlag, Berlin, 1974).

Two-dimensional scattering: The number of bound states from scattering data[a)]

Margaret Cheney
Department of Mathematics, Stanford University, Stanford, California 94305

(Received 7 December 1982; accepted for publication 28 October 1983)

Relations are found between scattering data and the spectrum for the two-dimensional Schrödinger operator $\Delta + V(x)$, where V is a local noncentral potential. In particular, a two-dimensional version of the Levinson theorem is obtained; this theorem gives the number of bound states in terms of the change in phase of the determinant of the scattering operator.

PACS numbers: 03.65.Nk

INTRODUCTION

In this paper we investigate the relation between scattering data and the spectrum for the two-dimensional Schrödinger operator $\Delta + V(x)$, where V is a local noncentral potential. The goal is to prove a generalized Levinson theorem, which gives a formula for the number of bound states in terms of the change in phase of the determinant of the scattering operator. This theorem, important to inverse scattering because it allows determination of the number of bound states from the scattering data, is somewhat different from its three-dimensional counterpart[1] and has a proof that is substantially more intricate.

In Sec. 1 we study the Lippmann–Schwinger equation corresponding to the two-dimensional Schrödinger equation, and we extract the scattering amplitude from the large x behavior of the wave function (see the Appendix). Although the Lippmann–Schwinger Green's function has a logarithmic singularity at zero energy, this results in an operator singularity that is only of rank one; a few facts are proved about the well-behaved remainder.

In Sec. 2 we consider the exceptional points and their relation to the Fredholm determinant. Because the two- and three-dimensional proofs are essentially the same, the bulk of this section is a short summary of the relevant results of Ref. 1.

Section 3 contains the Levinson theorem and its proof, which is based on spectral theory together with the argument principle and facts from Secs. 1 and 2. This version applies only to the case in which zero-energy exceptional points are absent; the case in which they are present is still under investigation.

1. PRELIMINARIES

Two-particle scattering in the center of mass system is governed by the time-independent Schrödinger equation

$$-\Delta\psi(k,x) + V(x)\psi(k,x) = k^2\psi(k,x).$$

Here $x \in R^2$, the potential $V(x)$ is real valued, and k is a positive scalar.

Scattering solutions are defined by the Lippmann–Schwinger equation

$$\psi(k,\theta,x) = \exp(ik\theta\cdot x) + \int G(k,|x-y|)V(y)\psi(k,\theta,y)d^2y, \tag{1.1}$$

where θ denotes a unit vector in R^2 and the function G is a fundamental solution of $\Delta + k^2$. We take G to be

$$G(k,r) = -(i/4)H_0^{(1)}(kr),$$

where H_0 is the zero-order Hankel function and $r = |x|$.

In order to apply Fredholm theory, we multiply the Lippmann–Schwinger equation by $|V(x)|^{1/2}$ and make the following definitions:

$$\xi(k,\theta,x) = |V(x)|^{1/2}\psi(k,\theta,x),$$
$$\xi^0(k,\theta,x) = |V(x)|^{1/2}\exp(ik\theta\cdot x),$$
$$V_{1/2}(y) = V(y)/|V(y)|^{1/2},$$
$$K(k)f(x) = \int |V(x)|^{1/2}G(k,|x-y|)V_{1/2}(y)f(y)d^2y.$$

With this notation, the Lippmann–Schwinger equation becomes

$$\xi(k,\theta,x) = \xi^0(k,\theta,x) + K(k)\xi(k,\theta,x). \tag{1.2}$$

For k bounded away from zero, we prove the following result concerning the operator $K(k)$:

Proposition 1-1: Suppose $V \in L^2$ with $\int\int |V(x)V(y)||x-y|^{-1}d^2x\,d^2y = M < \infty$. Then for each $k_0 > 0$, the estimate $\|K(k)\|_{\text{H.S.}} \leqslant ck^{-1/2}$ holds for $k > k_0$, where c depends only on k_0 and on V.

Proof: We apply the definition[2] of the Hilbert–Schmidt norm $\|\ \|_{\text{H.S.}}$ to the operator K defined above; we write $\|K(k)\|^2_{\text{H.S.}} = I_1 + I_2$, where I_1 and I_2 are pieces corresponding to regions of integration $k|x-y| < 1$ and $k|x-y| > 1$, respectively. Well-known estimates[3] on the behavior of the Hankel function give us the following estimates:

$$I_1 \leqslant c\int\int_{|x-y|<k^{-1}} |V(x)V(y)||\log k|x-y||^2\,d^2x\,d^2y, \tag{1.3}$$

$$I_2 \leqslant c\int\int_{|x-y|>k^{-1}} |V(x)V(y)|(k|x-y|)^{-1}\,d^2x\,d^2y.$$

In (1.3), we let $z = x - y$, interchange the order of inte-

[a)] This is based on the author's Ph.D. thesis, "Quantum Mechanical Scattering and Inverse Scattering in Two Dimensions," Indiana University, 1982.

 0022-2488/84/051449-07$02.50

gration, and apply the Schwarz inequality to obtain

$$I_1 \leqslant c \int |\log k\,|z||^2 \int |V(z+y)V(y)|\, d^2y\, d^2z$$
$$\leqslant c\|V\|_2^2 \int_0^{k^{-1}} \int_{S^1} |\log kr|^2\, d\theta\, r\, dr \leqslant ck^{-2}.$$

I^2 is bounded by $1/k \times M$, which was defined in the hypotheses. ■

Remark 1-2: A Sobolev inequality[4] implies finiteness of the integral

$$\int\int |V(x)V(y)|/|x-y| d^2x\, d^2y$$

provided $V \in L^{4/3}$. However, if V belongs to $L^1 \cap L^2$, it then belongs to $L^{4/3}$. Henceforth we will usually assume that V belongs to $L^1 \cap L^2$ because this assumption allows us to apply Fredholm theory to (1.2); we obtain a unique solution $\xi(k,\theta,x)$ provided the operator K does not have the eigenvalue 1. Note that for k large enough, the operator norm of $K(k)$ is less than 1, which certainly implies that (1.2) is uniquely solvable (by iteration, in fact).

Values of k for which $K(k)$ has eigenvalue 1 are called *exceptional points* or *exceptional values*. It will be shown in Sec. 2 that there are no nonzero real exceptional points. For nonexceptional values of k, we have the following:

Proposition 1-3: Let $V \in L^2$ with $\int |V(x)||x|^4\, d^2x < \infty$. Let $k > 0$, and let $\xi = |V|^{1/2}\psi$ solve (1.2). Then

$$\psi(k,\theta,x) = \exp(ik\theta\cdot x) + \exp(-3\pi i/4)(8\pi)^{-1/2} A(k,\hat{x},\theta) \times \exp(ik|x|)(k|x|)^{-1/2} + h(k,\theta,x),$$

where

$$\hat{x} = x/|x|,$$

$$A(k,\theta,\theta') = \int \exp(-ik\theta\cdot x) V(x)\psi(k,\theta',x)\, d^2x, \qquad (1.4)$$

and

$$h(k,\theta,x) \in L^2(x) \quad \text{uniformly in } \theta.$$

Proof: The proof is given in the Appendix. ■

Remark: If V is in L^2 with $\int |V(x)||x|^4\, d^2x < \infty$, then V is also in L^1.

The quantity $A(k,\theta,\theta')$ is called the *scattering amplitude*; it essentially gives us the large x behavior of the wave function.

For nonexceptional k (i.e., for $k > 0$), application of the Schwarz inequality to (1.4) shows us that for $V \in L^1 \cap L^2$, the scattering amplitude is bounded:

$$|A(k,\theta,\theta')| \leqslant \|V\|_1 \|\xi(k)\|_2 \leqslant \|V\|_1^2 \|(I-K)^{-1}\|.$$

We let the scattering amplitude act on $L^2(S^1)$ via

$$(A(k)f)(\theta) = \int_{S^1} A(k,\theta',\theta) f(\theta')\, d\theta'.$$

For each k, the operator $A(k)$ is then bounded and linear on $L^2(S^1)$. We also define the *scattering operator* or *S-matrix* $S(k)$ on $L^2(S^1)$ by

$$S(k) = I - i(4\pi)^{-1}(\operatorname{sgn} k) A(k).$$

For the Levinson theorem, we will need to know about the small-k behavior of $K(k)$. For this we use the small-argument behavior of the Hankel function to rewrite K in terms of the operators L and P defined as follows:

$$L(k)f(x) = -(i/4) \int |V(x)|^{1/2} (H_0^{(1)}(k|x-y|) - (2i/\pi)\log k) \times V_{1/2}(y) f(y)\, d^2y, \qquad (1.5)$$

$$Pf(x) = |V(x)|^{1/2} (2\pi)^{-1} (V_{1/2}, f), \qquad (1.6)$$

where $(\ ,\ )$ denotes the $L^2(R^2)$ inner product. With this notation, we can write $K(k)$ as

$$K(k)f = (\log k) Pf + L(k) f.$$

We note that P is of rank 1 and $L(k)$ is well behaved at $k = 0$.

Lemma 1-4: Let V belong to L^1 with $\int |V(x)||x|\, d^2x < \infty$ and $\int\int |V(x)V(y)||\log|x-y||^2\, d^2x\, d^2y = M < \infty$, and let k_0 be positive. Then the operator $L(k)$ defined by (1.5) is Hilbert–Schmidt, and for $k < k_0$ the Hilbert–Schmidt norm satisfies

$$\|L(k)\|_{\text{H.S.}} \leqslant c,$$

where c depends on V and k_0, but not on k.

Proof: We apply the definition of the Hilbert–Schmidt norm to (1.5) and use standard estimates for $H_0^{(1)}$:

$$\|L(k)\|_{\text{H.S.}} \leqslant \left(\frac{1}{4}\right) \left(\iint_{|x-y|<k^{-1}} |V(x)V(y)||\log|x-y||^2\, d^2x\, d^2y \right)^{1/2} + c\|V\|_1 + \left(c \iint_{k|x-y|>1} |V(x)V(y)|(k|x-y|)^{-1}\, d^2x\, d^2y \right)^{1/2} + \left(c \iint_{k|x-y|>1} |V(x)V(y)||\log k|^2\, d^2x\, d^2y \right)^{1/2}.$$

The last term can only be increased by inserting $k|x-y|$ in the integrand; this gives us

$$\|L(k)\|_{\text{H.S.}} \leqslant cM + c\|V\|_1 + ck^{1/2}|\log k\,| \times \left(\int\int |V(x)V(y)||x-y| d^2x\, d^2y \right)^{1/2}.$$ ■

Proposition 1-5: Suppose V belongs to L^2 with $\int |x||V(x)| d^2x < \infty$. Then we have

$$(V_{1/2}, [I - L(k)]^{-1}|V|^{1/2}) = a_0 + a_1 k,$$

where

$$a_0 = (V_{1/2}, [I - L(0)]^{-1}|V|^{1/2}),$$

$$a_1 = (V_{1/2}, (d/dk)[I - L(k)]^{-1}|_{k=0}\, |V|^{1/2}) + \delta(k),$$

and $\delta \to 0$ as $k \to 0$. (Note that only a_0 is independent of k.)

Proof: This follows from the definition of derivative; differentiability of $(I - L(k))^{-1}$ with respect to k can be proved as follows. First one shows that L is itself differentiable with $\|L'(k)\|$ uniformly bounded near 0. (This can be done by considering difference quotients and estimating.) It is then a straightforward generalization of the proof in Ref. 2, p. 201 to show that $(I - L)^{-1}$ is also differentiable. ■

Remark: If V satisfies the hypotheses of Proposition 1-5, then V also satisfies the hypotheses of Proposition 1-4.

2. EXCEPTIONAL POINTS AND THE FREDHOLM DETERMINANT

We now allow k to take on complex values; much of the foregoing theory is still valid for complex k. For example, we note that the operator GV is Hilbert–Schmidt in the open upper half k plane.

Proposition 2-1: Let $V \in L^2$. Then for $\mathrm{Im}\, k > 0$ the operator $G(k)V$ given by $G(k)Vf(x) = -(i/4)\int H_0(k|x-y|) \times V(y)f(y)d^2y$ is Hilbert–Schmidt, and $\|G(k)V\|_{\mathrm{H.S.}} \leqslant c|k|^{-1}$.

Proof: We apply the definition of the Hilbert–Schmidt norm to GV:

$$\|GV\|^2_{\mathrm{H.S.}} = c\int\int |H_0(k|x-y|)V(y)|^2 d^2x\, d^2y.$$

The change of variables $z = x - y$ in the inner x integral allows us to carry out the y integration to obtain

$$\begin{aligned}\|GV\|^2_{\mathrm{H.S.}} &= c\|V\|_2^2 \int |H_0(k|z|)|^2 d^2z \\ &= c\|V\|_2^2 |k|^{-2} \int |H_0(z'|k/|k||)|^2\, d^2z' \\ &\leqslant c|k|^{-2}. \end{aligned}$$

■

We recall from Sec. 1 that exceptional points are points in the complex k plane for which $I - K(k)$ is not invertible. Invertibility can be tested by calculating the modified Fredholm determinant[5] $\det_2$; examination of the Plemelj–Smithies formulas[6] shows that for $\mathrm{Im}\, k > 0$,

$$\det_2(I - K(k)) = \det_2(I - GV).$$

We denote by $D(k)$ either of the above determinants.

By the analytic Fredholm theorem,[7] $D(k)$ is therefore analytic in the upper half plane and approaches 1 as $|k|$ becomes infinite.

Analysis of the exceptional points now proceeds as in Ref. 1. Zeros of $D(k)$ in the upper half plane correspond to negative-energy bound states. The same hypotheses we have used in Lemma 1-4 in fact imply that these bound states are finite in number.[8,9] Furthermore, the degeneracy of a bound state eigenspace is equal to the multiplicity of the corresponding zero of $D(k)$.

Real zeros of $D(k)$, on the other hand, are of two types. Real exceptional points of the first kind correspond to positive-energy bound states, whose existence we rule out with the assumption $V(x) = O(|x|^{-1})$ at infinity.[10] Nonzero exceptional points of the second kind can also be ruled out; a two-dimensional version of the proof in Ref. 1 uses the assumptions $V \in L^1 \cap L^2$ and $\int |V(x)||x|^4\, d^2x < \infty$. The proof rests on the unitarity of the scattering operator, a fact that was proved by Agmon,[11] and on the proposition proved in the Appendix concerning exceptional points.

The upshot of this analysis is that we can apply the argument principle to $D(k)$ in order to calculate the number of bound states. Moreover, we may integrate along the real axis without trepidation provided we avoid the point $k = 0$.

3. THE LEVINSON THEOREM

We now know how $D(k)$ is related to the number of bound states; next we need a relation between $D(k)$ and $\det S(k)$.

Proposition 3-1: Let $V \in L^1 \cap L^2$ with no positive-energy bound states.[12] Then for positive k,

$$D(-k) = D(k)\det S(k)\exp\left((i/2)\int V(x)\, d^2x\right).$$

Proof: To prove such a relation we must find a factorization of the determinant. To do this, we begin by factoring $I - K$:

$$\begin{aligned} D(-k) &= \det_2(I - K(-k)) \\ &= \det_2\{(I - K(k)) \\ &\quad \times [I - (I - K(k))^{-1}(K(-k) - K(k))]\}. \end{aligned} \tag{3.1}$$

We recall[5] that the modified Fredholm determinant satisfies the following identity:

$$\begin{aligned}\det_2((I - A)(I - B)) &= \det_2(I - A)\det(I - B) \\ &\quad \times \exp(\mathrm{Tr}(I - A)B), \end{aligned} \tag{3.2}$$

where we have assumed B to be a trace class operator.

Since we shall show that $B = (I - K(k))^{-1}(K(-k) - K(k))$ does have finite trace, Eq. (3.1) becomes

$$D(-k) = D(k)\det(I - B)\exp[\mathrm{Tr}(I - K(k))B]. \tag{3.3}$$

We consider the last factor first. The product $(I - K(k))B$ is of course equal to $K(-k) - K(k)$, which in turn is given by

$$\begin{aligned} -\lim_{\epsilon\to 0} [&|V|^{1/2}(-\Delta - k^2 + i\epsilon)^{-1} \\ &- (-\Delta - k^2 - i\epsilon)^{-1}]V_{1/2}. \end{aligned}$$

These limits exist not only in the uniform operator topology of $B(L^{2,s}, H^{2,-s})$, as results of Agmon[13] imply, but also in the Hilbert–Schmidt norm on L^2. Stone's formula,[14] moreover, tells us

$$\begin{aligned} \lim_{\epsilon\to 0} (2\pi i)^{-1} &\int_k^b ((-\Delta - k'^2 - i\epsilon)^{-1} \\ &- (-\Delta - k'^2 + i\epsilon)^{-1})2k'\, dk' \\ &= P_{(k^2, b^2)}, \end{aligned}$$

where P denotes the spectral projection corresponding to the self-adjoint operator $-\Delta$ with domain H^2. We use Fourier inversion to represent the spectral projection as follows. For $f \in H^2$ we write

$$\begin{aligned} &(P_{(k^2,b^2)}f)(x) \\ &\quad = (2\pi)^{-2} \int_k^b \int_{S^1} \exp(ix\cdot k'\theta) \\ &\qquad \times \int \exp(-ik'\theta\cdot y)f(y)d^2y\, d\theta\, k'\, dk'. \end{aligned}$$

We eliminate the k' integral by multiplying by $1/(b-k)$ and letting b approach k. Providing that the ϵ and b limits can be interchanged, we have

$$(\pi i)^{-1}(K(-k)-K(k))f(x)$$
$$=(2\pi)^{-2}|V(x)|^{1/2}\int_{S^1}\int \exp\{ik\theta\cdot(x-y)\}V(y)f(y)d^2y\,d\theta. \tag{3.4}$$

Since the integral on the right converges absolutely, we may interchange the order of integration, which makes it easy to compute the trace:

$$\mathrm{Tr}(K(-k)-K(k))=(i/2)\int V(x)d^2x.$$

Thus the last factor of (3.3) is $\exp((i/2)\int V(x)d^2x)$.

To complete our discussion of the last factor of (3.2), it remains to show the legitimacy of interchanging the above b and ϵ limits. Our task is to show that the following expression goes to 0 as b approaches k:

$$\Big|\Big|\lim_{\epsilon\to 0}|V|^{1/2}(b-k)^{-1}\int_k^b((-\Delta-k'^2+i\epsilon)^{-1}-(-\Delta-k'^2-i\epsilon)^{-1})2k'\,dk'\,V_{1/2}$$
$$-2k\lim_{\epsilon\to 0}|V|^{1/2}((-\Delta-k^2+i\epsilon)^{-1}-(-\Delta-k^2-i\epsilon)^{-1})V_{1/2}\Big|\Big|_{\mathrm{H.S.}}.$$

We consider only the $+\epsilon$ term, the $-\epsilon$ term being similar.

$$\Big|\Big|\lim_{\epsilon\to 0}|V|^{1/2}(b-k)^{-1}\int_k^b(-\Delta-k'^2+i\epsilon)^{-1}2k'\,dk'\,V_{1/2}-2k\lim_{\epsilon\to 0}|V|^{1/2}(-\Delta-k^2+i\epsilon)^{-1}V_{1/2}\Big|\Big|$$
$$=\lim_{\epsilon\to 0}(b-k)^{-1}\int_k^b\|2k'|V|^{1/2}(-\Delta-k'^2+i\epsilon)^{-1}V_{1/2}-2k\,|V|^{1/2}(-\Delta-k^2+i\epsilon)V_{1/2}\|dk'$$
$$\leqslant\lim_{\epsilon\to 0}2\max_{(k,b)}\||V|^{1/2}(-\Delta-k'^2+i\epsilon)^{-1}V_{1/2}-|V|^{1/2}(-\Delta-k^2+i\epsilon)^{-1}V_{1/2}\|$$
$$\to 0.$$

Next we consider the second factor of (3.3), $\det(I-B)$. We use Eq. (3.4) to represent B:

$$Bf(x)=(I-K)^{-1}(i/4\pi)|V(x)|^{1/2}\int\int_{S^1}\exp\{ik\theta\cdot(x-y)\}d\theta\,V_{1/2}(y)f(y)d^2y$$
$$=(i/4\pi)\int\int|V(x)|^{1/2}\psi(k,\theta,x)\exp(-ik\theta\cdot y)d\theta\,V_{1/2}(y)f(y)d^2y.$$

We shall compare this operator to the operator $(-i/4\pi)A(k)$ defined by

$$-(i/4\pi)(Af)(\theta)=-\frac{i}{4\pi}\int_{S^1}\int\exp(-ik\theta'\cdot x)V(x)\psi(k,\theta,x)d^2x\,f(\theta')d\theta'.$$

Although these operators act on different spaces, it turns out that they have the same trace. In fact, for every integer m, we have $\mathrm{Tr}\,A^m=\mathrm{Tr}\,D^m$, where $D=-4\pi iB$.

Use of the Plemelj–Smithies formulas[15] for the Fredholm determinant now allows us to conclude that $\det(I-B)=\det(I-(i/4\pi)A)$. Thus for positive k, (3.3) is

$$D(-k)=D(k)\det S(k)\exp\Big((i/2)\int V(x)d^2x\Big).\quad\blacksquare$$

Remark: Proposition 3-1 uses only Definition (1.4) of the scattering amplitude; the relation between ψ and A stated in Proposition 1-3 is not needed.

We now turn to the Levinson theorem, which allows us to determine from scattering data the number of bound states.

Theorem 3-2 (The Levinson Theorem): Let $V\in L^2$ with $\int|x|^4|V(x)|d^2x<\infty$. Assume that V has no positive-energy bound states,[12] and assume that $(I-L(0))^{-1}$ exists. Denote by N the number of negative-energy bound states of $-\Delta+V(x)$. Then N satisfies

$$2\pi iN=\log\det S(0)-\log\det S(\infty)-(i/2)\int V(x)d^2x. \tag{3.5}$$

Proof: We apply the argument principle to $D(k)$: we integrate the logarithmic derivative along a contour $C(\epsilon,R)$ which extends along the real axis from $-R$ to $+R$, avoiding the origin by a small semicircle of radius ϵ in the upper half plane, and which closes in the upper half plane with a large semicircle of radius R.

Because in the large $|k|$ limit $D(k)$ approaches 1 and thus has unchanging argument, the integral over the large semicircle approaches 0 for large R. In the integral from $-R$ to $-\epsilon$, we use the preceding proposition:

$$2\pi iN=\int_{C(\epsilon,\infty)}d\log D(k)$$
$$=\int_{\gamma(\epsilon)}d\log D(k)+\int_\epsilon^\infty d\log D(k)$$
$$-\int_\epsilon^\infty d\log\det S(k)-\int_\epsilon^\infty d\log D(k)$$
$$-\int_\epsilon^\infty d\Big((i/2)\int V(x)d^2x\Big)$$
$$=\int_{\gamma(\epsilon)}d\log D(k)+\log\det S(\epsilon)-\log\det S(\infty).$$

We must now evaluate the $\gamma(\epsilon)$ integral. To do this we

recall from Sec. 1 that $K = L + P \log k$. We also use the Fredholm determinant multiplication formula (3.2) that appeared in the proof of the previous proposition:

$$\begin{aligned} D(k) &= \det_2(I - K(k)) \\ &= \det_2\{(I - L(k))[I - (I - L(k))^{-1} P \log k]\} \\ &= \det_2(I - L)\det(I - (I - L)^{-1} P \log k) \\ &\quad \times \exp(\mathrm{Tr}\, P \log k). \end{aligned} \tag{3.6}$$

The last factor of (3.6) is easily calculated to be $\exp(\mathrm{Tr}\, P \log k) = k^{\int V(x) d^2x/(2\pi)}$.

The second factor, which is the determinant of identity plus a rank 1 operator, is

$$\det(I - (I - L)^{-1} P \log k) = 1 - (\log k)(2\pi)^{-1}(V_{1/2}, [I - L(k)]^{-1}|V|^{1/2}).$$

We can now write the $\gamma(\epsilon)$ integral as

$$\int_{\gamma(\epsilon)} d \log D(k) = \int_{\gamma(\epsilon)} d \log \det_2(I - L(k)) + \int_{\gamma(\epsilon)} d \log\{1 - (\log k)(2\pi)^{-1}(V_{1/2}, [I - L(k)]^{-1}|V|)\} + (2\pi)^{-1} \int V(x) d^2x \int_{\gamma(\epsilon)} k^{-1}\, dk. \tag{3.7}$$

The last term of (3.7) is easily calculated to be $-(i/2)\int V(x) d^2x$. In the limit as ϵ goes to 0, the first term of (3.7) is zero because $(I - L)^{-1}$ is assumed to be well behaved near 0.

Finally we consider the second term of (3.7), whose integrand is

$$\frac{k^{-1}(V_{1/2}, [I - L(k)]^{-1}|V|^{1/2}) - (\log k)(V_{1/2}, (d/dk)[I - L(k)]^{-1}|V|^{1/2})}{2\pi - (\log k)(V_{1/2}, [I - L(k)]^{-1}|V|^{1/2})}. \tag{3.8}$$

We recall from Sec. 1 that we can write

$$(V_{1/2}, (I - L)^{-1}|V|^{1/2}) = a_0 + a_1 k,$$

where a_1 is a bounded function of k and a_0 is constant.

With this notation, we write the first term of (3.8) as

$$\frac{a_0 + a_1 k}{2\pi k - k(\log k)(a_0 + a_1 k)}.$$

If a_0 is 0, this expression is bounded and thus its contribution to the $\gamma(\epsilon)$ integral is 0 in the limit as ϵ goes to 0. If a_0 is nonzero then this expression is bounded by $c(k \log k)^{-1}$ and the following computation gives its contribution to the $\gamma(\epsilon)$ integral:

$$\int_{\gamma(\epsilon)} \frac{dk}{k \log k} = \log(1 + i\pi/\log \epsilon) \to 0 \quad \text{as } \epsilon \to 0.$$

Next we consider the second term of (3.8). If a_0 is nonzero, this term is bounded and thus gives a zero contribution to the limiting $\gamma(\epsilon)$ integral. If a_0 is 0, the second term of (3.8) is bounded by $c \log k$, which again integrates to 0 as ϵ approaches 0. ■

Remark 3-3: If V belongs to L^1 and satisfies

$$\int\int |V(x)V(y)||\log|x - y||^2\, d^2x\, d^2y < \infty,$$

then $L(0)$ is a Hilbert–Schmidt operator. Fredholm theory therefore applies to the operator $I - L(0)$, and tells us that if $(I - L(0))^{-1}$ does not exist, then the equation

$$\xi = L(0)\xi \tag{3.9}$$

has a nonzero solution. If we multiply this equation by $|V|^{1/2}$, operate on it with the Laplacian, and recall that convolution by $\log x$ is the inverse of the Laplacian, then we obtain the zero-k Schrödinger equation. This implies that the zero-k Schrödinger equation has a solution ψ with $\xi = |V|^{1/2}\psi$ belonging to L^2. If ψ itself belongs to L^2, it is a *bound state*; otherwise it is a *half-bound state*.

The assumption in Theorem 3-2 that $(I - L)^{-1}$ exists at $k = 0$ is thus equivalent to assuming the absence of zero-energy bound states and half-bound states. The work of Klaus and Simon[16] indicates that this assumption rules out consideration of λV for certain values of the coupling constant λ; these values are thresholds in the sense that new eigenvalues are born from the continuous spectrum as the coupling constant increases to these threshold values.

The assumption that there are no zero-energy bound or half-bound states could conceivably rule out consideration of other potentials as well; in fact, the conditions under which zero-energy bound and half-bound states are present and the implications of their presence have not yet been studied in two dimensions.

Note added in proof. In a recent preprint, "sum rules for few-particle scattering in two dimensions," D. Bollé, C. Danneels, and T. A. Osborn announce results that include those of this paper.

ACKNOWLEDGMENTS

I am grateful to my thesis advisor, Roger Newton, for advice, encouragement, and many helpful discussions. I would also like to thank Joe Keller for reading the manuscript and making useful suggestions.

This work was supported in part by the Air Force Office of Scientific Research, the National Science Foundation, the Office of Naval Research, and the Army Research Office.

APPENDIX: LARGE x BEHAVIOR OF ψ

Proposition 1-3: Let $V \in L^2$ with $\int |V(x)||x|^4\, d^2x < \infty$. Suppose $k > 0$ is not an exceptional point,[12,17] and let $\xi = |V|^{1/2}\psi$ solve Eq. (1.1). Then

$$\psi(k,\theta,x) = \exp(ik\theta\cdot x) + \exp(-3\pi i/4)(8\pi)^{-1/2} A(k,\hat{x},\theta) \times \exp(ik|x|)(k|x|)^{-1/2} + h(k,\theta,x),$$

where

$$A(k,\theta,\theta') = \int \exp(-ik\theta\cdot x) V(x) \psi(k,\theta',x)\, d^2x,$$

and $h(k,\theta,x)\in L^2(x)$ uniformly in θ. If $k>0$ is an exceptional point, then

$$\psi(k,x)=\exp(-3\pi i/4)(8\pi)^{-1/2}\exp(ik|x|)(k|x|)^{-1/2}\times\int\exp(-ik\hat{x}\cdot y)V(y)\psi(k,y)d^2y+h(k,x).$$

Proof: We begin with the Lippmann–Schwinger Eq. (1.1), which we write as

$$\psi(k,\theta,x)=\exp(ik\theta\cdot x)+I_1+I_2+I_3+I_4+I_5,$$

where I_1 is the piece of the integral of (1.1) corresponding to integration over the set $k|x-y|<1$, and the rest of the I's correspond to integration over the set $k|x-y|>1$. The exponential appearing in the large-argument asymptotic expansion for H_0 can be split into pieces corresponding to I_2, I_3, and I_4; I_5 is the large-argument remainder term. Thus the I's satisfy the following:

$$I_1\leqslant c\int_{k|x-y|<1}|\log k|x-y||\,|V(y)\psi(k,\theta,y)|d^2y,$$

$$I_2=\exp(-i\pi/4)(-i/4)(2/\pi)^{1/2}\exp(ik|x|)(k|x|)^{-1/2}\times\int_{k|x-y|>1}\exp(-ik\hat{x}\cdot y)V(y)\psi(k,\theta\cdot y)d^2y,$$

$$I_3=\exp(-3\pi i/4)(8\pi)^{-1/2}\times\int_{k|x-y|>1}\exp(ik|x-y|)(k|x-y|)^{-1/2}-(k|x|)^{-1/2})V(y)\psi(k,\theta,y)d^2y,$$

$$I_4=\exp(-3\pi i/4)(8\pi)^{-1/2}\times\int_{k|x-y|>1}[\exp(ik|x-y|)-\exp(ik(|x|-x\cdot y))]\times(k|x|)^{-1/2}V(y)\psi(k,\theta,y)d^2y,$$

$$I_5\leqslant c\int_{k|x-y|>1}(k|x-y|)^{-3/2}|V(y)\psi(k,\theta,y)|d^2y.$$

The claim is that $I_1+I_3+I_4+I_5$ belongs to L^2 as a function of x.

An application of Young's inequality[18] to I_1 shows that I_1 is in L^2.

We show that I_5 is in L^2 by applying the Schwarz inequality, interchanging the order of integration, and splitting the domain of x integration into parts where $|x-y|<1$ and $|x-y|>1$, respectively.

Next we calculate the L^2 norm of I_3:

$$\|I_3\|_2^2\leqslant k^{-1}\int\int_{k|x-y|>1}||x-y|^{-1/2}-|x|^{-1/2}|\times|V(y)\psi(k,\theta,y)|d^2y\times\int_{k|x-y'|>1}||x-y'|^{-1/2}-|x|^{-1/2}|\times|V(y')\psi(k,\theta,y')|d^2y'\,d^2x.$$

The positivity of the integrand allows us to do the x integral first; application of the Schwarz inequality then induces us to consider the integral

$$J=\int||x-y|^{-1/2}-|x|^{-1/2}|^2\,d^2x.$$

We make the substitutions $x=|y|z$, $y=|y|n$ with $|n|=1$:

$$J=\int||z-n|^{-1/2}-|z|^{-1/2}|^2\,d^2z=J_1+J_2,$$

where J_1 is the integral over $B(0,2)$, the disc of radius 2 centered at the origin, and J_2 is the integral over the complement. Then we have

$$J_1\leqslant\int_{B(0,2)}|z-n|^{-1}\,d^2z+2\int_{B(0,2)}(|z||z-n|)^{-1/2}\,d^2z+\int_{B(0,2)}|z|^{-1}{}^2d^2z<\infty.$$

In the integrand of J_2, first we add the fractions and then multiply numerator and denominator by both $|z|^{1/2}+|z-n|^{1/2}$ and $|z|+|z-n|$. Upon simplification, we have

$$J_2=\int_{B^c}\frac{|2z\cdot n-1|^2\,d^2z}{|z||z-n|||z|^{1/2}+|z-n|^{1/2}|^2||z|+|z-n||^2},$$

which converges. This shows that J is finite.

We return to I_3 having bounded the x integral by the constant c; an application of the Schwarz inequality then shows that $\|I_3\|_2$ is finite for $k\neq 0$.

Next we consider I_4; we calculate its L^2 norm:

$$\|I_4\|_2^2=\frac{c}{k}\int\int_{k|x-y|>1}|x|^{-1}|\exp(ik|x-y|)-\exp(ik(|x|-\hat{x}\cdot y))|\,|V(y)\psi(k,\theta,y)|d^2y\times\int_{k|x-y'|>1}|x|^{-1}\,|\exp(-ik|x-y'|)-\exp(-ik(|x|-\hat{x}\cdot y'))|\times|V(y')\,\overline{\psi(k,\theta,y')}|d^2y'\,d^2x.$$

The positivity of the integrand allows us to integrate over all space and then to do the x integral first; again an application of the Schwarz inequality induces us to estimate the integral

$$Q=\int|x|^{-1}|\exp(ik(|x-y|-|x|+\hat{x}\cdot y))-1|^2\,d^2x.$$

With the notation $y=|x|z$, $z=|x|\hat{x}$, the factor in the exponent can be written

$$g=|x|((1-2\hat{x}\cdot z+|z|^2)^{1/2}-1+\hat{x}\cdot z),$$

which behaves like $|x||z|$ for large z and like $|x||z|^2$ for small z. This gives us the estimate

$$g\leqslant c|x||z|^2(a+|z|)^{-1}=c|y|^2(a|x|+|y|)^{-1}.$$

This helps us to estimate the integrand of Q;

$$\begin{aligned}|\exp(ikg)-1|^2&=(\exp(ikg)-1)(\exp(-ikg)-1)\\&=4\sin^2(kg/2)\\&\leqslant 4|kg/2|^2(1+|kg/2|)^{-2}\\&\leqslant 4ck^2|y|^4(2a|x|+2|y|+ck|y|^2)^{-2}.\end{aligned}$$

We use this in Q, letting $x=|y|w$:

$$Q\leqslant 4c\int k^2|y|^4|w|^{-1}(2a|y||w|+2|y|+c|k||y|^2)^{-2}\,d^2w\leqslant ck^2|y|^2.$$

Using this fact to estimate I_4, we have

$$\|I_4\|_2^2\leqslant(c/k)\left(\int|V(y)\psi(k,\theta,y)|k^2|y|^2\,d^2y\right)^2<\infty.$$

We have now shown that $I_1 + I_3 + I_4 + I_5$ belongs to L^2; it remains to show that if we write $I_2 = I_6 + I_7$, where

$$I_6 = \exp(-3\pi i/4)(8\pi)^{-1/2} A(k,\hat{x},\theta) \exp(ik|x|)(k|x|)^{-1/2},$$

and

$$I_7 = -\exp(-3\pi i/4)(8\pi)^{-1/2} \times \int_{k|x-y|<1} \exp(-ik\hat{x}\cdot y) V(y)\psi(k,\theta,y) d^2y \times \exp(ik|x|)(k|x|)^{-1/2},$$

then I_7 is in L^2. To do this we write $\|I_7\|_2^2 = I_8 + I_9$, where I_8 and I_9 correspond to x integration over the unit ball $B(0,1)$ and its complement, respectively. The integral I_8 converges by the Schwarz inequality. Application of the Schwarz inequality to I_9 gives us

$$|I_9| \leqslant k^{-1} \int_{R^2 \setminus B(0,1)} \|\xi(k)\|_2^2 \int_{k|x-y|<1} |V(y)| d^2y\, d^2x;$$

we then let $z = x - y$ in the inner integral, change the order of integration, and carry out the x integration. This gives us

$$|I_9| \leqslant \|\xi\|_2^2 k^{-1} \int_{B(0,k^{-1})} \|V\|_1 dz < \infty. \quad \blacksquare$$

[1] R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I. Functional Analysis* (Academic, New York, 1972), p. 210.
[3] G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1970), pp. 504, 518.
[4] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton University, Princeton, 1971), p. 9.
[5] B. Simon, *Trace Ideals and Their Applications* (Cambridge University, Cambridge, England, 1979), p. 107.
[6] B. Simon, *Trace Ideals and Their Applications* (Cambridge University, Cambridge, England, 1979), p. 108.
[7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I. Functional Analysis* (Academic, New York, 1972), p. 201.
[8] R. G. Newton, J. Operator Theory **10**, 119 (1983).
[9] B. Simon, Ann. Phys. **97**, 279 (1976).
[10] T. Kato, Commun. Pure Appl. Math. **12**, 403 (1959).
[11] S. Agmon, Ann. Scuola Norm. Sup. Pisa IV **2**, 151 (1975).
[12] The absence of positive-energy bound states is assured when, for example, $V(x) = O(|x|^{-1})$ at infinity.
[13] S. Agmon, Ann. Scuola Norm. Sup. Pisa IV **2**, 151 (1975).
[14] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I. Functional Analysis* (Academic, New York, 1972), p. 237.
[15] B. Simon, *Trace Ideals and Their Applications* (Cambridge University, Cambridge, England, 1979), p. 68.
[16] M. Klaus and B. Simon, Ann. Phys. **130**, 251 (1980).
[17] There are no positive k exceptional points of the second kind under the hypotheses of Proposition 1-3. The additional assumption $V(x) = O(|x|^{-1})$ at infinity therefore implies that the conclusion of Proposition 1-3 holds for all $k > 0$.
[18] M. Reed and B. Simon, *Methods of Modern Mathematical Physics II. Fourier Analysis and Self-Adjointness* (Academic, New York, 1975), p. 32.

Some magnetohydrostatic models of cylindrical symmetry in general relativity

S. R. Roy
Department of Mathematics, Banares Hindu University, Varanashi, India
Raj Bali
Department of Mathematics, University of Rajasthan, Jaipur, India

(Received 15 June 1981; accepted for publication 19 August 1983)

The purpose of this paper is to study the problem of equilibrium of a cylindrically symmetric distribution of perfect fluid under the action of an incident magnetic field. The magnetic field is due to an electric current along the axis of symmetry surrounded by static fluid of infinite electric conductivity. Two models have been derived and their physical features have been discussed.

PACS numbers: 04.20. − q, 04.30. + x

1. INTRODUCTION

Static universes play an important role in the understanding of phenomena of cosmological and astrophysical significance. Static cylindrically symmetric space-times representing material distribution were obtained by Marder.[1] It is known that a strong magnetic field is present in galactic and intergalactic spaces and is significant even in the cosmological scale. This raises the interesting question about the possibility of static configurations of perfect fluid distribution in equilibrium with the magnetic field. Magnetostatic fields of cylindrical symmetry were considered by Witten,[2] Melvin,[3,4] Thorne,[5,6] and Safko and Witten.[7] A magnetic universe with matter has been studied by Khalatnikov.[8] Patel and Vaidya[9] have derived cosmological models describing a static magnetic universe. Safko and Witten[10] have derived static models of matter with different internal and external magnetic fields. Prakash and Roy[11] have recently obtained a cosmological model of cylindrical symmetry with incident magnetic field which is an inhomogeneous generalization of the Einstein universe.

In the present paper we have derived some magnetohydrostatic models of cylindrical symmetry in which the free gravitational field is of Petrov type D. The magnetic flux vector is assumed to be in the azimuthal direction. Einstein's field equations give rise to two models. In the second section we give the derivation of these two models. In the third section we discuss the physical properties of these models.

2. DERIVATION OF THE CYLINDRICALLY SYMMETRIC MODELS

We consider the cylindrically symmetric metric in the form

$$ds^2 = A^2(d\rho^2 - dt^2) + B^2\,dz^2 + C^2\,d\phi^2\,, \tag{2.1}$$

in which A, B, and C are functions of ρ alone. The distribution consists of an electrically neutral perfect fluid with an infinite electrical conductivity and magnetic field. The energy-momentum tensor of the composite field is assumed to be the sum of the corresponding energy-momentum tensors:

$$T_i^j = (\epsilon + p)v_i v^j + p g_i^j + E_i^j\,, \tag{2.2}$$

where the electromagnetic energy momentum tensor E_i^j is given by[12]

$$E_i^j = h_1 h^{-1}(v_i v^j + \tfrac{1}{2}g_i^j) - h_i h^j\,. \tag{2.3}$$

In the above ϵ, p are the proper density and pressure of the fluid and v^i is the unit velocity vector; μ is the magnetic permeability and h_i is the magnetic flux vector defined in terms of the electromagnetic field tensor F_{ij} by

$$h_i = (1/2\mu)\sqrt{-g}\,\epsilon_{ijkl}F^{kl}v^j\,. \tag{2.4}$$

The nonvanishing component of v^i is assumed to be $v^4 = 1/A$. Since the field is purely magnetic the components F_{14}, F_{24}, F_{34} are assumed to be zero, the coordinates x^1, x^2, x^3, x^4 being the same as ρ, z, ϕ, and t. The magnetic field is assumed to be in the azimuthal direction so that F_{12} is the only nonvanishing component of F_{ij}. Maxwell's equations

$$F_{(ij;k)} = 0\,, \tag{2.5}$$

$$\left(\frac{1}{\mu}F^{ij}\right)_{;j} = 0 \tag{2.6}$$

require that F_{12} be a function of ρ alone and that $F_{12}C/B$ be a constant, say H so that

$$h_3 = H/\mu A\,.$$

The field equations

$$R_i^{\ j} - \tfrac{1}{2}Rg_i^{\ j} - \Lambda g_i^{\ j} = -8\pi T_i^{\ j} \tag{2.7}$$

lead to the equation

$$\left(\frac{A_{,1}}{A}\right)_{,1} + \frac{C_{,11}}{C} - \frac{A_{,1}B_{,1}}{AB} - \frac{A_{,1}C_{,1}}{AC} - \frac{B_{,1}C_{,1}}{BC} = 0\,, \tag{2.8}$$

$$\frac{B_{,11}}{B} - \frac{C_{,11}}{C} + \frac{8\pi H^2}{\mu C^2} = 0\,, \tag{2.9}$$

$$8\pi p = \frac{1}{A^2}\left(\frac{A_{,1}B_{,1}}{AB} + \frac{A_{,1}C_{,1}}{AC} + \frac{B_{,1}C_{,1}}{BC} - \frac{4\pi H^2}{\mu C^2}\right) + \Lambda\,, \tag{2.10}$$

$$8\pi\epsilon = \frac{1}{A^2}\left(\frac{A_{,1}B_{,1}}{AB} + \frac{A_{,1}C_{,1}}{AC} - \frac{B_{,1}C_{,1}}{BC} - \frac{B_{,11}}{B} - \frac{C_{,11}}{C} - \frac{4\pi H^2}{\mu C^2}\right) - \Lambda\,. \tag{2.11}$$

These are four equations in five unknowns A, B, C, ϵ, and p. To get a determinate solution we require one additional con-

 0022-2488/84/051456-04$02.50

dition. We assume the free gravitational field to be Petrov type D, the degeneracy being in the axial plane. This requires that $C^{23}_{23} = C^{31}_{31}$, where C_{hijk} is the conformal curvature tensor. Thus

$$\left(\frac{A_{,1}}{A}\right)_{,1} - \frac{C_{,11}}{C} + \frac{A_{,1}C_{,1}}{AC} + \frac{B_{,1}C_{,1}}{BC} - \frac{A_{,1}B_{,1}}{AB} = 0. \tag{2.12}$$

The above equations give rise to two cases.

Case I: $A_1/A \neq 0$

In this case Eqs. (2.8), (2.9), and (2.12) lead to the solution

$$A = MC + q, \quad B = LC_{,1}/(MC + q)$$

with C determined from the equation

$$\frac{d}{d\rho}\left[\frac{C(C_{,11})}{MC+q} - \frac{MC(C_{,1})^2}{(MC+q)^2} - \frac{(C_{,1})^2}{MC+q}\right] + \frac{8\pi H^2}{\mu C(MC+q)} = 0, \tag{2.13}$$

L, M, and q being constants. Equation (2.13) on integration gives

$$\begin{aligned}\left(\frac{dC}{d\rho}\right)^2 = C^2(MC+q)^2\Bigg[&- \frac{16\pi H^2M^2}{\mu q^4}\Bigg\{\frac{1}{2}\left(\frac{C+q/M}{C}\right)^2 \log\frac{C+q/M}{C} \\ &- \frac{1}{4}\left(\frac{C+q/M}{C}\right)^2 - 2\left(\frac{C+q/M}{C}\right)\log\frac{C+q/M}{C} + 2\left(\frac{C+q/M}{C}\right) \\ &+ \frac{1}{2}\left(\log\frac{C+q/M}{C}\right)^2 - \frac{7}{4}\Bigg\} + \frac{P}{q^3}\Bigg\{-\frac{1}{2}\left(\frac{C+q/M}{C}\right)^2 \\ &+ 2\left(\frac{C+q/M}{C}\right) - \log\frac{C+q/M}{C} - \frac{3}{2}\Bigg\} + Q\Bigg],\end{aligned} \tag{2.14}$$

where P and Q are constants.

After suitable transformation of coordinates and renaming of the constants the metric takes the form

$$ds^2 = \frac{d\xi^2}{\xi^2 R^2(\xi)} + \xi^2 R^2(\xi) dZ^2 + \xi^2 d\phi^2 - (\xi+b)^2 dt^2, \tag{2.15}$$

where

$$\begin{aligned}R^2(\xi) = \frac{2K}{b^4}\Bigg[&\frac{1}{2}\left(\frac{\xi+b}{\xi}\right)^2 - 2\left(\frac{\xi+b}{\xi}\right)\log\frac{\xi+b}{\xi} - \frac{1}{4}\left(\frac{\xi+b}{\xi}\right)^2 + 2\left(\frac{\xi+b}{\xi}\right) \\ &+ \frac{1}{2}\left(\log\frac{\xi+b}{\xi}\right)^2 - \frac{7}{4}\Bigg] + \frac{k}{b^3}\Bigg\{-\frac{1}{2}\left(\frac{\xi+b}{\xi}\right)^2 + 2\left(\frac{\xi+b}{\xi}\right) - \log\frac{\xi+b}{\xi} - \frac{3}{2}\Bigg\} + Q.\end{aligned} \tag{2.16}$$

In the above k can take the values 0, 1, or -1.

Case II: $A_1 = 0$

In this case we have the solution

$$A = \beta, \quad B = (\alpha C^2 + (8\pi H^2/\mu)\gamma^2 \log C + \delta)^{1/2}$$

with C determined from the equation

$$\frac{dC}{d\rho} = \frac{1}{\gamma}\left(\alpha C^2 + \frac{8\pi H^2}{\mu}\gamma^2 \log C + \delta\right)^{1/2},$$

α, β, γ, and δ being constants. After suitable transformation of coordinates and renaming of the constants the metric takes the form

$$ds^2 = d\xi^2/\psi(\xi) + \psi(\xi) dZ^2 + \xi^2 d\phi^2 - dT^2, \tag{2.17}$$

where

$$\psi(\xi) = 1 - \xi^2/\bar{\rho}^2 - K\log\xi. \tag{2.18}$$

3. PHYSICAL FEATURES OF THE MODELS

Case I: The first model

The pressure and the density for the model (2.15) are given by

$$\begin{aligned}8\pi p = \frac{3\xi+b}{\xi+b}R^2(\xi) + \frac{2\xi+b}{\xi^2(\xi+b)^2}\left(\frac{K}{b}\log\frac{\xi}{\xi+b} + \frac{k}{2}\right) \\ + \frac{K}{2\xi^2(\xi+b)^2} + \Lambda,\end{aligned} \tag{3.1}$$

$$\begin{aligned}8\pi\epsilon = -3R^2(\xi) - \frac{2\xi+3b}{\xi^2(\xi+b)^2}\left(\frac{K}{b}\log\frac{\xi}{\xi+b} + \frac{k}{2}\right) \\ - \frac{K}{2\xi^2(\xi+b)^2} - \Lambda.\end{aligned} \tag{3.2}$$

The nonvanishing component F_{12} of the tensor F_{ij} is given by

$$F_{12} = H/\xi^2(\xi+b)^2, \tag{3.3}$$

where the constant H is related to K by the equation

$$K = -8\pi H^2/\mu.$$

In the above the coordinates x^1, x^2, x^3, x^4 stand for

ρ, Z, ϕ, and T.

As $\xi \to \infty$, $8\pi p \to 3Q + \Lambda$ and $8\pi\epsilon \to -(3Q + \Lambda)$. Also C_{hijk} vanishes asymptotically. We set $3Q + \Lambda$ equal to zero. In order to understand the significance of the constant b, we note that when $K = k = 0$, it gives rise to an unrealistic model with zero density and nonzero pressure. Further, with $K \neq 0$, $k = 0$, the density arises due to the magnetic field alone. Owing to this unsatisfactory feature, we shall assume that for $b \neq 0$, $k = \pm 1$.

When b is zero, both density and pressure vanish. In this case the line element takes the form

$$dS^2 = \frac{\xi^2 d\xi^2}{\frac{1}{2}K - \frac{1}{3}k\xi + Q\xi^4}$$

$$+\frac{\frac{1}{2}K-\frac{1}{3}k\xi+Q\xi^4}{\xi^2}dZ^2+\xi^2\,d\phi^2-\xi^2\,dT^2\,. \quad (3.4)$$

The following cases arise:

(a) $Q=0$, $k=-1$. The solution is a special case of Witten's solution representing the external field due to an infinite wire carrying an electric current and having gravitational mass per unit length equal to 8π.

(b) $Q<0$, $k=-1$. The metric (3.4) has the proper signature when $K^3Q<\frac{1}{96}$, and in that case, the metric is valid in the region $\xi_1<\xi<\xi_2$ where ξ_1 and ξ_2 are the positive roots of the equation $\frac{1}{2}K+\frac{1}{3}\xi+Q\xi^4=0$. In the absence of magnetic field, we get an empty space valid in $0<\xi<(1/3Q)^{1/3}$.

(c) $Q>0$. The metric is valid in the region $\xi>\xi_1>0$ where ξ_1 is the positive root of the equation $\frac{1}{2}K-\frac{1}{3}k\xi+Q\xi^4=0$. In the absence of the magnetic field, there are three possibilities corresponding to the three values of k. These are

$k=-1$. The metric is valid in $\xi>0$.

$k=+1$. The metric is valid in $\xi>(1/3Q)^{1/3}$.

$k=0$. The metric is that of anti-de Sitter space-time. A study of the scalar invariants (see Appendix) shows that these are finite everywhere in case (a). In case (b), the invariants are finite at both $\xi=\xi_1$ and $\xi=\xi_2$, while in case (c) they are finite at $\xi=\xi_1$. In this sense, the above solution has no singularity anywhere in the region in which it is valid. Real singularities appear at $\xi=0$ in the absence of the magnetic field in all the cases except when $k=0$ or 1 in case (c). One can define a new coordinate $r^2=\xi-\xi_1$ in both the cases (b) and (c) and take the axis as $r=0$. It is easy to see that the elementary flatness condition is not satisfied on the axis. The hypersurface $\xi=\xi_2$ in case (b) is null and forms a horizon. Each of the models (a), (b), and (c) represents the field due to an electric current flowing along the axis which also carries a distribution of matter.

For a timelike radial geodesic, we have

$$\frac{dt}{ds}=\frac{E}{(r^2+\xi_1)^2},$$

$$\frac{dr}{ds}=\pm\frac{\sqrt{E^2-(r^2+\xi_1)^2}\sqrt{\frac{1}{2}K-\frac{1}{3}k(r^2+\xi_1)+Q(r^2+\xi_1)^4}}{(\xi_1+r^2)^2r},$$

E being a constant. In case (b), the velocity vanishes at the horizon at which there is repulsion. It is clear that a pair of such geodesics may cross each other twice at the most. It takes finite proper time for a neutral particle to reach the axis. Similarly the proper time to reach the horizon radially is finite. When $b\neq0$ and $K=0$, the signature of the metric (2.16) remains valid for $0\leqslant\xi<\infty$ provided $Q\geqslant0$, $k=-1$, $b>0$. There exists a real singularity at $\xi=0$. Assuming b to be small and $Q=0$, we find that the density and pressure are given by

$$\epsilon=3p=b/32\xi^4.$$

Thus, up to the first order in b, the matter behaves as disordered radiation.

When $b\neq0$, $K\neq0$, we find that $R^2(\xi)\to-\infty$ as $\xi\to0$ and $R^2(\xi)\to Q$ as $\xi\to\infty$. If $Q>0$, there exists a positive root $\xi=\xi_1$ of $R^2(\xi)$. In this case the metric is valid for $\xi_1<\xi<\infty$. The invariants are finite at $\xi=\xi_1$ and there are no apparent singularities (see Appendix). Defining the coordinates by $\xi-\xi_1=r^2$, we find that there is lack of elementary flatness at the axis $r=0$. Equation (3.3) tells us that the magnetic field is transverse and is due to an electric current along the axis. Following Witten, we find the total current flowing along the axis to be $2\pi H/\mu$. The lack of elementary flatness shows that there is matter distribution along the axis. Density and pressure are finite everywhere including the axis. The model therefore, represents a wire of nonvanishing mass density carrying an electric current and is surrounded by a perfect fluid.

Case II: The second model

The pressure and density for the model (2.17) are given by

$$8\pi p=\Lambda-1/\bar{\rho}^2, \quad (3.5)$$

$$8\pi\epsilon=3/\bar{\rho}^2+K/\xi^2-\Lambda. \quad (3.6)$$

The nonvanishing component F_{12} of the tensor F_{ij} is given by

$$F_{12}=H/\xi, \quad (3.7)$$

where

$$K=-8\pi H^2/\mu.$$

The reality conditions $\epsilon\geqslant p>0$ require that

$$\frac{K}{2\xi^2}+\frac{2}{\bar{\rho}^2}>\Lambda>\frac{1}{\bar{\rho}^2}. \quad (3.8)$$

The function $\psi(\xi)=1-\xi^2/\bar{\rho}^2-K\log\xi$ has a maximum at $\xi=\xi m=(-K\bar{\rho}^2/2)^{1/2}$. The model will be realistic when

$$\psi(\xi_m)=1+\tfrac{1}{2}K-\tfrac{1}{2}K\log(-\frac{K\bar{\rho}^2}{2})>0. \quad (3.9)$$

In this case $\psi(\xi)$ vanishes when $\xi=\xi_1$ and $\xi=\xi_2$ $(0<\xi_1<\xi_2)$. A glance at the scalar invariants (see Apendix) shows that there is no singularity at $\xi=\xi_1$ and $\xi=\xi_2$. Putting $\xi-\xi_1=r^2$, we find that the elementary flatness condition is not satisfied on the axis $r=0$. The hypersurface $\xi=\xi_2$ is null. It is a horizon of the model. In this case also, the magnetic field is due to an electric current along the axis of strength $2\pi H/\mu$ with a nonvanishing matter distribution along the axis represented by the absence of elementary flatness there.

For a timelike radial geodesic, we have

$$\frac{dT}{ds}=E,$$

$$\frac{dr}{ds}=\pm\frac{\sqrt{E^2-1}}{2r}\sqrt{1-\frac{(r^2+\xi_1)^2}{\bar{\rho}^2}-K\log(r^2+\xi_1)},$$

E being a constant. The velocity vanishes at the horizon at which there is repulsion. A pair of such geodesics may cross each other twice at the most. A similar result holds for null geodesics. A neutral particle moving radically reaches the axis in finite proper time. Similarly the horizon is reached in finite proper time.

In the absence of the magnetic field, the metric (2.17) reduces to the form

$$dS^2 = \frac{d\xi^2}{1-\xi^2/\bar{\rho}^2} + \left(1 - \frac{\xi^2}{\bar{\rho}^2}\right)dZ^2 + \xi^2 d\phi^2 - dT^2 \tag{3.10}$$

for which the pressure and density are given by

$$8\pi p = \Lambda - \frac{1}{\bar{\rho}^2}, \tag{3.11}$$

$$8\pi\epsilon = \frac{3}{\bar{\rho}^2} - \Lambda. \tag{3.12}$$

The metric (3.10) is that of an Einstein universe. The effect of the magnetic field is to introduce inhomogeneity in the density while the pressure remains homogeneous. It also introduces singularity on the axis which shows that a current is possible only when there is a nonvanishing distribution of matter along the axis.

ACKNOWLEDGMENT

The authors are grateful to the referee for his valuable comments and suggestions.

APPENDIX

Taking the orthogonal tetrad in the direction of coordinate axes, the physical components of C_{hijk} and R_{ij} are defined by

$$C_{(abcd)} = C_{hijk}\lambda^{h}{}_{(a)}\lambda^{i}{}_{(b)}\lambda^{j}{}_{(c)}\lambda^{k}{}_{(d)},$$
$$R_{(ab)} = R_{ij}\lambda^{i}{}_{(a)}\lambda^{j}{}_{(b)}.$$

The nonvanishing components of these for the two models are given below.

I. First model

$$C_{(2323)} = C_{(1313)} = -\frac{1}{2}C_{(1212)} = \frac{1}{6(\xi+b)^2}\left[-\frac{2K}{b\xi}\log\frac{\xi+b}{\xi} - \frac{K}{\xi^2} + \frac{k}{\xi}\right],$$

$$R_{(11)} = \xi^2 R^2(\xi)\left\{\frac{1}{R}\frac{d^2R}{d\xi^2} + \frac{2}{\xi R}\frac{dR}{d\xi} + \left(\frac{1}{\xi} + \frac{1}{R}\frac{dR}{d\xi}\right)\left(\frac{2}{\xi}\frac{1}{\xi+b} + \frac{1}{R}\frac{dR}{d\xi}\right)\right],$$

$$R_{(22)} = \xi^2 R^2(\xi)\left[\frac{1}{R}\frac{d^2R}{d\xi^2} - \frac{1}{R}\left(\frac{dR}{d\xi}\right)^2 + \frac{1}{R}\frac{dR}{d\xi}\left(\frac{1}{\xi} + \frac{1}{\xi+b}\right) + \frac{1}{\xi(\xi+b)} + \left(\frac{1}{\xi} + \frac{1}{R}\frac{dR}{d\xi}\right)\times\left\{2\xi R^2 + 2\xi^2 R\left(2 - \frac{dR}{d\xi}\right)\right\}\right],$$

$$R_{(33)} = 2R^2 + 2\xi R\frac{dR}{d\xi} + \frac{\xi R^2}{\xi+b},$$

$$R_{(44)} = -\frac{2\xi^2 R}{\xi+b}\frac{dR}{d\xi} - \frac{\xi R^2(3b+R\xi)}{(\xi+b)^2},$$

$$\lambda^{i}{}_{(a)} = \mathrm{Diag}\left\{\xi R(\xi), \frac{1}{\xi R(\xi)}, \frac{1}{\xi}, \frac{1}{\xi+b}\right\}.$$

II. Second model

$$C_{(2323)} = C_{(1313)} = -\tfrac{1}{2}C_{(1212)} = -K/\xi^2,$$
$$R_{(11)} = -2/\bar{\rho}^2,$$
$$R_{(22)} = -2/\bar{\rho}^2,$$
$$R_{(33)} = -(2/\bar{\rho}^2 + K/\xi^2),$$
$$R_{(44)} = 0,$$
$$\lambda^{i}{}_{(a)} = \mathrm{Diag}\{\psi^{1/2}(\xi), \psi^{-1/2}(\xi), 1/\xi, 1\}.$$

The scalar invariants can be expressed in terms of these quantities.

[1] L. Marder, Proc. Roy. Soc. London, Ser. A **244**, 524 (1958).
[2] L. Witten, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 382.
[3] M. A. Melvin, Phys. Lett. **8**, 65 (1964).
[4] M. A. Melvin, Phys. Rev. **139**, B225 (1965).
[5] K. S. Thorne, Phys. Rev. **138**, B251 (1965).
[6] K. S. Thorne, Phys. Rev. **139**, B244 (1965).
[7] J. L. Safko and L. Witten, J. Math. Phys. **12**, 257 (1971).
[8] I. M. Khalatnikov, Zh. Eksp. Teor. Fiz. Pis'ma **5**(6), 195 (1967) [JETP Lett. **5**, 155 (1967)].
[9] L. K. Patel and P. C. Vaidya, Curr. Sci. **40**, (11), 288 (1971).
[10] J. L. Safko and L. Witten, Phys. Rev. D **5**, 293 (1972).
[11] S. Prakash and S. R. Roy, J. Phys. A **13**, 3773 (1980).
[12] A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967), p 93.

A class of shear-free perfect fluids in general relativity. II

C. B. Collins
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
A. J. White
Materials and Mechanics Branch, Atomic Energy of Canada Limited, Whiteshell Nuclear Research Establishment, Pinawa, Manitoba R0E 1L0, Canada

(Received 26 September 1983; accepted for publication 16 December 1983)

We continue our previous investigation of shear-free perfect fluids in general relativity, under the assumptions that the fluid satisfies an equation of state $p = p(\mu)$ with $\mu + p \neq 0$, and that the vorticity and acceleration of the fluid are parallel (and possibly zero). We classify algebraically the set of such solutions into thirteen invariant nonempty cases. In each case, we investigate the allowed isometry groups and Petrov types, and invariantly characterize the special subcases that arise. We also show how the various subcases are related to each other and to the works of previous authors.

PACS numbers: 04.20.Cv, 02.40. + m

1. INTRODUCTION

In this article, we continue our previous investigation (Ref. 1; hereinafter referred to as I) of shear-free perfect fluids in general relativity, under the assumptions that the fluid satisfies an equation of state $p = p(\mu)$ with $\mu + p \neq 0$, and that the vorticity (ω) and acceleration ($\dot{u}$) of the fluid are parallel. Our notation, units, and conventions follow our previous work, with which a familiarity will be assumed, and our previous results can be summarized in the following theorems (labeled as in I):

Theorem 3.1: Consider a shear-free perfect fluid in general relativity, with an equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$. Suppose that the vorticity and acceleration are non-zero and parallel. Then the fluid's volume expansion is zero.

Theorem 3.2 (cf. Theorem 5.1, Ellis[2]): Consider a shear-free perfect fluid in general relativity, in which the fluid flow is geodesic and in which $\mu + p \neq 0$. Suppose that the vorticity is nonzero. Then the fluid's volume expansion scalar vanishes.

These two theorems can be combined, to give

Theorem 3.3: Any shear-free perfect fluid in general relativity with an equation of state $p = p(\mu)$, such that $\mu + p \neq 0$, has either vanishing vorticity or vanishing expansion (i.e., $\sigma = 0 \Rightarrow \omega\theta = 0$), provided that the vorticity and acceleration are parallel (and possibly zero).

The result of Theorem 3.3, together with analogous results by several authors (see references cited in I), suggests the validity of the following conjecture:

Conjecture: Any shear-free perfect fluid in general relativity with an equation of state $p = p(\mu)$, such that $\mu + p \neq 0$, has either vanishing vorticity or vanishing expansion, i.e., $\sigma = 0 \Rightarrow \omega\theta = 0$.

As previously observed (I), the case in which both the acceleration and vorticity vanish is well-understood, giving rise to the spatially homogeneous and isotropic Friedmann–Robertson–Walker (FRW) models,[3] and in particular to the generalized Einstein static model when the expansion also vanishes. We shall therefore ignore this case, and assume that $\dot{\mathbf{u}}^2 + \boldsymbol{\omega}^2 \neq 0$. If, as seems plausible, the above Conjecture is true, then the following situations are therefore of interest:

(i) $\sigma = 0$, $\boldsymbol{\omega} = \mathbf{0}$, $\theta \neq 0$, $\dot{\mathbf{u}} \neq \mathbf{0}$,

(ii) $\sigma = \theta = 0$, $\boldsymbol{\omega} = \mathbf{0}$, $\dot{\mathbf{u}} \neq \mathbf{0}$,

and

(iii) $\sigma = \theta = 0$, $\boldsymbol{\omega} \neq \mathbf{0}$,

where $\dot{\mathbf{u}}$ is possibly, but not necessarily, zero in case (iii). If the Conjecture were false, there would of course be a fourth possibility

(iv) $\sigma = 0$, $\boldsymbol{\omega}\theta \neq \mathbf{0}$, $\dot{\mathbf{u}} \neq \mathbf{0}$

(note that $\dot{\mathbf{u}} \neq \mathbf{0}$ in case (iv), in view of Theorem 3.2).

Case (i) is well understood, since it was recently treated by Collins and Wainwright.[4] It was shown that the only possibilities are the spherically symmetric Wyman[5] metrics, or a class of new exact solutions that are plane symmetric and hypersurface homogeneous. These results can be considered as generalizations of the works of Mansouri,[6] Glass,[7] Mashhoon and Partovi,[8] and Srivastava and Prasad,[9] who investigated case (i) solutions assuming spherical symmetry at the outset. As discussed by Collins and Wainwright,[4] the existence of the class of new exact solutions shows that Theorem 5.1 and Theorem 5.2 of King and Ellis,[10] wherein it is claimed that there are no tilted spatially homogeneous shear-free perfect fluids, and deduced that the only spatially homogeneous shear-free perfect fluid models are FRW, are not valid. The technique of Collins and Wainwright[4] involved assuming the coordinate approach and results of Barnes,[11] who investigated shear-free perfect fluids with $\boldsymbol{\omega} = \mathbf{0}$, and then the conditions relating to the existence of an equation of state and to the assumption $\theta \neq 0$ were imposed. It was first discovered that the solutions were necessarily "locally rotationally symmetric",[2,12] and the two classes of models were then determined. We shall discuss these results from the somewhat more satisfactory standpoint of orthonormal tetrads, in which the geometrical properties are more transparently displayed. Moreover, we find that the discussion of cases (ii) and (iii) shares some common feature with that of case (i), and we therefore commence our analysis by providing a fairly uniform treatment of all three cases. This has the merit of indicating clearly both the common and the divergent features of the solutions in the different cases.

While the solutions are known completely in case (i), this is not the situation for cases (ii), (iii), and (iv). In case (ii),

 0022-2488/84/051460-13$02.50

the flow is static and irrotational, but, as far as we are aware, not all solutions are known in this case (the article by Barnes[11] involves all such solutions which are algebraically degenerate and which do not necessarily possess an equation of state). In case (iii), the situation is similar; there are some known exact solutions, perhaps the most familiar being due to Gödel,[13] in which $\dot{\mathbf{u}}\equiv\mathbf{0}$, although there are others given in the literature,[2,12,14] in which the acceleration is not necessarily zero.

In the present article, we shall consider all cases, but in case (iii) we shall impose the additional restriction that the acceleration, if nonzero, be parallel to the vorticity. It then follows from Theorem 3.3 that all shear-free perfect fluids in general relativity with an equation of state $p=p(\mu)$, such that $\mu+p\neq 0$, and with the acceleration and vorticity parallel (and possibly zero), are then encompassed by our discussion of cases (i), (ii), and (iii), i.e., that case (iv) is disallowed. Although we still cannot obtain all exact solutions in cases (ii) and (iii), we are able to derive a set of tetrad equations which is self-consistent. We then investigate the allowed isometry groups and Petrov types for such space-times, both in the most general case, and in all various special cases, and so we feel that some progress has been made in understanding the nature of the solutions.

The plan of this article is as follows. In Sec. 2, we obtain some preliminary results that are of assistance in the compilation of a "specialization diagram," which shows how various subclasses (particularly those involving the solutions given by other authors) are interrelated. In Sec. 3, we present a technical discussion of the isometry groups and Petrov classification of the solutions. Section 4 concludes with miscellaneous comments and with ideas for further study. Throughout we shall use the tetrad constructed in Sec. 2 of I, and at first we shall refer to the Jacobi identities, Einstein field equations, Bianchi identities, and commutation relations as listed in the Appendix of I, but incorporating the result of Theorem 3.3, viz., that $\omega\theta\equiv 0$. After obtaining Proposition 2.1, we incorporate its results, viz., that $n=0$ and $\partial_\alpha\theta=0$, and rewrite these equations in Appendix A. As in our previous work, in the case $\dot{u}\neq 0$ we shall need to refer to the result of applying the commutation relations [(A28)–(A33) of I] to the function F, defined by

$$F(\mu):=-\int\frac{p'}{\mu+p}\,d\mu,$$

where $p'=dp/d\mu$ (apart from a multiplicative constant, e^F is then the thermodynamic "enthalpy" of the fluid). We thereby obtain

$$\partial_0\dot{u}=-G(\mu)\dot{u}\theta, \tag{1.1}$$

$$\partial_2\theta=\partial_3\theta=0, \tag{1.2}$$

$$\partial_2\dot{u}=d_2\dot{u},\quad \partial_3\dot{u}=d_3\dot{u}, \tag{1.3}$$

and

$$n=0. \tag{1.4}$$

Use has been made of Theorem 3.1 and the Bianchi identity (A24) of I in deriving (1.4), and (1.1) has been simplified using (1.4) and the (01) field equation (A15) of I, together with the definition

$$G(\mu):=(p''/p')(\mu+p)-p'+\tfrac{1}{3}, \tag{1.5}$$

where a prime (′) denotes differentiation with respect to μ, and $p'\neq 0$ by the Bianchi identity (A25) of I.

2. PRELIMINARY RESULTS AND A SPECIALIZATION DIAGRAM

The following five propositions pertain to shear-free perfect fluids in general relativity with an equation of state $p=p(\mu)$, such that $\mu+p\neq 0$, and with the vorticity and acceleration parallel (and possibly zero). The tetrad is that defined in Sec. 2 of I, and we assume that $\dot{u}^2+\omega^2\neq 0$ throughout.

Proposition 2.1: $n=0$ and $\partial_\alpha\theta=0$ $(\alpha=1,2,3)$.

Proof: If $\dot{u}\neq 0$, we have $n=0$ and $\partial_2\theta=\partial_3\theta=0$ by Eqs. (1.2) and (1.4), and then the (01) field equation (A15) of I requires that $\partial_1\theta=0$. On the other hand, if $\dot{u}\equiv 0$, then, by assumption, $\omega\neq 0$, so by Theorem 3.2, $\theta\equiv 0$, and the (01) field equation (A15) of I then implies that $n=0$. Hence in either case the proposition is proved. □

In view of Proposition 2.1, the basic tetrad equations are rewritten in Appendix A of the present article.

Proposition 2.2: If $\dot{u}\neq 0$, then $(d_2^2+d_3^2)\theta\equiv 0$.

Proof: We show first that if $\dot{u}\neq 0$, it follows that

$$(d_2^2+d_3^2)\theta\,[(G'/p')(\mu+p)-2G+1]=0, \tag{2.1}$$

and then deduce that the possibility $(d_2^2+d_3^2)\theta\neq 0$ is untenable.

Let the function H be defined by

$$H:=\partial_1\dot{u}+\dot{u}^2+\dot{u}\hat{\theta}-\tfrac{1}{2}(\mu+3p-2\Lambda), \tag{2.2}$$

so that the (00) field equation (A13) becomes

$$\partial_0\theta=-\tfrac{1}{3}\theta^2+2\omega^2+H. \tag{2.3}$$

Applying the $[\mathbf{e}_0,\mathbf{e}_\alpha]$ commutators (A27) and (A28) to θ, and using Proposition 2.1 with Eqs. (2.3), (A14), and (A15), we obtain

$$\partial_2 H=-4d_2\omega^2 \tag{2.4}$$

and

$$\partial_3 H=-4d_3\omega^2, \tag{2.5}$$

whereas, applying the $[\mathbf{e}_0,\mathbf{e}_1]$ commutator (A26) to $\dot{u}$, we have

$$\begin{aligned}\partial_0 H=&-(G+\tfrac{1}{3})[H+\tfrac{1}{2}(\mu+3p-2\Lambda)]\theta\\&+\dot{u}^2\theta\,[(G'/p')(\mu+p)-2G+1]\\&+\tfrac{1}{2}(1+3p')(\mu+p)\theta,\end{aligned} \tag{2.6}$$

where we have made use of Eqs. (1.1), (2.2), (A6), (A22), and (A23), and of Proposition 2.1. We now apply that $[\mathbf{e}_0,\mathbf{e}_2]$ commutator (A27) to the function H, and use Eqs. (1.3), (2.4), (2.5), (A7), and (A24), together with Proposition 2.1 and Theorem 3.1, to deduce that

$$d_2\theta\,[(G'/p')(\mu+p)-2G+1]=0. \tag{2.7}$$

Similarly, we can apply the $[\mathbf{e}_0,\mathbf{e}_3]$ commutator (A28) to H and use Eqs. (1.3), (2.5), (2.6), (A8), and (A25) to conclude that

$$d_3\theta\,[(G'/p')(\mu+p)-2G+1]=0. \tag{2.8}$$

[Alternatively, we can regard Eq. (2.8) being derived from

(2.7) by the allowed rotations $\mathbf{e}_1 \to \mathbf{e}_1$, $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to -\mathbf{e}_2$, or $\mathbf{e}_1 \to -\mathbf{e}_1$, $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to \mathbf{e}_2$; cf I.] Equations (2.7) and (2.8) are equivalent to Eq. (2.1).

In the case where $(d_2{}^2 + d_3{}^2)\theta \neq 0$ in (2.1), Theorem 3.1 requires that $\omega \equiv 0$, and Eq. (2.1) implies that

$$(G'/p')(\mu + p) - 2G + 1 = 0. \tag{2.9}$$

Eliminating $\partial_1 A_2$ from equations (A3) and (A17), we obtain

$$2\partial_2 \hat{\sigma}_{22} - \partial_2 \hat{\theta} + 4A_2 \hat{\sigma}_{22} - 2d_2 \dot{u} = 0. \tag{2.10}$$

Upon propagating Eq. (2.10) along $\mathbf{e}_0$, we obtain

$$d_2 \dot{u} \theta (G - \tfrac{2}{3}) = 0, \tag{2.11}$$

where Eqs. (1.1), (1.3), (A7), (A9), (A12), and (A27) and Proposition 2.1 have been employed. Similarly,

$$d_3 \dot{u} \theta (G - \tfrac{2}{3}) = 0, \tag{2.12}$$

so, under the present assumption that $(d_2{}^2 + d_3{}^2)\dot{u}\theta \neq 0$, it follows from (2.11) and (2.12) that $G \equiv \frac{2}{3}$, which contradicts Eq. (2.9). □

Proposition 2.3: If $d_2 = d_3 = 0$ and $\theta \neq 0$, then $\hat{\sigma}_{22} = 0$.

Proof: Since $\theta \neq 0$, it follows by Theorem 3.3 that $\omega \equiv 0$, and hence, by our assumption, that $\dot{u} \neq 0$. We shall suppose in the following that $\hat{\sigma}_{22} \neq 0$, and derive a contradiction. We first observe from (A21) that $\widehat{\Omega} = 0$. Now eliminating $\partial_2 \hat{\sigma}_{22}$ between (A3) and (A17), we obtain

$$\partial_1 A_2 = A_2(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}). \tag{2.13}$$

If we apply the $[\mathbf{e}_0, \mathbf{e}_2]$ commutator (A27) to θ, and use Proposition 2.1, we find that

$$\partial_2(\partial_0 \theta) = 0, \tag{2.14}$$

whereas applying the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A29) to $\dot{u}$, and using (1.3), we see that

$$\partial_2(\partial_1 \dot{u}) = 0. \tag{2.15}$$

Hence, differentiating the (00) field equation (A13) along $\mathbf{e}_2$, and employing Eqs. (1.3), (2.14), (2.15), and (A24) and Proposition 2.1, we obtain $\dot{u}\partial_2 \hat{\theta} = 0$, or, since $\dot{u} \neq 0$,

$$\partial_2 \hat{\theta} = 0. \tag{2.16}$$

Eliminating $\partial_1 \hat{\theta}$ between Eqs. (A19) and (A20), there results

$$\partial_1 \hat{\sigma}_{22} = -\hat{\sigma}_{22}(\dot{u} + \hat{\theta}), \tag{2.17}$$

and so, applying the $[\mathbf{e}_0, \mathbf{e}_1]$ commutator (A26) to $\hat{\sigma}_{22}$ and using (1.1), (2.17), (A6), and (A12) in conjunction with Proposition 2.1, we have $\dot{u}\hat{\sigma}_{22}(G - \frac{2}{3})$, or, since $\dot{u}\hat{\sigma}_{22} \neq 0$,

$$G = \tfrac{2}{3}. \tag{2.18}$$

The next step is to eliminate $\partial_1 A_2$ between (A3) and (A17), recalling (2.16). Thus

$$\partial_2 \hat{\sigma}_{22} = -2\hat{\sigma}_{22} A_2. \tag{2.19}$$

We show that $A_2 = 0$, by applying the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A29) to $\hat{\sigma}_{22}$, and using (1.3), (2.16), (2.17), and (2.19), to obtain

$$\partial_1 A_2 = -A_2(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}). \tag{2.20}$$

Upon comparing Eqs. (2.13) and (2.20), it follows immediately that $A_2 = 0$. Similarly, $A_3 = 0$, i.e.,

$$A_2 = A_3 = 0. \tag{2.21}$$

With Eq. (2.21) in force, we consider Eqs. (A13), (A16), and the sum of (A19) and (A20). If we eliminate $\partial_1 \hat{\theta}$ and $\partial_1 \dot{u}$, we obtain

$$\partial_0 \theta = -\tfrac{3}{2}(p - \Lambda) + \tfrac{3}{2}\dot{u}\hat{\theta} - \tfrac{3}{2}\hat{\sigma}_{22}^2 + \tfrac{3}{8}\hat{\theta}^2 - \tfrac{1}{2}\theta^2, \tag{2.22}$$

whereas if we eliminate $\partial_1 \hat{\theta}$ and $\partial_0 \theta$, we have

$$\partial_1 \dot{u} = -\dot{u}^2 + \tfrac{1}{2}\dot{u}\hat{\theta} - \tfrac{1}{6}\theta^2 + \tfrac{1}{2}(\mu + \Lambda) - \tfrac{3}{2}\hat{\sigma}_{22}^2 + \tfrac{3}{8}\hat{\theta}^2. \tag{2.23}$$

Finally, we apply the $[\mathbf{e}_0, \mathbf{e}_1]$ commutator (A26) to $\dot{u}$ and obtain, using Eqs. (1.1), (2.18), (2.22), (2.23), (A6), (A12), and (A22) and Proposition 2.1, that $\dot{u} \equiv 0$, a contradiction. Hence the proposition is proved. □

In view of the results of Theorems 3.1–3.3 and of Propositions 2.2 and 2.3, it will be convenient to classify the solutions under consideration according to whether or not the quantities ω, θ, $\dot{u}$, $\hat{\sigma}_{22}$, and $d_2{}^2 + d_3{}^2$ vanish. Of the 32 possibilities, eight (those in which $\omega\theta \neq 0$) are immediately ruled out, by Theorem 3.3. Of the remaining 24 cases, in which $\omega\theta \equiv 0$, a further eight are forbidden (those in which $\omega \equiv \dot{u} \equiv 0$) by our assumption that $\dot{u}^2 + \omega^2 \neq 0$. This leaves 16 possibilities, in which $\omega\theta \equiv 0$ and $\dot{u}^2 + \omega^2 \neq 0$. Now Proposition 2.2 disallows two more cases [for which $(d_2{}^2 + d_3{}^2)\dot{u}\theta \neq 0$ and $\omega = 0$], and there are therefore 14 possibilities remaining, in all of which $\omega\theta \equiv 0$, $\dot{u}^2 + \omega^2 \neq 0$, and if $\dot{u} \neq 0$ then $(d_2{}^2 + d_3{}^2)\theta \equiv 0$. Finally, by Proposition 2.3, the case $\dot{u}\theta\hat{\sigma}_{22} \neq 0$, $\omega \equiv d_2{}^2 + d_3{}^2 \equiv 0$ is ruled out, and there are 13 cases remaining. It will be convenient to assign a dimension to each of these 13 possibilities, using the facts that it requires one parameter to specify each of ω, θ, and $\dot{u}$ at a point, and two parameters to specify each of $\hat{\sigma} := \hat{\sigma}_{AB}$ and $d := d_A$ $(A,B = 2,3)$, since (in a general frame) $\hat{\sigma}_{AB}$ is 2×2, symmetric, and trace-free. Respecting the results of Theorem 3.3 and Proposition 2.2, it follows that the "most general" of the 13 possibilities is that in which $\theta \equiv 0$ and $\dot{u}\omega\hat{\sigma}d \neq 0$, which has a dimension of 6, whereas, by our assumption that $\dot{u}^2 + \omega^2 \neq 0$, the "most special" cases are those in which $\dot{u} \neq 0$ and $\omega \equiv 0 \equiv \hat{\sigma} \equiv d \equiv 0$ and in which $\omega \neq 0$ and $\dot{u} \equiv \theta \equiv \hat{\sigma} \equiv d \equiv 0$, each possessing a dimension of 1. All other cases can be assigned intermediate dimensions, and the various possibilities linked by the specializations involved. It is important to note that this algebraic classification is now to be regarded as being based on the vanishing of ω, θ, $\dot{u}$, $\hat{\sigma}$, and d *on an open set*. The results are summarized in Table I. The models are labeled I, II, III, according as they belong to the cases

(i) $\sigma \equiv \omega \equiv 0$, $\theta\dot{u} \neq 0$,
(ii) $\sigma \equiv \theta \equiv \omega \equiv 0$, $\dot{u} \neq 0$,

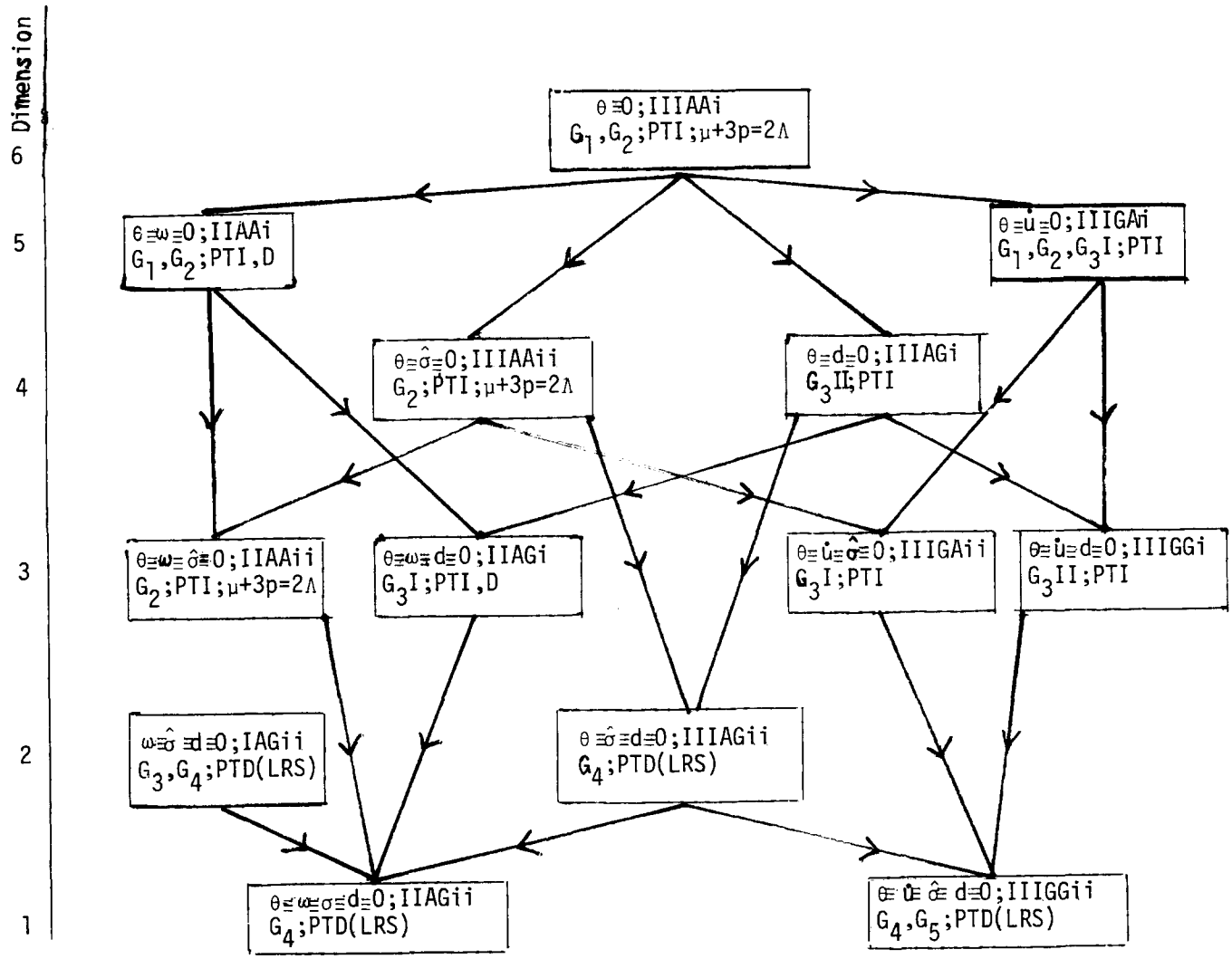
and

(iii) $\sigma \equiv \theta \equiv 0$, $\omega \neq 0$

(cf. Sec. 1). There then follows a letter, *A* or *G*, according as the fluid flow is accelerating or geodesic; a subsequent letter, *A* or *G*, similarly indicates whether the $\mathbf{e}_1$ congruence is accelerating or geodesic. Finally, models in which $\hat{\sigma} \neq 0$ are labeled "i", while "ii" indicates that $\hat{\sigma} \equiv 0$. (Note that with $\hat{\sigma}$ defined as $\hat{\sigma}_{22}$ in the chosen tetrad, we have $2\hat{\sigma}^2 = \hat{\sigma}^{AB}\hat{\sigma}_{AB}$ and $\hat{\sigma} = 0 \Leftrightarrow \hat{\sigma}_{AB} = 0$, in direct analogy to the relationships between σ and σ_{ij}.)

In Sec. 3, we shall study these 13 types of solutions. However, before doing so it is of interest to obtain some further results. Firstly, the four most special types (in which

TABLE I. Specialization diagram for the class of shear-free general relativistic perfect fluids, with an equation of state $p = p(\mu)$, $\mu + p \neq 0$, and with parallel vorticity (ω) and acceleration ($\dot{u}$). This includes the cases where $\dot{u} \equiv 0 \neq \omega$ and where $\omega \equiv 0 \neq \dot{u}$, but excludes the FRW solutions in which $\dot{u} \equiv \omega \equiv 0$. Algebraic dimension is assigned on the left and arrows denote specialization to models of lower dimension. Locally rotationally symmetric (LRS) models are labelled, and in each case the Petrov type (PT) is given. The admissible isometry groups G_n ($n = 1,2,3,4,5$) are also indicated. In all non-LRS cases the G_n acts simply transitively on T_n ($n = 1,2,3$); in LRS types IIAGii, IIIAGii, and IIIGGii the G_n acts multiply transitively on T_{n-1} ($n = 4,5$) and in LRS type IAGii the G_n acts multiply transitively on S_{n-1} ($n = 3,4$) (S_n = spacelike orbit of dimension n; T_n = timelike orbit of dimension n). The notation "G_3I" denotes a Bianchi type I (abelian) isometry group, whereas "G_3II" signifies Bianchi type II.



$\hat{\sigma} \equiv d \equiv 0$) are locally rotationally symmetric (LRS), whereas none of the most general types (in which $\hat{\sigma}^2 + d^2 \neq 0$, where $2\hat{\sigma}^2 = \sigma^{AB}\sigma_{AB}$ and $d^2 = d^A d_A$) is LRS.

Proposition 2.4: The solutions are LRS if and only if $\hat{\sigma} \equiv d \equiv 0$.

Proof. It is clear that if the solutions are LRS then, since $\dot{u}^2 + \omega^2 \neq 0$, the $\mathbf{e}_1$ axis is the spatial symmetry axis, and that hence $\hat{\sigma} \equiv d \equiv 0$, because otherwise there would be a preferred direction in the $\mathbf{e}_2 - \mathbf{e}_3$ plane in the tangent space at any point.

Now suppose that $\hat{\sigma} \equiv d \equiv 0$. It follows from Eqs. (1.3), (A14), and (A15) that $\partial_2 \dot{u} = \partial_3 \dot{u} = \partial_2 \omega = \partial_3 \omega = 0$. Moreover, by Proposition 2.1, $\partial_2 \theta = \partial_3 \theta = 0$, and, eliminating $\partial_1 A_2$ between (A3) and (A17), and $\partial_1 A_3$ between (A4) and (A18), we have $\partial_2 \hat{\theta} = \partial_3 \hat{\theta} = 0$. With the allowed tetrad freedom, $\mathbf{e}_1 \to \pm \mathbf{e}_1$ and the rotation

$$\mathbf{e}_2 \to \mathbf{e}_2 \cos \Theta + \mathbf{e}_3 \sin \Theta, \qquad \mathbf{e}_3 \to -\mathbf{e}_2 \sin \Theta + \mathbf{e}_3 \cos \Theta, \tag{2.24}$$

with $\partial_0 \Theta = 0$ (cf. Sec. 2 of I), it is possible to arrange for $\partial_2 \hat{\Omega} = \partial_3 \hat{\Omega} = \partial_2 A_2 = \partial_3 A_2 = \partial_2 A_3 = \partial_3 A_3 = 0$, and one can therefore deduce (cf. Ref. 15) that the solutions are LRS. However, this involves the investigation of several compatibility requirements for the propagation of Θ. We find it more convenient, particulary in view of our analysis in Sec. 3, to show that any solution admits an isometry group which acts multiply transitively either in the hypersurfaces orthogonal

to $\mathbf{e}_1$ (in the case $\omega \neq 0$) or in the two-surfaces orthogonal to $\mathbf{e}_0$ and $\mathbf{e}_1$ (in the case $\omega \equiv 0$), and it then follows (cf. Ref. 16) that the solutions are LRS, with $\mathbf{e}_1$ as the axis of symmetry. Following Ellis,[2] we investigate the existence of solutions to the equations

$$[\xi,\mathbf{e}_0] = [\xi,\mathbf{e}_1] = 0, \quad [\xi,\mathbf{e}_2] = b\mathbf{e}_3,$$

and

$$[\xi,\mathbf{e}_3] = -b\mathbf{e}_2, \tag{2.25}$$

which are equivalent to the equations for the Killing vector ξ. Here, b is a scalar function on the space-time. Under the transformation (2.24), $\hat{\Omega} \to \hat{\Omega} + \partial_1\Theta$, and so propagating Θ according to the requirement $\partial_1\Theta = -\hat{\Omega}$, we may arrange for $\hat{\Omega} = 0$ [note that this is permitted, since the $[\mathbf{e}_0, \mathbf{e}_1]$ commutator (A26) applied to Θ is satisfied, by virtue of (A11)]. Expanding ξ in the form $\xi = \xi^0\mathbf{e}_0 + \xi^2\mathbf{e}_2 + \xi^3\mathbf{e}_3$ (i.e., demanding that $\xi^1 := \xi_0 \cdot \mathbf{e}_1 \equiv 0$), substituting into (2.25), and employing the commutation relations (A26)–(A31), we obtain propagation equations for ξ^0, ξ^2, ξ^3 (see Appendix B for the equations in the general case, with ξ^1 not necessarily zero). In order for the equations to be compatible, the commutation relations acting on ξ^a must be satisfied, and this leads to the restrictions [cf. Eqs. (A48)–(A59) of Appendix B]

$$\begin{aligned} &\xi^0\partial_0\dot{u} = \xi^0\partial_0\theta = \xi^0\partial_0\hat{\theta} = 0, \\ &\partial_0 b = \partial_1 b = 0, \\ &\partial_2 b + A_2 b + \xi^a\partial_a A_3 = 0, \quad \partial_3 b + A_3 b - \xi^a\partial_a A_2 = 0, \end{aligned} \tag{2.26}$$

where use is made of Eqs. (A1), (A3)–(A6), (A9), and (A10), and Proposition 2.1 and Theorem 3.3. Equations (2.26) themselves provide propagation equations for the function b, which, by Eqs. (A3), (A4), (A9), (A10), and (A19) are compatible. In the case $\omega \neq 0$, we have $\theta \equiv 0$ by Theorem 3.3. We therefore have a self-consistent system of equations for the four quantities ξ^0, ξ^2, ξ^3, and b, all being determined uniquely by fixing their values at one arbitrary point, i.e., there is a G_4 isometry group acting multiply transitively on the hypersurfaces orthogonal to $\mathbf{e}_1$. In the case $\omega \equiv 0$, there are solutions with $\xi^0 \equiv 0$, and the three quantities ξ^2, ξ^3, and b are determined uniquely by their values at one point, i.e., there is a G_3 isometry group acting multiply transitively on the two-surfaces orthogonal to $\mathbf{e}_0$ and $\mathbf{e}_1$. Note that in each case the first three of Eqs. (2.26) are satisfied, by virtue of (1.1) and (A6), and that the groups whose existence we have thus determined are not necessarily the *maximal* isometry groups.□

It is also of interest to note that in three of the nine non-LRS models, the matter necessarily satisfies the unphysical equation of state $\mu + 3p - 2\Lambda = 0$. This is summarized in

Proposition 2.5: In cases IIAAii, IIIAAi, and IIIAAii, the matter content satisfies the unphysical equation of state $\mu + 3p - 2\Lambda = 0$, i.e., if $\dot{u}d \neq 0$ and $\theta \equiv 0$, and either $\omega \neq 0$ or $\omega \equiv \hat{\sigma} \equiv 0$, then $\mu + 3p - 2\Lambda = 0$.

Proof: We begin by applying the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A29) to $\dot{u}$, thereby obtaining

$$\begin{aligned} &\dot{u}[\partial_1 d_2 + \partial_2\hat{\theta} + d_2(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - d_3\hat{\Omega}\,] \\ &\quad + d_2(\mu + 3p - 2\Lambda) = 0, \end{aligned} \tag{2.27}$$

where use is made of Eqs. (1.3), (A13), (A14), and (A24). Similarly, if we apply the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A29) to ω, and use (1.3), (A1), (A14), and (A15), we have

$$\omega[\partial_1 d_2 + \partial_2\hat{\theta} + d_2(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - d_3\hat{\Omega}\,] = 0. \tag{2.28}$$

In the case where $\omega \neq 0$, it now follows from (2.27) and (2.28) that $d_2(\mu + 3p - 2\Lambda) = 0$. Similarly, we have (if $\omega = 0$), $d_3(\mu + 3p - 2\Lambda) = 0$, and so, since $d \neq 0$, it follows that $\mu + 3p - 2\Lambda = 0$.

Now suppose that $\omega \equiv \hat{\sigma}_{22} \equiv 0$. We can rotate the tetrad so that $d_3 \equiv 0$, and then the difference between (A19) and (A20) shows that

$$\partial_2 d_2 = d_2(d_2 + A_2), \tag{2.29}$$

and (A2) and (A21) together show that $\partial_3 d_2 = A_3 = 0$. Forming the difference between (A3) and (A17), we obtain $\partial_2\hat{\theta} = -2d_2\dot{u}$, and from the difference between (A4) and (A18), $\partial_3\hat{\theta} = 0$. Now applying the $[\mathbf{e}_3, \mathbf{e}_1]$ commutator (A31) to $\dot{u}$ [cf. the derivation of (2.27)], we have $\hat{\Omega} = 0$, since $\dot{u}d_2 \neq 0$. Thus Eq. (2.27) becomes

$$\dot{u}[\partial_1 d_2 - 2d_2\dot{u} - \tfrac{1}{2}d_2\hat{\theta}\,] + d_2(\mu + 3p - 2\Lambda) = 0,$$

whose compatibility with equation (2.29) is assured provided

$$\dot{u}(2\dot{u} + \hat{\theta}) = 3(\mu + 3p - 2\Lambda), \tag{2.30}$$

where use is made of the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A29) and Eqs. (1.3), (A3), and (A24). Differentiation of (2.30) along $\mathbf{e}_2$ now implies that $2\dot{u} + \hat{\theta} = 0$, since $\dot{u}d_2 \neq 0$, and so by (2.30) itself, $\mu + 3p - 2\Lambda = 0$, as required.

3. EXAMINATION OF THE VARIOUS CASES

In this section we consider each of the 13 cases in Table I. We shall investigate the possible isometry groups admitted by space-times in each case, and we shall also examine the Petrov types. Wherever possible, we link our results to those of other authors. It follows from Proposition 2.4 that all solutions of types IAGii, IIAGii, IIIAGii, and IIIGGii are locally rotationally symmetric, whereas no other solution is. In these non-LRS cases, there is a geometrically invariantly defined tetrad, and so the study of isometry groups involves an investigation of (2.25) with the function $b \equiv 0$ (cf. Ref. 2). In Appendix B, we have written out in equations (A32)–(A47) the components of (2.25) for a general $\xi = \xi^a\mathbf{e}_a$ (cf. the proof of Proposition 2.4), using the commutation relations (A26)–(A31). Appendix B also contains in (A48)–(A59) the set of equations obtained as consistency conditions on (A32)–(A47), where use is made of the equations in Appendix A. In the non-LRS case ($b \equiv 0$), these conditions reduce to the standard requirement that the commutation functions $\gamma^a{}_{bc}$ be constant on the group orbits.

We shall find that, among the 12 types of solutions in which $\theta \equiv 0$, the most general case admits only a G_1 isometry group. An abelian G_2 isometry group may arise under special circumstances. In certain other cases, the abelian G_2 group is not the maximal isometry group, but it arises as a subgroup of the maximal group. All such cases in which there is an abelian G_2 isometry group acting on timelike orbits may be regarded *locally* as stationary axisymmetric solutions of Einstein's field equations, which have been extensively studied in view of their astrophysical applications (see, e.g., the discussion and references cited in Ref. 16). Since in our case the fluid flow vector lies in the group orbits, and since the fluid is

both shear-free and expansion-free, the system is thus undergoing rigid rotation. However, it is important to recognize that there are also solutions under consideration which *cannot* be regarded as rigidly rotating stationary axisymmetric space-times *in the conventional sense*. The reason for this is that in order to be axisymmetric, a solution must admit a spacelike Killing vector field whose trajectories are closed curves. Given an abelian G_2 isometry group we can select a spacelike Killing vector field and identify points along the trajectories (cf. Ref. 17), thus making the space-time axisymmetric. In general, however, there will not be a rotation axis on which the Killing vector field vanishes, and even if there is, the space-time need not necessarily be regular on the axis (cf. the simple examples of conical singularities constructed in Minkowski space-time[18]). Thus the conventional study of stationary axisymmetric gravitational fields involves the imposition of global requirements, not only on the compactness of the trajectories of a certain spacelike Killing vector field, but also on the existence of a well-behaved rotation axis. That not all of our solutions would satisfy these requirements, even under appropriate identifications, is to be expected, and is made particularly evident in view of a well-known result of Kundt and Trümper,[19] which states that stationary axisymmetric gravitational fields are orthogonally transitive if and only if $\xi^{d}R_{d[a}\xi_b\eta_{c]} = \eta^{d}R_{d[a}\xi_b\eta_{c]} = 0$, where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are the Killing vectors involved. In the case of a perfect fluid whose flow vector is tangential to the orbits of the group generated by $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, these conditions are automatically satisfied. We shall find in some cases that, while there is an abelian G_2 isometry group with timelike orbits, this group is not necessarily orthogonal transitive. This does not contradict the result of Kundt and Trümper,[19] since that result is proved using the conventional definition of axisymmetry (indeed, given the conditions $\xi^{d}R_{d[a}\xi_b\eta_{c]} = \eta^{d}R_{d[a}\xi_b\eta_{c]} = 0$, orthogonal transitivity is proved by using the rotation axis in a crucial manner; see Refs. 16 and 20). In a similar manner (under appropriate identifications, but ignoring the question of the rotation axis), we may regard certain specializations as being static and/or as possessing cylindrical symmetry. Some further remarks, concerning the existence of a rotation axis, are given in Sec. 4.

The investigation of the consistency of the system of tetrad equations in each of the 13 types involves some very lengthy and tedious calculations. For the sake of brevity, since the actual techniques are similar to those already presented both in this article and also in I, we shall not give the details, but merely present the results. Similar remarks apply, in general, to the calculations of the isometry groups, although a somewhat distinct situation arises when, in a given type, additional symmetries are allowed. Such cases are discussed in a slightly extended manner. The calculations of the allowed Petrov types involves an investigation of the admissible canonical forms of the trace-free complex tensor $Q_{ab} = E_{ab} + iH_{ab}$, where E_{ab} and H_{ab} are, respectively, the electric and magnetic parts of the Weyl tensor (see, e.g., Ref. 16). This is achieved by first determining the *general* (i.e., $\hat{\sigma}_{23}$ is not set equal to zero) tetrad components, $Q_{\alpha\beta}$, of Q_{ab}, using, for example, the expressions given by MacCallum,[21] and then applying some established results in order to narrow down the range of possibilities. For instance, there are no Petrov type O (conformally flat) models in our class, since if $E_{ab} = H_{ab} = 0$ for a perfect fluid with an equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$, it necessarily follows that $\dot{\mathbf{u}} = \boldsymbol{\omega} = \mathbf{0}$ (and the space-time is FRW; this result, credited to Trümper, is proved by Ellis[3]). Also, if $d = 0$, it follows that $E_{\alpha\beta} + iH_{\alpha\beta}$ is diagonal in a frame with respect to which $\hat{\sigma}_{23} = 0$, and so the space-times are of Petrov types I or D. Moreover, if both $\hat{\sigma}$ and d vanish, the models are LRS, and therefore Petrov type D, whereas if both θ amd ω vanish, the models are static, and therefore either Petrov type I or D.[16]

These results therefore show that, at least apart from the four types IIIAAi ($\theta \equiv 0$), IIIAAii ($\theta \equiv \hat{\sigma} \equiv 0$), IIIGAi ($\theta \equiv \dot{u} \equiv 0$), and IIIGAii ($\theta \equiv \dot{u} \equiv \hat{\sigma} \equiv 0$), the space-times in our class are of Petrov type I or D. In fact, this is true of all space-times in our class. We first considered the Petrov type III and N cases, in which $Q_{\alpha\beta}Q^{\beta\gamma}Q_{\gamma}{}^{\delta} = 0$, and showed that this requirement could not hold in an open set. We then considered the Petrov type II possibility, in which $Q_{\alpha\beta}$ possesses one real eigenvector, and one complex null eigenvector. With respect to any frame, at least one of $d_2^2 + \hat{\sigma}_{23}^2$ and $d_3^2 + \hat{\sigma}_{23}^2$ is nonzero; without loss of generality (renumbering if necessary) we can assume that $d_3^2 + \hat{\sigma}_{23}^2 \neq 0$, that the real eigenvector is in a direction parallel to the nonzero vector $(0, -\hat{\sigma}_{23}, d_3, 0)$ and that the null eigenvector is therefore (parallel to) the vector $(0, d_3, \hat{\sigma}_{23}, \pm i(d_3^2 + \hat{\sigma}_{23}^2)^{1/2})$. Writing out the components of the eigenvalue equations for both eigenvectors, and appealing to the fact that the null eigenvector is associated with a repeated eigenvalue it follows by propagation of the resulting algebraic equations that this situation is untenable. We may thus conclude that all models of our class are either Petrov type I or Petrov type D, and it remains to investigate the circumstances under which the solutions can be of the more special Petrov type D. This now corresponds to the situations in which $Q_{\alpha\beta}$ possesses a pair of equal eigenvalues, and we may therefore rotate the frame about the $\mathbf{e}_1$ axis [as in (2.24)] to arrange for the preferred eigenvector to be in the $\mathbf{e}_1$-$\mathbf{e}_2$ plane, and thus for $\mathbf{e}_3$ to be an eigenvector associated with the repeated eigenvalue. In this case, $d_3 = \hat{\sigma}_{23} = 0$, and the problem becomes one of investigating the compatibility of the system of equations relative to this tetrad and subject to the restriction on $Q_{\alpha\beta}$. Once again this is a fairly standard but tedious calculation, whose details are omitted. The results are contained in the subsequent discussion of the various types of models in our class.

1. Type IIIAAi ($\theta \equiv 0$). By Proposition 2.5, the fluid has the unrealistic equation of state $\mu + 3p - 2\Lambda = 0$. Equation (2.27) provides a propagation equation for d_2 along $\mathbf{e}_1$:

$$\partial_1 d_2 + \partial_2\hat{\theta} + d_2(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - d_3\hat{\Omega} = 0, \tag{3.1}$$

and similarly

$$\partial_1 d_3 + \partial_3\hat{\theta} - d_3(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}) + d_3\hat{\Omega} = 0. \tag{3.2}$$

The remaining equations are consistent [e.g., the derivative of (A2) along $\mathbf{e}_1$ is compatible with (3.1) and (3.2)]. The solutions are of Petrov type I, and in general admit a G_1 isometry group, with Killing vector parallel to the fluid flow vector $\mathbf{u}$. The only special case that can arise is when there is an abelian G_2 isometry group whose orbits are orthogonal to both

ω (or, equivalently, to $\dot{\mathbf{u}}$) and $\mathbf{d}$; this G_2 is orthogonally transitive if and only if $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{\alpha\beta}$ (in which case $d_3 = 0$, without loss of generality, and then $\widehat{\Omega} = A_3 = 0$ and $\partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$; a proof of this statement is provided in Appendix C). The space-times are stationary, and in the special case of a G_2 may be regarded as also being axisymmetric (cf. our previous remarks in this section).

2. Type IIAAi ($\theta \equiv \omega \equiv 0$). This is the specialization of type IIIAAi to the static case. The commutation relations (A29) and (A31) applied to $\dot{u}$ result in propagation equations for d_2 and d_3 along $\mathbf{e}_1$:

$$\partial_1 d_2 + \partial_2 \hat{\theta} + d_2(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - d_3\widehat{\Omega} + (d_2/\dot{u})(\mu + 3p - 2\Lambda) = 0$$

and

$$\partial_1 d_3 + \partial_3 \hat{\theta} - d_3(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}) + d_2\widehat{\Omega} + (d_3/\dot{u})(\mu + 3p - 2\Lambda) = 0,$$

and the remaining equations are consistent. In general, the solutions are of Petrov type I, but may be of Petrov type D in special cases. In order for a solution to be of Petrov type D, it is necessary that $\mathbf{d}$ should be an eigenvector of $\hat{\sigma}_{\alpha\beta}$. In a frame in which $d_3 = \hat{\sigma}_{23} = 0$, it then follows that in all Petrov type D solutions,

$$\frac{dp}{d\mu} = 1, \quad A_3 = \widehat{\Omega} = 0, \quad \partial_3 \gamma^a{}_{bc} = 0,$$

$$4\dot{u}\hat{\sigma}_{22}^2 + \dot{u}d_2^2 + (\mu + p)\hat{\sigma}_{22} = 0, \tag{3.3}$$

$$-2\dot{u}\hat{\sigma}_{22}^2 + \dot{u}d_2^2 + 3\dot{u}\hat{\theta}\hat{\sigma}_{22} - (\mu + 3p - 2\Lambda)\hat{\sigma}_{22} = 0,$$

and

$$2d_2(\hat{\sigma}_{22} + \dot{u}) - d_2\hat{\theta} - 4A_2\hat{\sigma}_{22} = 0,$$

where $\gamma^a{}_{bc}$ refers to any commutation function. In such a tetrad, Eqs. (3.3) together with the field equations and Jacobi identities are self-consistent, and in fact Eqs. (3.3) provide the necessary and sufficient conditions for type IIAAi models to be of Petrov type D in such a frame. The Petrov type D models have already been determined as exact solutions by Barnes[11,22] in his studies of shear-free normal flows, the present type corresponding to Barnes' class IV solutions. In the general case, type IIAAi solutions admit a G_1 isometry group, with Killing vector parallel to the fluid flow. As in type IIIAAi, there is the possibility of an additional symmetry, in which case there is an abelian G_2 isometry group whose orbits are orthogonal to both $\dot{\mathbf{u}}$ and $\mathbf{d}$; and again the G_2 is orthogonally transitive if and only if $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{AB}$ (in which case $d_3 = 0$, without loss of generality, and then $\widehat{\Omega} = A_3 = 0$ and $\partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$; a proof of this is provided in Appendix C). Thus the Petrov type D solutions admit an orthogonally transitive abelian G_2 isometry group with timelike orbits (which agrees with the results of Barnes[11,22]), although the converse is not true. When there is an abelian G_2 isometry group, the space-times may be considered as being static and axisymmetric (cf. our previous remarks).

3. Type IIIGAi ($\theta \equiv \dot{u} \equiv 0$). The commutators (A29) and (A31) applied to ω imply Eqs. (3.1) and (3.2), and the (00) field equation, (A13), requires that $4\omega^2 = \mu + 3p - 2\Lambda$ (which may be regarded as determining μ). The field equations and Jacobi identities are now self-consistent. The solutions are always of Petrov type I, in general admitting a G_1 isometry group with Killing vector parallel to $\mathbf{u}$, the fluid-flow vector. The admissible special cases are as follows. First, there could be a second Killing vector, $\xi^a \mathbf{e}_a$, where $\xi^1 \hat{\theta} = \xi^2 d_2 + \xi^3 d_3$ and $(\xi^2)^2 + (\xi^3)^2 \neq 0$. This case gives rise to an abelian G_2 isometry group acting on timelike 2-surfaces. This group is orthogonally transitive if and only if $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{AB}$ *and* the group orbits are orthogonal to both ω and $\mathbf{d}$ (in which case $d_3 = 0$, without loss of generality, and then $\widehat{\Omega} = A_3 = 0$ and $\partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$ (cf. type IIIAAi); a proof is given in Appendix C). In addition, it is possible for solutions of type IIIGAi to admit a G_3 isometry group of Bianchi type I (whose orbits are timelike hypersurfaces orthogonal to $\mathbf{d}$), and this occurs if and only if $\hat{\theta} = 0$ and $\hat{\sigma}_{22} = \pm\widehat{\Omega}$ (in which case, $d_3 = \pm d_2$, $A_3 = \pm A_2$, $4\hat{\sigma}_{22}^2 + 8d_2A_2 = \mu - p + 2\Lambda$, and $\partial_1 \gamma^a{}_{bc} = \partial_2 \gamma^a{}_{bc} \mp \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions, $\gamma^a{}_{bc}$). Since the acceleration is zero, and we are assuming an equation of state $p = p(\mu)$, it follows from the Bianchi identities (A22)–(A25) that p is identically constant, and the solutions are therefore most readily interpreted as dust ($p \equiv 0$) models. The solutions admitting a G_3 isometry group correspond to the general class Ciii (i.e., $c \neq 0$) of Ellis[2]; the abelian G_2 case with $\xi^1 \equiv 0$ corresponds to the general class Cii (i.e., $c \neq 0$) of Ellis,[2] whereas the orthogonally transitive abelian G_2 case is the special case of Class Cii with $c = 0$ in Ellis,[2] and so (in the case of a regular rotation axis) is the case treated by Van Stockum.[23] When there is an abelian G_2 isometry group, the solutions can be regarded as being stationary axisymmetric dust solutions, and as being stationary cylindrically symmetric dust solutions when there is an abelian G_3 isometry group (cf. our previous remarks). Winicour[24] provides a construction of the exact solution of all stationary axisymmetric dust solutions in which the flow vector is tangential to the group orbits and in which there is a well-behaved axis of rotation.

4. Type IIIAAii ($\theta \equiv \hat{\sigma} \equiv 0$). This represents a specialization of type IIIAAi in which the requirement $\mu + 3p - 2\Lambda = 0$ is preserved, i.e., the equation of state is unphysical. Without loss of generality, $d_3 = 0$, and then $\widehat{\Omega} = 0$, $A_3 = 0$ and $\partial_3 \gamma^a{}_{bc} = 0$, for all commutation functions $\gamma^a{}_{bc}$. We must now append to (3.1) the extra propagation equation

$$\partial_2 d_2 = d_2(d_2 + A_2) \tag{3.4}$$

and the algebraic constraints

$$\hat{\theta} = -2\dot{u} \tag{3.5}$$

and

$$4\omega^2 + 4\dot{u}^2 - 4d_2A_2 + \mu - p + 2\Lambda = 0. \tag{3.6}$$

The system of equations is then self-consistent. The space-times are of Petrov type I, and there is an orthogonally transitive abelian G_2 isometry group whose orbits are orthogonal to both ω (or, equivalently, $\dot{\mathbf{u}}$) and $\mathbf{d}$. They may therefore be regarded as stationary axisymmetric space-times (cf. our remarks above).

5. Type IIIAGi ($\theta \equiv d \equiv 0$). In this case $\widehat{\Omega} = A_2 = A_3 = 0$, and

$$\tfrac{1}{4}\hat{\theta}^2 - \hat{\sigma}_{22}^2 + \dot{u}\hat{\theta} - \omega^2 - (p - \Lambda) = 0. \tag{3.7}$$

Furthermore, $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc}$ for all commutation functions

$\gamma^a{}_{bc}$. The resulting system of equations is now consistent. These solutions are of Petrov type I and admit a G_3 isometry group of Bianchi type II acting transitively on hypersurfaces orthogonal to ω and $\dot{\mathbf{u}}$.

6. Type IIAAii ($\theta \equiv \omega \equiv \hat{\sigma} \equiv 0$). Consistency of the constraints requires the unphysical equation of state $\mu + 3p - 2\Lambda = 0$. Without loss of generality, we may choose a tetrad in which $d_3 = 0$, and then $\partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$, and $A_3 = \hat{\Omega} = 0$. The quantity d_2 is propagated along $\mathbf{e}_1$ in accordance with Eq. (3.1), and furthermore, Eqs. (3.4)–(3.6), specialized to the case $\omega = 0$, hold as additional constraints. The system of equations is now self-consistent. The space-times are static, of Petrov type I and admit an orthogonally transitive abelian G_2 isometry group whose orbits are timelike 2-surfaces orthogonal to $\dot{\mathbf{u}}$ and $\mathbf{d}$. They may be regarded locally as static axisymmetric solutions (cf. our previous remarks).

7. Type IIAGi ($\theta \equiv \omega \equiv d \equiv 0$). This is a specialization of type IIIAGi. Once again, $\hat{\Omega} = A_2 = A_3 = 0$, and Eq. (3.7), with $\omega = 0$, must hold. Furthermore, $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc}$ for all commutation functions $\gamma^a{}_{bc}$, and the resulting system of equations is now consistent. In general, solutions of type IIAGi are of Petrov type I, but they may also be of Petrov type D. In the Petrov type D case, the commutation functions are constrained by the further relationships

$$(2\hat{\sigma}_{22} + 3\hat{\theta})\dot{u} = \mu + 3p - 2\Lambda,$$
$$\hat{\theta}^2 - 4\hat{\sigma}_{22}^2 + 4\dot{u}\hat{\theta} = 4(p - \Lambda),$$
$$2\hat{\sigma}_{22}\hat{\theta} - 4\hat{\sigma}_{22}^2 - 4\dot{u}\hat{\sigma}_{22} = -(\mu + p)/p',$$

and

$$4\hat{\sigma}_{22}(p' - 1)p' + \dot{u}[(\mu + p)(p''/p') + p' - 1] = 0,$$

from which it follows that the pressure and density are related to each other either by the second-order differential equation

$$(\mu + p)(\mu + 3p - 2\Lambda)\frac{d^2p}{d\mu^2} = \frac{dp}{d\mu}\left(3\frac{dp}{d\mu} - 1\right)(\mu - p + 2\Lambda) \quad \left(\text{with } \frac{dp}{d\mu} \neq \frac{1}{3}\right)$$

or the first-order differential equation

$$\frac{dp}{d\mu} = 1$$

[cf. the similar differential equations obtained by Collins and Wainwright[4] for our case IAGii treated below; these equations are (3.10) and (3.13) below]. The Petrov type D solutions have been obtained in exact form by Barnes,[11,22] and correspond to his class III. All solutions of type IIAGi admit a G_3 isometry group of Bianchi type I, whose orbits are timelike hypersurfaces orthogonal to $\dot{\mathbf{u}}$. All of the solutions are static; with appropriate identifications, they may be regarded as static *cylindrically* symmetric solutions (cf. our previous remarks), and with regular rotation axes fall into the class studied by Evans[25] and Bronnikov.[26] Note that the class of solutions with an ultrarelativistic equation of state $p = \frac{1}{3}\mu$ investigated by Teixera, Wolk, and Som[27] cannot be of Petrov type D, since as we have shown, $dp/d\mu \neq \frac{1}{3}$ in our Petrov type D solutions.

8. Type IIIGAii ($\theta \equiv \dot{u} \equiv \hat{\sigma} \equiv 0$). This is a specialization of both types IIIGAi and IIIAAii. Without loss of generality, $d_3 = 0$, and then $\hat{\Omega} = A_3 = \hat{\theta} = 0$, and $\partial_1 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$. The commutation function d_2 is propagated along $\mathbf{e}_2$ according to (3.4); note that (3.5) is still valid when specialized to this type. Equation (3.6) is replaced by the relations

$$4\omega^2 = \mu + 3p - 2\Lambda \tag{3.8}$$

and

$$4d_2A_2 = \mu - p + 2\Lambda,$$

and the entire system is now self-consistent. The solutions are of Petrov type I and admit a G_3 isometry group of Bianchi type I acting transitively on timelike hypersurfaces orthogonal to $\mathbf{d}$. When interpreted as dust solutions, they correspond to the case Bii of Ellis,[2] and, when interpreted as cylindrically symmetric dust solutions, to the special solution of Van Stockum,[23] which was first determined by Lanczos,[28] rediscovered by Wright,[29] who established the Petrov type and discussed the isometries, and later obtained by Krasinski[14] in his search for rigidly rotating perfect fluids possessing a Killing vector parallel to the vorticity vector.

9. Type IIIGGi ($\theta \equiv \dot{u} \equiv d \equiv 0$). This is a specialization of types IIIGAi and IIIAGi, and again we have $\hat{\Omega} = A_2 = A_3 = 0$ and $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions $\gamma^a{}_{bc}$. Furthermore, Eq. (3.7), specialized to the case $\dot{u} \equiv 0$, is still valid, and in addition we have the relation (3.8). The system of field equations and Jacobi identities is now self-consistent. The solutions are of Petrov type I and admit a G_3 isometry group whose orbits are timelike hypersurfaces orthogonal to ω. With the interpretation that the source is dust, this type corresponds to the exact solution of cases Aii and Ci of Ellis.[2]

10. Type IAGii ($\omega \equiv \hat{\sigma} \equiv d \equiv 0$). This type is LRS (cf. Proposition 2.4), and in view of its high symmetry, it has been discussed extensively in the literature. Without loss of generality, $\hat{\Omega} = 0$, and we obtain $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions save possibly A_2 and A_3 (cf. the proof of Proposition 2.4). In addition, $\partial_1 \theta = 0$, by Proposition 2.1. The system of equations is consistent provided

$$(3G - 2)p'[\mu + \Lambda + \tfrac{3}{4}\hat{\theta}^2 - \tfrac{1}{3}\theta^2 + 3(\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2)] + 6\dot{u}^2[2Gp' - G'(\mu + p)] = 0 \tag{3.9}$$

whose preservation requires a succession of constraints, depending on whether or not $\dot{u} = \dot{u}(\mu)$, as follows:

(i) $\dot{u} \neq \dot{u}(\mu)$: $9p'G = 9p'^2 - 1$, (3.10)

$$10(1 - 3p')\dot{u}^2 - 15p'(1 - 3p')\dot{u}\hat{\theta} - 27p'^2(u + p) = 0, \tag{3.11}$$

and

$$r \equiv -(\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2) = \left(\frac{\dot{u}}{3p'} - \frac{\hat{\theta}}{2}\right)^2 > 0. \tag{3.12}$$

(ii) $\dot{u} = \dot{u}(\mu)$:

$$G(3G - 2) + 3(1 + 3p')[(\mu + p)G' - 2Gp'] = 0, \tag{3.13}$$

$$2G\dot{u}^2[9Gp' - 9p'^2 + 1] + p'(\mu + p)(1 + 3p')^2 = 0, \tag{3.14}$$

$$(1 + 3p')\hat{\theta} = 2[1 + 3p' - 3G]\dot{u}, \tag{3.15}$$

and

$$r \equiv -(\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2) = 0. \quad (3.16)$$

Here G is the function defined in Eq. (1.5), in terms of μ, p, p', and p''. Thus in case (i), Eq. (3.10) represents a second-order differential equation for p as a function of μ, and by (3.12) the solutions are spherically symmetric [when $\Lambda = 0$, the differential equation thus obtained is equivalent to Eq. (3.18b) of Collins and Wainwright[4]]. In this case we may regard Eqs. (3.9)–(3.12) as determining the quantities θ, $\hat{\theta}$, and r in terms of basic independent variables, $\dot{u}$ and μ. In case (ii), Eq. (3.13) represents a third-order differential equation for p as a function of μ, and by (3.16) the solutions are plane symmetric [cf. Eq. (3.18a) of Ref. 4]. In this case, θ is identically constant, and Eqs. (3.9), (3.14), and (3.15) may be regarded as determining the quantities θ, $\dot{u}$, and $\hat{\theta}$ in terms of the one basic variable, μ [note also that in case (i) there is a special case in which θ is identically constant, but that this requires $p' = -\frac{1}{6}$, so $G = \frac{1}{2}$ and Eq. (3.10) is redundant; cf. Eq. (3.18c) of Ref. 4].

In case (i), the solutions admit a G_3 isometry group acting multiply transitively on the spatial 2-surfaces orthogonal to both $\mathbf{u}$ and $\dot{\mathbf{u}}$ and are spherically symmetric. They were first given in exact form by Wyman.[5] Further details may be found in Refs. 4 and 14, and references cited therein. In case (ii), the solutions admit a G_4 isometry group whose orbits are hypersurfaces tangential to $\mathbf{e}_2$ and $\mathbf{e}_3$ and which may be timelike or spacelike. These solutions are plane symmetric and hypersurface homogeneous, and are given in exact form by Collins and Wainwright.[4] They are of interest since they show that shear-free spatially homogeneous perfect fluids are not necessarily FRW, contrary to the claims of King and Ellis.[10]

Since all solutions of type IAGii are LRS, they are of Petrov type D (recall from above that they cannot be of Petrov type O, since we are assuming that $\dot{u}^2 + \omega^2 \neq 0$). They belong to case IIc of Stewart and Ellis.[12]

11. Type IIIAGii ($\theta \equiv \hat{\sigma} \equiv d \equiv 0$). These solutions are LRS and we may again choose a tetrad in which $\widehat{\Omega} = 0$; then $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions except possibly A_2 and A_3 (cf. the proof of Proposition 2.4). The system of field equations and Jacobi identities is self-consistent. The solutions are of Petrov type D, and admit a G_4 isometry group acting multiply transitively on timelike hypersurfaces orthogonal to $\boldsymbol{\omega}$ (and to $\dot{\mathbf{u}}$). In general, they belong to the case Id of Stewart and Ellis,[12] but may also belong to case Ib; this alternative is possible if and only if $\hat{\theta} \equiv 0$, which requires $2\omega^2 = (\mu + p)$ and an unphysical equation of state satisfying $dp/d\mu = -\frac{1}{3}$. Solutions of type IIIAGii have recently been studied by Lukács, Newman, Sparling, and Winicour.[30]

12. Type IIAGii ($\theta \equiv \omega \equiv \hat{\sigma} \equiv d \equiv 0$). This type is static and LRS. The commutation functions are constrained by

$$\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2 + \tfrac{1}{4}\hat{\theta}^2 + \dot{u}\hat{\theta} - (p - \Lambda) = 0 \quad (3.17)$$

(cf. Eq. (3.8) of Ref. 12), and $\partial_2 \gamma^a{}_{bc} = \partial_3 \gamma^a{}_{bc} = 0$ for all commutation functions except possibly A_2 and A_3. The solutions are of Petrov type D and belong to case IIc of Stewart and Ellis.[12] They admit a G_4 isometry group acting multiply-transitively on timelike hypersurfaces orthogonal to $\dot{\mathbf{u}}$. A construction of all static *plane-symmetric* solutions has been given by Taub,[31] and some solutions for $p = \frac{1}{3}\mu$ were obtained by Teixera, Wolk, and Som.[32] Plane symmetric solutions are governed by the further specification that $r \equiv (\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2) = 0$ in (3.17), which is automatically preserved throughout space-time. Static *spherically symmetric* solutions, in which $r > 0$, also belong to this class, and have been extensively studied for a variety of equations of state (see the discussion in Ref. 16).

13. Type IIIGGii ($\theta \equiv \dot{u} \equiv \hat{\sigma} \equiv d \equiv 0$). These solutions are LRS and of Petrov type D. The system of equations, with (3.8) appended, is self-consistent. In the general case ($\hat{\theta} \neq 0$), the solutions admit a G_4 isometry group, multiply transitive on the timelike hypersurfaces orthogonal to $\boldsymbol{\omega}$. These solutions belong to case Ic of Stewart and Ellis[12] and are given in exact form by these authors (but note that there are some errors, most notably in their expression for μ). If the fluid is interpreted as dust ($p = 0$), the solutions belong to case Ib of Ellis,[2] who also provides an exact form. If however $\hat{\theta} \equiv 0$, it follows that $p = \mu + \Lambda$, $r = -(\partial_2 A_2 + \partial_3 A_3 + A_2^2 + A_3^2) = -2\omega^2 < 0$, and $\omega^2 = \mu + \Lambda$ in agreement with Eq. (3.1) of Stewart and Ellis,[12] and, for dust, with (4.24) of Ellis.[2] In this case p, μ, r, and ω are constants, and there is a G_5 isometry group acting transitively on space-time. In the case of dust, this is the well-known Gödel[13] solution [cases Ia, Ai, Bi, and Ci of Ellis[2]], and for nonzero pressure it is the "generalized Gödel" solution, apparently first discovered (for $\Lambda = 0$) by Raval and Vaidya,[33] and (for arbitrary Λ) by Stewart and Ellis[12] in their case Ia. These solutions, in which $\hat{\theta} \equiv 0$, arose in the search by Krasinski[14] for stationary perfect fluid solutions of Einstein's equations in which there is a Killing vector parallel to the vorticity.

4. GENERAL REMARKS

Ellis[2] has shown that if, in the case of shear-free dust, there are spacelike hypersurfaces $\{\mu = \text{constant}\}$, then the space-time is locally either a Gödel model ($\boldsymbol{\omega} \neq \mathbf{0}$) or FRW ($\boldsymbol{\omega} \equiv \mathbf{0}$). This result can be easily generalized to

Theorem 4.1 (cf. Theorem 5.3, Ellis[2]): Any shear-free perfect fluid in general relativity with an equation of state $p = p(\mu)$, such that $\mu + p \neq 0$, for which there are hypersurfaces $\{\mu = \text{constant}\}$ *transverse* to the fluid flow, and for which the acceleration is parallel to a nonzero vorticity vector, is necessarily a generalized Gödel model ($\theta \equiv \dot{u} \equiv \hat{\sigma} \equiv d \equiv \hat{\theta} \equiv 0$).

Proof: By Theorem 3.3 of I, it follows that under the assumptions of the present theorem, $\theta \equiv 0$. Since μ is constant on hypersurfaces transverse to the flow, we may propagate off these hypersurfaces using the Bianchi identity (A22) to obtain that μ is constant throughout space-time, in which case, by the assumption of an equation of state, so also is p, and hence, by (A23), $\dot{u} \equiv 0$. The fluid is therefore effectively dust, and the proof of Ellis[2] can be used from this point on. [Equation (A13) shows that $4\omega^2 = \mu + 3p - 2\Lambda$, and hence ω is identically constant. From (A1), (A14) and (A15), we have $\hat{\theta} \equiv d_2 \equiv d_3 \equiv 0$, and so the solutions are of type IIIGGi ($\hat{\sigma} \neq 0$) or IIIGGii ($\hat{\sigma} \equiv 0$). Suppose that $\hat{\sigma} \neq 0$. By (A16), $2\hat{\sigma}_{22}^2 = -\frac{1}{2}(\mu - p) - \Lambda$ and hence $\hat{\sigma}_{22}$ is constant; the difference between (A3) and (A17) shows that $A_2 = 0$, and similarly,

$A_3 = 0$. Then (A19) shows that $2\omega^2 = -\frac{1}{2}(\mu - p) - \Lambda$, implying that $\mu + p \equiv 0$, a contradiction. Thus type IIIGGi is inadmissible, and the only case in type IIIGGii in which $\hat{\theta} \equiv 0$ is the generalized Gödel model, as we saw in Sec. 3.] □

Remark. This theorem provides a characterization of the Gödel model, i.e., a uniqueness theorem for the Gödel model among the class of solutions being studied. We have shown its validity irrespective of the *spacelike* nature of the hypersurfaces on which μ is constant. Since locally there are always hypersurfaces on which μ is constant, it follows that the only cases excluded from consideration in the theorem are those in which there are hypersurfaces $\{\mu = \text{constant}\}$ to which the flow is *tangential*.

Ellis[2] has also shown that if for shear-free rotating dust there is a Killing vector independent of the flow vector $\mathbf{u}$ in the 2-surfaces spanned by $\mathbf{u}$ and ω, then the space-time either admits a G_3 isometry group of Bianchi type I acting transitively on timelike hypersurfaces, or admits a G_5 isometry group acting multiply transitively on space-time (this being the Gödel model). We can now re-derive and generalize this result, in the context of our present study. Thus we obtain

Theorem 4.2 (cf. Theorem 5.4, Ellis[2]): Consider a shear-free perfect fluid with an equation of state $p = p(\mu)$, satisfying $\mu + p \neq 0$, in which the acceleration, $\dot{\mathbf{u}}$, and vorticity, ω, are parallel (and possibly zero, but not simultaneously zero). Suppose that there is a Killing vector independent of $\mathbf{u}$ and tangent to the 2-surfaces spanned by $\mathbf{u}$, $\dot{\mathbf{u}}$ and ω. Then *either* $\dot{\mathbf{u}} \neq \mathbf{0}$ and $\omega \neq 0$, and the solutions admit a G_3 isometry group of Bianchi type II acting transitively on timelike hypersurfaces orthogonal to $\dot{\mathbf{u}}$ and ω (type IIIAGii with $\hat{\theta} = 0$ and hence $dp/d\mu = -\frac{1}{3}$) *or* $\dot{\mathbf{u}} \equiv \mathbf{0}$ and $\omega \neq \mathbf{0}$, in which case either there is a G_3 isometry group of Bianchi type I acting transitively on timelike hypersurfaces (type IIIGAii), or there is a G_5 isometry group acting multiply transitively on space-time (the generalized Gödel model; type IIIGGii with $\hat{\theta} \equiv 0$).

Proof: We see from Appendix B that in order to have two independent Killing vector fields in the $\mathbf{e}_0 - \mathbf{e}_1$ plane, it is necessary that $\theta \equiv \hat{\sigma}_{22} \equiv \hat{\theta} \equiv 0$. If $\omega \equiv 0$ and $\dot{u} \neq 0$, the only candidates are types IIAAii and IIAGii. However, in type IIAAii, $\hat{\theta} = -2\dot{u}$ [cf. Eq. (3.5)] and so $\hat{\theta} \neq 0$, whereas if $\hat{\theta} \equiv 0$ in type IIAGii, Eqs. (A13) and (A16) imply that $\mu + p \equiv 0$. If, on the other hand, $\dot{u} \neq 0$ and $\omega \neq 0$, type IIIAAii is ruled out by Eq. (3.5); the only viable alternative is type IIIAGii, and, as we have seen, $\hat{\theta} \equiv 0$ is possible only if $dp/d\mu = -\frac{1}{3}$. Finally, suppose that $\dot{u} \equiv 0$ and $\omega \neq 0$. Then the only possibilities are type IIIGAii (in which $\hat{\theta}$ automatically vanishes) and type IIIGGii, where, as discussed in Sec. 3, the requirement that $\hat{\theta} \equiv 0$ leads to the generalized Gödel model. The group properties claimed in the theorem can be immediately deduced by reference to the discussion in Sec. 3. □

Another result of Ellis[2] shows that if for rotating shear-free dust, μ is constant on the hypersurfaces orthogonal to ω, then the space-time admits an isometry group acting transitively on these hypersurfaces. That this is so follows readily from our discussion in Sec. 3, since if $\theta \equiv \dot{u} \equiv 0$ it follows that $4\omega^2 = \mu + 3p - 2\Lambda$, and differentiation of this along $\mathbf{e}_2$ and $\mathbf{e}_3$, with the assumptions $\partial_2\omega = \partial_3\omega = 0 \neq \omega$ requires $d = 0$; then we have types IIIGGi and IIIGGii, which both admit isometry groups that are transitive on the hypersurfaces orthogonal to ω. The fact that ω is indeed hypersurface-orthogonal follows since $n = 0$, which is also true if $\dot{\mathbf{u}} \neq \mathbf{0}$ and ω and $\dot{\mathbf{u}}$ are parallel (Proposition 2.1). This result of Ellis[2] does not generalize directly to the case $\dot{\mathbf{u}} \neq \mathbf{0}$, since types IIIAAi and IIIAAii provide counterexamples. However, in these types the matter obeys the unrealistic equation of state $\mu + 3p - 2\Lambda = 0$; moreover, in the class of models under consideration, they are the only exceptional cases, since space-time *is* locally homogeneous on the hypersurfaces orthogonal to ω (and $\dot{\mathbf{u}}$) in the case of the other candidates, viz., IIIAGi and IIIAGii. If in addition μ is constant along ω, then $\dot{\mathbf{u}}$ must be zero. We therefore obtain

Theorem 4.3 (cf. Theorem 5.2, Ellis[2]): Consider a rotating shear-free perfect fluid with an equation of state $p = p(\mu)$, satisfying $\mu + p \neq 0$ and $\mu + 3p \neq 2\Lambda$, in which the acceleration is parallel to the vorticity (and possibly zero). Then μ is constant on the hypersurfaces orthogonal to ω if and only if space-time is invariant under an isometry group acting transitively on these hypersurfaces. If in addition μ is constant along the vortex lines, then space-time is homogeneous.

Wainwright[34] has constructed a classification scheme for Petrov type D perfect fluid solutions. Since some of our solutions are of Petrov type D, it is of interest to classify them according to Wainwright's scheme. The LRS solutions are of Wainwright's class IA and the acceleration vector, if not zero, lies in the 2-space spanned by the principal null directions at each point. The only other admissible solutions are static, and belong to our types IIAAi and IIAGi; these are the static Barnes[11,22] class III and IV solutions, and are of class IC.[34] In the context of Wainwright's classification scheme, the static Barnes' solutions differ from the LRS solutions not only in regard to their subclass (i.e., IC as opposed to IA), but also in the fact that in these solutions the acceleration vector can not lie in the 2-space spanned by the principal null directions.

It is to be noted that in three of our types (viz., IIAAii, IIIAAi, and IIIAAii) the fluid necessarily satisfies the unrealistic equation of state $\mu + 3p - 2\Lambda = 0$; a proof was given in Proposition 2.5. While this equation of state is physically unreasonable, it is reminiscent of that in the well-known Wahlquist[35] solution; $\Lambda = 0$ and $\mu + 3p$ is constant. More particularly, there is a limiting case[16] of the Wahlquist[35] solution, due to Vaidya,[36] in which $\Lambda = 0$ and $\mu + 3p = 0$ (so that $\mu + 3p - 2\Lambda = 0$, as above). These solutions are of Petrov type D, the fluid flow is *shear-free*, expansion-free, rotating and accelerating, and the space-time admits an abelian G_2 isometry group acting on timelike orbits.[16,34,35] The solutions belong to Wainwright's class II, in which the fluid flow vector does not lie in the 2-space spanned by the principal null directions. They are the only class II solutions of which Wainwright[34] was aware, and the only static axisymmetric perfect fluid (not dust, and with no higher symmetries) solutions listed by Kramer, Stephani, MacCallum, and Herlt.[16] In these solutions, the acceleration and vorticity are not parallel, and therefore do not fall into the class presently under study.

Krasinski[14] has considered stationary perfect fluids with an equation of state $p = p(\mu)(\mu + p \neq 0)$ having

$\sigma \equiv \theta \equiv 0$ and $\omega \neq \mathbf{0}$, and with the additional restriction that there is a Killing vector parallel to the vorticity vector, ω. He arrived at three cases. In the first case, the only one in which $\dot{\mathbf{u}} \neq \mathbf{0}$, he obtained a new exact solution. In the other two cases, $\dot{\mathbf{u}} \equiv \mathbf{0}$, and the possibilities can be deduced from our discussion in Sec. 3, viz., types IIIGAi, IIIGAii, and IIIGGii (the generalized Gödel solution). In all of these cases, there is a G_3 isometry group of Bianchi type I that acts transitively on timelike hypersurfaces to which ω is tangential. However, type IIIGAi is ruled out, since it follows from (A42), (A43), and (A47) that in order for there to be a Killing vector parallel to ω, it is necessary that $\hat{\sigma} \equiv \hat{\theta} \equiv \hat{\Omega} \equiv 0$. In both types IIIGAii and IIIGGii there is indeed a Killing vector parallel to ω. Krasinski's first case is a new exact solution in which the vorticity and acceleration are *orthogonal*, and therefore does not belong to the class of solutions presently under discussion.

It is of interest to observe that in the 12 stationary types of solution, the space-times are necessarily *analytic*, provided that there is an analytic equation of state.[37] This means that, having determined locally a solution which admits an orthogonally transitive abelian G_2 isometry group with timelike orbits (e.g., in type IIIGAi), such a space-time must (provided that the equation of state is analytic) be extended analytically, so that in principle one could determine *uniquely* whether the initial region is locally part of a space-time which contains a regular rotation axis, and so is stationary and axisymmetric in the conventional sense.

In connection with shear-free solutions of Einstein's field equations, in which the source is a perfect fluid with equation of state $p = p(\mu)$, and in which $\mu + p \neq 0$, a number of outstanding problems clearly remains. Among these, the most pressing concern is, in our opinion, whether or not the Conjecture of Sec. 1 is true. Since this seems to be difficult to determine, a less ambitious pursuit would be to attempt to establish the truth (or falsity) of the Conjecture under a variety of additional special conditions (cf. I, and the many special results cited therein). Among these extra conditions could be the requirement that the vorticity be *orthogonal* to the acceleration (cf. the results of Krasinski[14] discussed above), or that the space-time admit certain types of isometry groups, such as an abelian G_3 with timelike orbits (as in the Krasinski[14] solution), or an abelian G_2 with timelike orbits (cf. the solutions of Wahlquist[35] and Vaidya[36]).

Regardless of whether the Conjecture is true (or can be proved to be true!), a thorough examination of the cases (i)–(iv) of Sec. 1 would be of interest, particularly with a view to obtaining exact solutions. Case (i) ($\sigma \equiv 0$, $\omega \equiv \mathbf{0}$, $\theta \neq 0$, $\dot{\mathbf{u}} \neq \mathbf{0}$) has been discussed herein, and the solutions are known in exact form.[4] Case (ii) ($\sigma \equiv \theta \equiv 0$, $\omega \equiv \mathbf{0}$, $\dot{\mathbf{u}} \neq \mathbf{0}$) has also been discussed in the present article. The solutions are all static, and the non-LRS Petrov type D ones are known exactly.[11,22] While there are many known LRS (and necessarily Petrov type D) solutions, such as those with static spherical symmetry, the most general solution is not known in exact form. We know of no exact static Petrov type I space-times, of which the most special (and hence presumably most amenable to exact solution) belong to our type IIAGi ($\theta \equiv \omega \equiv d \equiv 0$) and admit an abelian G_3 isometry group. In case (iii) ($\sigma \equiv \theta \equiv 0$, $\omega \neq \mathbf{0}$), we have examined in this article the special case in which $\dot{\mathbf{u}}$ is parallel to ω (and possibly zero). When $\dot{\mathbf{u}} \equiv \mathbf{0}$, the solutions are all known in exact form if $\hat{\sigma} d \equiv 0$ (corresponding to our types IIIGAii, IIIGGi, and IIIGGii; see Ref. 2), but the general type IIIGAi remains obscure. Even in the most special subcase, when type IIIGAi admits an abelian G_3 isometry group, there appears to be no known exact solution. When $\dot{\mathbf{u}}$ is nonzero and parallel to ω, there are four types; the most special (in terms of isometry groups and Petrov types) and also the most relevant (allowing a realistic equation of state) are type IIIAGi ($\theta \equiv d \equiv 0$), which is Petrov type D and admits a G_3 isometry group, and type IIIAGii ($\theta \equiv \hat{\sigma} \equiv d \equiv 0$), which is LRS (and hence Petrov type D), and admits a G_4 isometry group. The only exact solution of which we are aware in the entire class is the special case Ib of Stewart and Ellis,[12] for which $\theta \equiv \hat{\sigma} \equiv d \equiv \hat{\sigma} \equiv 0$ and $dp/d\mu = -\frac{1}{3}$, i.e., we know of no solutions admitting a physically reasonable equation of state. However, we have succeeded[38] in constructing a general metric form for all solutions of type IIIAGi and its various specializations (IIAGi, IIAGii, IIIAGi, IIIGGi, and IIIGGii), viz.,

$$ds^2 = -T^2(x^1)(dx^0 - y(x^2,x^3)dx^3)^2 + (dx^1)^2 + Y^2(x^1)(dx^2)^2 + Z^2(x^1)t^2(x^2)(dx^3)^2,$$

where, in the orthonormal tetrad

$$\mathbf{e}_0 = \frac{1}{T}\frac{\partial}{\partial x^0}, \quad \mathbf{e}_1 = \frac{\partial}{\partial x^1}, \quad \mathbf{e}_2 = \frac{1}{Y}\frac{\partial}{\partial x^2},$$

and

$$\mathbf{e}_3 = \frac{1}{tZ}\left(y\frac{\partial}{\partial x^0} + \frac{\partial}{\partial x^3}\right),$$

we have

$$\dot{u} = \frac{T'}{T}, \quad \omega = \frac{-T}{2YZ}\frac{\partial y}{\partial x^2}, \quad \hat{\sigma}_{22} = \frac{1}{2}\left(\frac{Y'}{Y} - \frac{Z'}{Z}\right),$$

$$\hat{\theta} = \frac{Y'}{Y} + \frac{Z'}{Z}, \quad A_2 = \frac{1}{tY}\frac{dt}{dx^2},$$

and

$$\theta = d_2 = d_3 = \hat{\Omega} = A_3 = 0.$$

Without loss of generality, we can make further specializations. In the non-LRS models, we put $t \equiv 1$; if $\omega \equiv 0$, we make $y \equiv 0$, whereas if $\omega \neq 0$, we put $y \equiv x^2$. If $\dot{u} \equiv 0$, we put $T \equiv 1$, and in the LRS models, $Y \equiv Z$, with $t(x^2)$ satisfying the equation $d^2t/d(x^2)^2 + Kt = 0$, where $K = -1, 0$ or $+1$. This coordinate system is equivalent to that of Barnes[11,22] for the Petrov type D solutions of our type IIAGi, to that of Ellis[2] for the type IIIGGi and type IIIGGii solutions, and to that of Stewart and Ellis[12] for the type IIAGii, type IIIAGii, and type IIIGGii solutions. Finally, there is case (iv) ($\sigma \equiv 0$, $\omega \neq \mathbf{0}$, $\theta \neq 0$, $\dot{\mathbf{u}} \neq \mathbf{0}$), and we know of no solutions in this class (indeed, the exhibition of one solution in case (iv) amounts to the provision of a counterexample to the Conjecture of Sec. 1). In addition to the above list of problems, the question of the global properties of the solutions remains outstanding, particularly with regard to whether or not any can represent reasonable stellar or cosmological models.

ACKNOWLEDGMENTS

This research was partially supported by an Individual Operating Grant (#A3978; C.B.C.) and by a Postgraduate Scholarship (A.J.W.), both from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Helen Warren for her diligent typing of a difficult manuscript.

APPENDIX A

Jacobi identities

$$\partial_1\omega = \omega[\dot{u} - \hat{\theta}], \tag{A1}$$

$$\partial_2 d_3 - \partial_3 d_2 - d_2 A_3 + d_3 A_2 = 0, \tag{A2}$$

$$\partial_1 A_2 + \partial_2\hat{\sigma}_{22} - \tfrac{1}{2}\partial_2\hat{\theta} - \partial_3\hat{\Omega} + \tfrac{1}{2}(d_2 + A_2)\hat{\theta} - (d_2 - A_2)\hat{\sigma}_{22} + \hat{\Omega}(d_3 - A_3) = 0, \tag{A3}$$

$$\partial_1 A_3 - \partial_3\hat{\sigma}_{22} - \tfrac{1}{2}\partial_3\hat{\theta} + \partial_2\hat{\Omega} + \tfrac{1}{2}(d_3 + A_3)\hat{\theta} + (d_3 - A_3)\hat{\sigma}_{22} - \hat{\Omega}(d_2 - A_2) = 0, \tag{A4}$$

$$\partial_0\omega = 0, \tag{A5}$$

$$\partial_0\hat{\theta} = -\tfrac{1}{3}\theta(\hat{\theta} - 2\dot{u}), \tag{A6}$$

$$\partial_0 d_2 = -\tfrac{1}{3}d_2\theta, \tag{A7}$$

$$\partial_0 d_3 = -\tfrac{1}{3}d_3\theta, \tag{A8}$$

$$\partial_0 A_2 = -\tfrac{1}{3}A_2\theta, \tag{A9}$$

$$\partial_0 A_3 = -\tfrac{1}{3}A_3\theta, \tag{A10}$$

$$\partial_0\hat{\Omega} = -\tfrac{1}{3}\hat{\Omega}\theta, \tag{A11}$$

$$\partial_0\hat{\sigma}_{22} = -\tfrac{1}{3}\hat{\sigma}_{22}\theta. \tag{A12}$$

Field equations

$$(00)\quad \partial_0\theta + \tfrac{1}{3}\theta^2 - 2\omega^2 - \partial_1\dot{u} - \dot{u}^2 - \dot{u}\hat{\theta} + \tfrac{1}{2}(\mu + 3p - 2\Lambda) = 0, \tag{A13}$$

$$(0\alpha)\quad \partial_2\omega = d_2\omega, \tag{A14}$$

$$\partial_3\omega = d_3\omega, \tag{A15}$$

$$(\alpha\beta)\quad \partial_1\hat{\theta} - \partial_2 d_2 - \partial_3 d_3 + 2\hat{\sigma}_{22}^2 + \tfrac{1}{2}\hat{\theta}^2 + d_2^2 + d_3^2 - d_2A_2 - d_3A_3 = \tfrac{1}{3}\partial_0\theta - \partial_1\dot{u} - \dot{u}^2 + \theta^2/3 - \tfrac{1}{2}(\mu - p) - \Lambda, \tag{A16}$$

$$\partial_1 A_2 - \partial_2\hat{\sigma}_{22} + \tfrac{1}{2}\partial_2\hat{\theta} - \partial_3\hat{\Omega} + d_3\hat{\Omega} - A_3\hat{\Omega} + \tfrac{1}{2}(d_2 + A_2)\hat{\theta} - (d_2 + 3A_2)\hat{\sigma}_{22} = -2d_2\dot{u}, \tag{A17}$$

$$\partial_1 A_3 + \partial_3\hat{\sigma}_{22} + \tfrac{1}{2}\partial_3\hat{\theta} + \partial_2\hat{\Omega} - d_2\hat{\Omega} + A_2\hat{\Omega} + \tfrac{1}{2}(d_3 + A_3)\hat{\theta} + (d_3 + 3A_3)\hat{\sigma}_{22} = -2d_3\dot{u}, \tag{A18}$$

$$\partial_1\hat{\sigma}_{22} + \tfrac{1}{2}\partial_1\hat{\theta} - \partial_2 d_2 + \partial_2 A_2 + \partial_3 A_3 + \hat{\sigma}_{22}\hat{\theta} + \tfrac{1}{2}\hat{\theta}^2 - d_3A_3 + A_2^2 + A_3^2 + d_2^2 = -\dot{u}(\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}) - \tfrac{1}{2}(\mu - p) - \Lambda + \theta^2/3 + \partial_0\theta/3 + 2\omega^2, \tag{A19}$$

$$-\partial_1\hat{\sigma}_{22} + \tfrac{1}{2}\partial_1\hat{\theta} - \partial_3 d_3 + \partial_2 A_2 + \partial_3 A_3 - \hat{\sigma}_{22}\hat{\theta} + \tfrac{1}{2}\hat{\theta}^2 - d_2A_2 + A_2^2 + A_3^2 + d_3^2 = \dot{u}(\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta}) - \tfrac{1}{2}(\mu - p) - \Lambda + \theta^2/3 + \partial_0\theta/3 + 2\omega^2, \tag{A20}$$

$$\partial_2 d_3 + \partial_3 d_2 - 4\hat{\Omega}\hat{\sigma}_{22} - 2d_2d_3 - d_2A_3 - d_3A_2 = 0. \tag{A21}$$

Equations (1.3), which were derived in the case $\dot{u} \neq 0$, but which also hold trivially in the case $\dot{u} \equiv 0$, have been employed in equations (A17) and (A18).

Bianchi identities

$$\partial_0\mu + (\mu + p)\theta = 0, \tag{A22}$$

$$\partial_1 p + (\mu + p)\dot{u} = 0, \tag{A23}$$

$$\partial_2 p = 0, \tag{A24}$$

$$\partial_3 p = 0. \tag{A25}$$

Commutation relations

$$[\mathbf{e}_0,\mathbf{e}_1] = \dot{u}\mathbf{e}_0 - (\theta/3)\mathbf{e}_1, \tag{A26}$$

$$[\mathbf{e}_0,\mathbf{e}_2] = -(\theta/3)\mathbf{e}_2, \tag{A27}$$

$$[\mathbf{e}_0,\mathbf{e}_3] = -(\theta/3)\mathbf{e}_3, \tag{A28}$$

$$[\mathbf{e}_1,\mathbf{e}_2] = -d_2\mathbf{e}_1 - (\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta})\mathbf{e}_2 + \hat{\Omega}\mathbf{e}_3, \tag{A29}$$

$$[\mathbf{e}_2,\mathbf{e}_3] = -2\omega\mathbf{e}_0 + A_3\mathbf{e}_2 - A_2\mathbf{e}_3, \tag{A30}$$

$$[\mathbf{e}_3,\mathbf{e}_1] = d_3\mathbf{e}_1 + \hat{\Omega}\mathbf{e}_2 - (\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta})\mathbf{e}_3. \tag{A31}$$

APPENDIX B

Components of Eq. (2.25) for $\boldsymbol{\xi} = \xi^a\mathbf{e}_a$:

$$\partial_0\xi^0 = -\dot{u}\xi^1, \tag{A32}$$

$$\partial_1\xi^0 = \dot{u}\xi^0, \tag{A33}$$

$$\partial_2\xi^0 = 2\omega\xi^3, \tag{A34}$$

$$\partial_3\xi^0 = -2\omega\xi^2, \tag{A35}$$

$$\partial_0\xi^1 = \tfrac{1}{3}\theta\xi^1, \tag{A36}$$

$$\partial_1\xi^1 = -\tfrac{1}{3}\theta\xi^0 + d_2\xi^2 + d_3\xi^3, \tag{A37}$$

$$\partial_2\xi^1 = -d_2\xi^1, \tag{A38}$$

$$\partial_3\xi^1 = -d_3\xi^1, \tag{A39}$$

$$\partial_0\xi^2 = \tfrac{1}{3}\theta\xi^2, \tag{A40}$$

$$\partial_1\xi^2 = (\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta})\xi^2 + \hat{\Omega}\xi^3, \tag{A41}$$

$$\partial_2\xi^2 = -\tfrac{1}{3}\theta\xi^0 - (\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta})\xi^1 - A_3\xi^3, \tag{A42}$$

$$\partial_3\xi^2 = -\hat{\Omega}\xi^1 + A_3\xi^2 + b, \tag{A43}$$

$$\partial_0\xi^3 = \tfrac{1}{3}\theta\xi^3, \tag{A44}$$

$$\partial_1\xi^3 = -\hat{\Omega}\xi^2 - (\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta})\xi^3, \tag{A45}$$

$$\partial_2\xi^3 = \hat{\Omega}\xi^1 + A_2\xi^3 - b, \tag{A46}$$

$$\partial_3\xi^3 = -\tfrac{1}{3}\theta\xi^0 + (\hat{\sigma}_{22} - \tfrac{1}{2}\hat{\theta})\xi^1 - A_2\xi^2. \tag{A47}$$

Compatibility requirements for (A 32)–(A 47):

$$\xi^a\partial_a\dot{u} = 0, \tag{A48}$$

$$\xi^0\partial_0\theta = 0, \tag{A49}$$

$$\xi^a\partial_a d_2 = bd_3, \tag{A50}$$

$$\xi^a\partial_a d_3 = -bd_2, \tag{A51}$$

$$\xi^a\partial_a\hat{\theta} = 0, \tag{A52}$$

$$\xi^a\partial_a\hat{\sigma}_{22} = 0, \tag{A53}$$

$$\xi^a\partial_a A_2 = \partial_3 b + A_3 b, \tag{A54}$$

$$\xi^a\partial_a A_3 = -\partial_2 b - A_2 b, \tag{A55}$$

$$\xi^a\partial_a\hat{\Omega} = \partial_1 b, \tag{A56}$$

$$\xi^a\partial_a\omega = 0, \tag{A57}$$

$$b\hat{\sigma}_{22} = 0, \tag{A58}$$

$$\partial_0 b = 0. \tag{A59}$$

APPENDIX C

In this appendix, we derive necessary and sufficient conditions for an abelian G_2 isometry group in types IIAAi, IIIAAi, and IIIGAi to be orthogonally transitive. In all three types, $\theta \equiv 0$ and $\hat{\sigma} \neq 0$, and in types IIIAAi and IIIGAi, $\omega \neq 0$. Since $\theta \equiv 0$, there is a Killing vector parallel to $\mathbf{e}_0$; suppose now that $\boldsymbol{\xi} = \xi^0 \mathbf{e}_0 + \xi \mathbf{e}_1 + \eta \mathbf{e}_2 + \nu \mathbf{e}_3$ is an independent Killing vector field (so that $\xi^2 + \eta^2 + \nu^2 \neq 0$). If $\eta \equiv \nu \equiv 0$, so that $\xi \neq 0$, Eqs. (A42) and (A47) show that $\hat{\sigma} = 0$, a contradiction, and hence $\eta^2 + \nu^2 \neq 0$. We construct the orthonormal tetrad $\mathbf{E}_a$ defined as follows:

$$\mathbf{E}_0 = \mathbf{e}_0,$$

$$\mathbf{E}_1 = \frac{(\eta^2 + \nu^2)\mathbf{e}_1 - \xi\eta\mathbf{e}_2 - \xi\eta\mathbf{e}_3}{(\eta^2 + \nu^2)^{1/2}(\xi^2 + \eta^2 + \nu^2)^{1/2}},$$

$$\mathbf{E}_2 = \frac{-\nu\mathbf{e}_2 + \eta\mathbf{e}_3}{(\eta^2 + \nu^2)^{1/2}},$$

and

$$\mathbf{E}_3 = \frac{\xi\mathbf{e}_1 + \eta\mathbf{e}_2 + \nu\mathbf{e}_3}{(\xi^2 + \eta^2 + \nu^2)^{1/2}},$$

so that the two Killing vector fields lie in the 2-space spanned by $\mathbf{E}_0$ and $\mathbf{E}_3$ at each point. In types IIAAi and IIIAAi it follows that $\xi \equiv 0$ since $\boldsymbol{\xi}(\mu) \equiv 0$, and in these cases $\dot{u} \neq 0$. By equations (A32), (A40) and (A44), we have that $\mathbf{E}_0$ and $\mathbf{E}_3$ (and hence the two Killing vectors) commute. We show that $\mathbf{E}_1$ and $\mathbf{E}_2$ are 2-surface forming if and only if $\xi = 0$ and $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{AB}$, and this completes the proof. The necessary and sufficient condition for $\mathbf{E}_1$ and $\mathbf{E}_2$ to be 2-surface forming is that $\mathbf{E}_0 \cdot [\mathbf{E}_1, \mathbf{E}_2] = \mathbf{E}_3 \cdot [\mathbf{E}_1, \mathbf{E}_2] = 0$. Now by direct calculation using (A31), we obtain $\mathbf{E}_0 \cdot [\mathbf{E}_1, \mathbf{E}_2] = 0 \Leftrightarrow \omega\xi(\eta^2 + \nu^2) = 0 \Leftrightarrow \omega\xi = 0$. But in types IIAAi and IIIAAi, $\xi \equiv 0$, as we have just shown, so $\mathbf{E}_0 \cdot [\mathbf{E}_1, \mathbf{E}_2] \equiv 0$, whereas in type IIIGAi, $\omega \neq 0$, so $\mathbf{E}_0 \cdot [\mathbf{E}_1, \mathbf{E}_2] = 0 \Leftrightarrow \xi = 0$. In this case, the above expressions for $\mathbf{E}_1$, $\mathbf{E}_2$, and $\mathbf{E}_3$ simplify considerably and a direct calculation using (A29), (A31), (A41) and (A45) shows that $\mathbf{E}_3 \cdot [\mathbf{E}_1, \mathbf{E}_2] = 0 \Leftrightarrow \eta\nu\hat{\sigma} = 0 \Leftrightarrow \eta\nu = 0$, since $\hat{\sigma} \neq 0$, as we observed previously. Now $\eta d_2 + \gamma d_3 = 0$, since in types IIAAi and IIIAAi, $\dot{u} \neq 0$, so $\boldsymbol{\xi}(\mu) \equiv 0 \Leftrightarrow \eta d_2 + \nu d_3 = 0$ using Eq. (1.3), whereas in type IIIGAi, $\boldsymbol{\xi}(\omega) \equiv 0 \Leftrightarrow \eta d_2 + \nu d_3 = 0$ by (A14) and (A15). Thus if $\eta = 0$ and $\nu \neq 0$, $d_3 = 0$, and so $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{AB}$ (and similarly if $\nu = 0$ and $\eta \neq 0$), whereas if $\mathbf{d}$ is an eigenvector of $\hat{\sigma}_{AB}$ then, without loss of generality, $d_3 = 0$ (and $d_2 \neq 0$ in the three types under consideration), so $\eta = 0$, implying $\mathbf{E}_3 \cdot [\mathbf{E}_1, \mathbf{E}_2] = 0$.

Note also that if $d_3 = 0$ and $d_2 \neq 0$, the requirement that $\eta = 0$ and $\nu \neq 0$ in (A41) and (A42) proves that $\hat{\Omega} = A_3 = 0$.

[1] A. J. White and C. B. Collins, J. Math. Phys. **25**, 332 (1984).
[2] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).
[3] G. F. R. Ellis, "Relativistic Cosmology," in *General Relativity and Cosmology*, Proceedings of the International School of Physics "Enrico Fermi" Course XLVII, 1969, edited by R. K. Sachs (Academic, New York, 1971), p. 104.
[4] C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[5] M. Wyman, Phys. Rev. **70**, 396 (1946).
[6] R. Mansouri, Ann. Inst. Henri Poincaré A **27**, 175 (1977).
[7] E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[8] B. Mashhoon and M. H. Partovi, Ann. Phys. (N.Y.) **130**, 99 (1980).
[9] D. C. Srivastava and S. S. Prasad, "Perfect Fluid Spheres in General Relativity," preprint, University of Gorakhpur, India (1981).
[10] A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[11] A. Barnes, Gen. Relativ. Gravit. **4**, 105 (1973).
[12] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[13] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[14] A. Krasinski, Acta Phys. Polon. **B5**, 411 (1974); **B6**, 223 (1975).
[15] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[16] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, England, 1980).
[17] G. F. R. Ellis, Gen. Relativ. Gravit. **2**, 7 (1971).
[18] G. F. R. Ellis and B. G. Schmidt, Gen. Relativ. Gravit. **8**, 915 (1977).
[19] W. Kundt and M. Trümper, Z. Physik **192**, 419 (1966).
[20] B. Carter, "Black Hole Equilibrium States," in *Black Holes*, Les Houches Lectures, 1972, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973), p. 57.
[21] M. A. H. MacCallum, "A Class of Homogeneous Cosmological Models," Ph.D. thesis, Cambridge University (1969).
[22] A. Barnes, J. Phys. **A5**, 374 (1972).
[23] W. J. Van Stockum, Proc. R. Soc. Edinburgh **A57**, 135 (1937).
[24] J. Winicour, J. Math. Phys. **16**, 1806 (1975).
[25] A. B. Evans, J. Phys. **A10**, 1303 (1977).
[26] K. A. Bronnikov, J. Phys. **A12**, 201 (1979).
[27] A. F. da F. Teixera, I. Wolk, and M. M. Som, Nuovo Cimento **B41**, 387 (1977).
[28] K. Lanczos, Z. Physik **21**, 73 (1924).
[29] J. P. Wright, J. Math. Phys. **6**, 103 (1965).
[30] B. Lukács, E. T. Newman, G. Sparling, and J. Winicour, Gen. Relativ. Gravit. **15**, 567 (1983).
[31] A. H. Taub, "Plane-symmetric Similarity Solutions for Self-Gravitating Fluids," in *General Relativity, Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Clarendon, Oxford, 1972), p. 133.
[32] A. F. da F. Teixera, I. Wolk, and M. M. Som, J. Phys. **A10**, 1679 (1977).
[33] H. M. Raval and P. C. Vaidya, Ann. Inst. Henri Poincaré **A4**, 21 (1966).
[34] J. Wainwright, Gen. Relativ. Gravit. **8**, 797 (1977).
[35] H. D. Wahlquist, Phys. Rev. **172**, 1291 (1968).
[36] P. C. Vaidya, Pramana **8**, 512 (1977).
[37] L. Lindblom, J. Math. Phys. **17**, 1810 (1976).
[38] A. J. White, "Shear-Free Perfect Fluids in General Relativity," M. Math. thesis, University of Waterloo, Waterloo, Ontario, Canada (1981).

Phase space of general relativity revisited: A canonical choice of time and simplification of the Hamiltonian

Abhay Ashtekar[a)]
Physics Department, Syracuse University, Syracuse, New York 13210 and Physique Théorique, IHP, Université de Paris VI, 75231 Paris, France[b)]

Gary T. Horowitz[c)]
Institute for Advanced Study, Princeton, New Jersey 08540

(Received 2 September 1983; accepted for publication 30 December 1983)

The phase space of general relativity is considered in the asymptotically flat context. Using spinorial techniques introduced by Witten, a prescription is given to transport rigidly the space-time translations at infinity to the interior of the (spatial) three-manifold. This yields a preferred four-parameter family of lapses and shifts and hence reduces the infinite-dimensional freedom in the choice of "time" to the restricted freedom available in special relativity. The corresponding Hamiltonians are computed and are shown to have an especially simple form: the Hamiltonians are "diagonal" in the (spatial) derivatives of variables which define "time." Furthermore, the Hamiltonians (generating timelike translations) are shown to be positive in a *neighborhood* of the constraint submanifold of the phase space, even at points at which the ADM energy is negative.

PACS numbers: 04.20.Fy, 04.60. + n, 11.10.Ef

1. INTRODUCTION

The multifingered nature of time in general relativity has been a major problem in the canonical quantization of the theory. In fact, in a recent review, Kuchař[1] has singled out this issue as the primary obstacle to canonical quantization. He summarizes the situation as follows. "Where we have a priviledged time, as in classical mechanics, relativistic particle theory in flat space-time, or free field theory in flat space-time, we know how to build the Hilbert space of states. When we are not sure what to use as time, as in relativistic particle theory and field theory in curved space-time, or in geometrodynamics, we lose our way.... I thus believe that the main problem which we face in an attempt to build quantum geometrodynamics is not a technical problem, but a conceptual one. It consists in reconciling the diametrically opposite ways in which relativity and quantum mechanics view the concept of time." General relativity by itself does not select a preferred time, in fact it insists that the choice can be essentially arbitrary. Quantum mechanics, on the other hand, knows how to make physically interesting predictions only when time is separated from the rest of the variables and treated as a c-number.

The purpose of this paper is to suggest a way in which the necessary reconciliation can be made in the context of asymptotically flat gravitational fields. More precisely, we shall see that it is possible to single out, among all possible lapse and shift fields, a preferred four-parameter family, thereby selecting the preferred time variables which can be eventually used in the quantum theory. However, this family is not universal: it depends on the gravitational Cauchy data. Thus, time is not introduced as an external parameter, insensitive to the choice of the gravitational field. Rather, it is a function of where one is in the gravitational phase space. This expresses the fact that we are dealing with general relativity rather than a field theory on a background geometry; the spirit of general relativity is not blatantly violated in the introduction of the extra structure required for quantum mechanics. What is more, the Hamiltonians generating evolutions along the preferred lapses and shifts have an especially simple form in terms of the spatial derivatives of the variables which define these lapses and shifts. In exactly soluble quantum models, the variables which simplify the Hamiltonian play a crucial role. One may therefore expect that quantization of the gravitational field would be simpler in terms of the new variables introduced here. There are many indications that this is the case. Furthermore, the canonical quantization program based on these variables is closely related to that involving Newman's[2] H-spaces and Penrose's[3] nonlinear gravitons. These issues, however, will be discussed elsewhere; this paper will deal only with classical general relativity.

The main ideas of this paper may be summarized as follows. Consider an asymptotically flat initial data set for general relativity. In terms of this data, one can write down an elliptic differential equation on a spinor field, the equation introduced by Witten[4] in his proof of the positive energy theorem. The equation enables one to transport rigidly constant spinors at infinity to the interior of the three-manifold (on which the initial data are defined.) The resulting spinor fields will be called the *Witten spinors*. The required preferred family of lapses and shifts can be constructed by "taking squares" of these Witten spinors; one can interpret them as the projections of the four-dimensional vector fields, obtained by transporting rigidly the space-time translations at infinity. Next, on the phase space of general relativity, one

[a)] Alfred P. Sloan Research Fellow. Supported in part by the NSF contract PHY80 08154 with Syracuse University.
[b)] Current address.
[c)] Albert Einstein Fellow. Address from September 1, 1983: Physics Department, University of California, Santa Barbara, CA 93106.

 0022-2488/84/051473-08$02.50

can compute Hamiltonians corresponding to these lapses and shifts. Although these Hamiltonians—like all others—have a complicated form (involving volume and surface integrals) in terms of the usual canonical variables, they are simply the volume integrals of squares of derivatives of the Witten spinors. In particular, the Hamiltonians generating Witten-time translations are manifestly positive. Thus, the squares of Witten spinors define "time" while the squares of their derivatives "diagonalize" the Hamiltonian. It is this intertwining that makes the present prescription interesting. Indeed, one can imagine other prescriptions to select preferred families of lapses and shifts; in essence, all one needs is an elliptic equation to rigidly transport translations from infinity. However, it seems unlikely that other prescriptions will select the preferred lapses and shifts *and* cause such a drastic simplification in the Hamiltonian in the same stroke.

Section 2 recalls the Witten equation and, under the assumption that it admits asymptotically constant solutions, introduces the preferred lapse-shift fields and the corresponding Hamiltonians. Section 3 discusses the issue of existence of solutions: using methods introduced by Reula,[5] we show that there exists a neighborhood of the constraint submanifold of the phase space for which the Witten equation admits solutions and that the solutions are uniquely determined by their asymptotic values. (The ADM energy is *not* everywhere positive in this neighborhood.) In broad terms, what we have done in this paper is to use certain techniques asssociated with the Witten equation to obtain new insights into the Hamiltonian description of general relativity.

A. Sen, J. Nester, and G. A. J. Sparling have independently obtained results similar to those contained in Sec. 2C. However, since canonical quantization is not the main concern in these analyses, their motivation and the overall viewpoint is different from ours. In particular, the problem of separating gauge and dynamics and the use of the Witten equation to rigidly transport lapses and shifts do not play a central role in these works.

2. GRAVITATIONAL PHASE SPACE AND WITTEN SPINORS

This section is divided into three parts. The first recalls the phase space description of general relativity, fixes notation and formulates the problem. The second introduces the preferred family of lapses and shifts under the assumption that the solutions to the Witten equation exist and are uniquely determined by their asymptotic (constant) values. The third analyzes the Hamiltonians corresponding to these lapses and shifts.

A. Preliminaries

Fix a C^∞ three-manifold Σ, the complement of a compact set of which is diffeomorphic with the complement of a ball in R^3. Denote by $\mathscr{C}$ the space of positive definite metrics q_{ab} on Σ, with respect to which Σ is complete, and which are asymptotically flat in a suitable sense. (Since the existence theorems of Sec. 3 hold under quite weak fall-off conditions, we shall not tie ourselves down to a specific choice here.) The phase space Γ consists of pairs, (q_{ab}, P^{cd}), where P^{cd} is a symmetric tensor density of weight 1, also subject to a suitable asymptotic condition. Not all points of Γ are accessible to the physical gravitational field: there are constraints. The constraint submanifold $\bar{\Gamma}$ of Γ consists of points (q_{ab}, P^{cd}) satisfying the constraint equations

$$C^a \equiv \mathscr{D}_b (q^{-1/2} P^{ab}) = 0, \tag{2.1}$$

$$C \equiv R - (q_{ac} q_{bd} - \tfrac{1}{2} q_{ab} q_{cd})\, q^{-1} P^{ab} P^{cd} = 0, \tag{2.2}$$

where $\mathscr{D}$, R, and q are, respectively, the derivative operator, the scalar curvature, and the determinant of q_{ab}. Every point (q,P) of $\bar{\Gamma}$ provides an initial datum for Einstein's equation, q_{ab} being the intrinsic three-metric and $\pi^{ab} := q^{-1/2}(P^{ab} - \frac{1}{2} P q^{ab})$ being the extrinsic curvature of Σ, where $P = P^{ab} q_{ab}$ is the trace of P^{ab}.

To define the Hamiltonian we need to introduce the notion of time. In the four-dimensional space-time picture, this corresponds to introducing a foliation by spacelike three-manifolds (given by $t = \text{const}$, say) and an evolution vector field ξ^a which "relates one leaf of the foliation to the next one." Such a vector field can be projected into and perpendicular to the leaves: $\xi^a = T t^a + T^a$, where t^a is the future directed unit normal to the leaves and $t_a T^a = 0$. The fields T and T^a refer only to the spatial three-manifolds and can be, therefore, carried over to the phase-space framework; T is the *lapse* field and T^a is the *shift*. Given any pair (T,T^a) (or, in the four-dimensional picture, the vector field ξ^a) on Σ, one can define a Hamiltonian $H_{T,T^a}(q,P)$ on Γ:

$$\begin{aligned} H_{T,T^a}(q,P) = \frac{-1}{16\pi} \int_\Sigma (TC + 2T_a C^a) dV_q \\ + \frac{1}{16\pi} \oint_S T(\partial_a q_{bc} - \partial_b q_{ac}) f^{ac} dS^b \\ + \frac{1}{8\pi} \oint_S T^a q^{-1/2} P_{ab}\, dS^b, \end{aligned} \tag{2.3}$$

where S is the (two-sphere) boundary of Σ at spatial infinity, f^{ac} a flat metric on Σ in a neighborhood of infinity, and ∂_a the derivative compatible with f^{ab}.[6] This Hamiltonian is a differentiable function on Γ provided T and T^a are asymptotically constants, i.e., ξ^a is an asymptotic translation. *From now on, we shall use the symbols T and T^a only to denote such fields.* Since H_{T,T^a} is differentiable, it generates canonical transformations on Γ. The restriction to $\bar{\Gamma}$ of these transformations gives us the required evolution: they correspond to the changes undergone by q_{ab} and P^{ab} under the diffeomorphism generated by ξ^a in the four-dimensional picture.

The multifingered nature of time expesses itself in the infinite-dimensional freedom available in the choice of T and T^a even when their asymptotic values—i.e., the asymptotic translation to which they correspond—are fixed. Let (T,T^a) and $(\tilde{T},\tilde{T}^a)$ agree asymptotically[7] and set $T - \tilde{T} = N$ and $T^a - \tilde{T}^a = N^a$. Then, since N and N^a vanish asymptotically, the Hamiltonian to which they correspond is given by

$$H_{N,N^a} = \frac{-1}{16\pi} \int_\Sigma (NC + 2N^a C_a) dV_q\,; \tag{2.4}$$

it is a constraint function. In the Dirac theory of constrained systems, constraint functions generate gauge. Hence, the canonical transformations generated by H_{N,N^a}—which correspond, in the four-dimensional picture, to the evolution

along the asymptotically zero vector field $Nt^a + N^a$—can be considered as gauge. Thus, the multifingered nature of time expresses itself in the Hamiltonian framework via gauge ambiguities: one can always add to the generator of a translation a generator of a gauge transformation and obtain a generator of the same asymptotic translation. Let us consider two lapse-shift pairs—or, alternatively, two Hamiltonians—as equivalent if they correspond to the same asymptotic translation. Then, we have a four parameter family of equivalence classes. To reduce the infinite-dimensional freedom in the choice of lapses and shifts to the restricted freedom of special relativity, we need to give a prescription to choose, from each equivalence class, a preferred element. That is, we need to give a "gauge fixing prescription" for lapses and shifts, or, for Hamiltonians. Once this is done, one would be able to decompose canonically any Hamiltonian $H_{T,T^a}(q,P)$ into two parts, one generating a "pure" translation, and the other, "pure" gauge. In the transition to quantum theory, one can then concentrate on generators of pure translations and e.g., look for variables which diagonalize, or, more generally, simplify them.

Remarks: (i) On the constraint submanifold $\bar{\Gamma}$, C and C^a of Eq. (2.1) and (2.2) vanish, and hence, the numerical value of $H_{T,T^a}(q,P)$ is given just by the surface terms of Eq. (2.3). Note, however, that, even if one wishes to reproduce the dynamical trajectories just on $\bar{\Gamma}$, one cannot just work with these surface terms; the entire expression of Eq. (2.3) is needed. This is because equations of motion involve *derivatives* of the Hamiltonian and it turns out that the dynamical trajectories *within* $\bar{\Gamma}$ crucially depend upon the derivatives of H_{T,T^a} off $\bar{\Gamma}$.

(ii) Since Einstein's equation determines dynamics only on $\bar{\Gamma}$, there is some ambiguity in the expression of the Hamiltonian. Consider another candidate, $H'_{T,T^a} = H_{T,T^a}(q,P) + f_{T,T^a}(q,P)$. Since the value of $H_{T,T^a}(q,P)$ on $\bar{\Gamma}$,i.e., the surface terms in Eq. (2.3) yields the correct expression of the component of the ADM four-momentum along the asymptotic translation defined by (T,T^a), and since this is a desirable property of any Hamiltonian associated to (T,T^a), $f(q,P)$ should vanish on $\bar{\Gamma}$. Furthermore, if H'_{T,T^a} is to yield the correct dynamical trajectories on $\bar{\Gamma}$, the derivatives of $f(q,P)$ must also vanish on $\bar{\Gamma}$. Thus, the most general permissible expression of $f(q,P)$ is of the type

$$f_{T,T^a}(q,P) = g_{T,T^a}(q,P)\int_\Sigma (NC + N^aC_a)dV_q \times \int_\Sigma (\tilde{N}C + \tilde{N}^aC_a)dV_q$$

for some function $g(q,P)$ on Γ and some choice of lapse-shift pairs (N,N^a) and $(\tilde{N},\tilde{N}^a)$.[8] In spite of this ambiguity, throughout the literature, it is $H_{T,T^a}(q,P)$ that has been used as the Hamiltonian because it has many attractive properties. For example, if we set $T=0$, it generates canonical transformations which correspond, on all of Γ, to diffeomorphisms within Σ generated by T^a. This property is desirable from just geometric grounds, without any appeal to field equations and can be satisfied by H'_{T,T^a} only if $g_{0,T^a}(q,P)=0$ for all T^a. A second and more important property comes from the study of the modifications required to incorporate matter fields (as sources of the gravitational field). $H_{T,T^a}(q,P)$ has the property that we only have to add to it the Hamiltonian of the matter field (propagating on a background geometry) to obtain a Hamiltonian which yields the correct dynamical trajectories for the coupled Einstein-matter system. This property is not shared by other Hamiltonians H'_{T,T^a} (except perhaps for very special choices of matter fields.) In this paper, we shall use $H_{T,T^a}(q,P)$ only.

B. The Witten equation and the preferred lapses and shifts

Fix a point (q_{ab},P^{cd}) of the phase space. We shall work with the SU(2) spin structure defined by the metric q_{ab}. The spinors will carry upper case latin indices; e.g., λ^A and μ_B denote spinor fields. At any point of Σ, contravariant and covariant spinors each form a two-dimensional complex vector space. These spaces are naturally equipped with second-rank skew spinors ϵ^{AB} and ϵ_{AB}, respectively, which enable one to raise and lower the indices. Associated with each spinor λ^A is its Hermitian conjugate $(\lambda^A)^+ \equiv \lambda^{+A}$ such that $((\lambda^A)^+)^+ = -\lambda^A$ and $\lambda^{+A}\lambda_A \geqslant 0$, equality holding if and only if $\lambda_A = 0$. Finally, there is an isomorphism σ_a^{AB} between the complexified tangent space at any point of Σ and the three-dimensional complex vector space of symmetric, second-rank spinors at that point: the vector S^a corresponds to the symmetric spinor $S^{AB} = :\sigma_a^{AB}S^a$. Under this isomorphism, the real vectors correspond to Hermitian spinors, i.e., spinors S^{AB} satisfying $(S^{AB})^+ = S^{AB}$. Thus, $i\lambda^{+(A}\lambda^{B)}\sigma_{AB}^a$ is a real vector. The metric q_{ab} is related to ϵ_{AB} by $q_{am} = -\sigma_a^{AB}\sigma_m^{MN}\epsilon_{A(M}\epsilon_{N)B}$. Since, in what follows, calculations will involve (an arbitrarily chosen but) fixed metric—and hence, a fixed isomorphism between the symmetric second-rank spinors and the complexified tangent vectors—we shall suppress the symbol σ_a^{AB} and simply write $S^a = S^{(AB)}$, $q_{am} \equiv q_{(AB)(MN)} = -\epsilon_{A(M}\epsilon_{N)B}$, etc.

To introduce Witten's equation let us first define a derivative operator $D_a = D_{(AB)}$ on the SU(2) spinors:

$$\begin{aligned} D_{AB}\lambda_C &:= \mathscr{D}_{AB}\lambda_C + [i/(2q)^{1/2}] \\ &\quad \times (P_{ABC}{}^D + \tfrac{1}{2}P_m{}^m\epsilon_{C(A}\epsilon_{B)}{}^D)\lambda_D \\ &\equiv \mathscr{D}_{AB}\lambda_C + (i/\sqrt{2})\pi_{ABC}{}^D\lambda_D\,, \end{aligned} \tag{2.5}$$

where $\mathscr{D}_{AB}$, P_{ABCD}, and π_{ABCD} are the spinorial forms of the derivative operator $\mathscr{D}_a$ compatible with q_{ab}, of the momentum P_{ab}, and of the extrinsic curvature π_{ab}. Witten's equation can now be written as

$$D_{AB}\lambda^B = 0. \tag{2.6}$$

To define the preferred four-parameter family of lapses and shifts, we proceed as follows. Fix any constant spinor $\mathring{\lambda}^A$ in a neighborhood of infinity of Σ (i.e., a spinor field $\mathring{\lambda}^A$ defined near infinity, satisfying $\partial_a\mathring{\lambda}^A = 0$ where ∂_a is the derivative operator compatible with a fixed flat metric f_{ab} to which q_{ab} approaches[9]). We shall see in Sec. 3 that (at least) when the point (q_{ab},P^{cd}) is chosen to lie in a suitable neighborhood of the constraint submanifold of the phase space, Eq. (2.6) admits a unique solution λ^A which asymptotically approaches $\mathring{\lambda}^A$. That is, Eq. (2.6) enables one to transport constant spinors at infinity to the interior of Σ and the rotation of the

spinor at infinity causes a rigid rotation on the entire spinor field. Consequently, there is available on Σ a distinguished, complex two-dimensional vector space of asymptotically constant spinor fields.[10] With each of these spinor fields λ^A we shall associate a lapse-shift pair (T,T^a) given by $T=\lambda^{+A}\lambda_A$ and $T^a = -\sqrt{2}i\lambda^{+(A}\lambda^{B)}$. (In the four-dimensional picture, this pair defines a null vector field which is an asymptotic translation.) Let α^A and β^A be two linearly independent Witten transported spinor fields. Then, by consecutively substituting α^A, β^A, $(\alpha^A+\beta^A)$, and $(\alpha^A+i\beta^A)$ for λ^A in the above prescription, one obtains four pairs, $(T_{(k)}, T^a_{(k)})$ with $(k)=1,2,3,4$, of lapse-shift pairs each of which defines an asymptotic null translation. Consider the real four-dimensional[10] vector space generated by these pairs. This space is independent of the initial choice of α^A and β^B. [*Proof*: Set $\tilde{\alpha}^A = a\alpha^A + b\beta^A$ and $\tilde{\beta}^A = c\alpha^A + d\beta^A$. It is easy to check that the resulting $(\tilde{T}_{(k)}, \tilde{T}^a_{(k)})$ are expressible as constant linear combinations of $(T_{(k)}, T^a_{(k)})$.] This is the preferred family of lapses and shifts. Each element of this family defines an asymptotic translation and is, in turn, completely determined by this translation. [*Proof*: Use the basis $(T_{(k)}, T^a_{(k)})$.] Thus, we have here a prescription for rigid transport of the asymptotic translations to the interior of Σ.

Remarks: (i) Although the transport prescription given above is unambiguous, the reason behind the choice of the specific expressions of T and T^a in terms of the Witten spinors is somewhat obscure in the SU(2) framework. To see that the choice follows naturally from the use of spinors, it is more convenient to use the SL(2,C) spinors.[11] Fix a space-time (M, g_{ab}) and an asymptotically flat spacelike submanifold Σ therein. Denote by $q_a{}^b$ the projection operator associated with Σ. Set $\mathbf{D}_a = q_a{}^b\nabla_b$, where ∇ is the derivative operator compatible with the four-metric g_{ab}. Then, in the SL(2,C) language, the Witten equation becomes

$$\mathbf{D}_{AA'}\lambda^A = 0. \qquad (2.6)'$$

As before, provided the (q_{ab},P^{cd}) pair defined by g_{ab} on Σ lies in a neighborhood of the constraint surface $\bar{\Gamma}$ of the phase space, (2.6)′ admits a unique solution, given the boundary value $\mathring{\lambda}^A$ at infinity. This provides an isomorphism between the SL(2,C) spin space at infinity and the SL(2,C) spin space at any point of Σ which, in turn, defines an isomorphism between the space of (four-dimensional) asymptotic translations and the four-dimensional tangent space at any point in Σ. That is, we can rigidly transport translations at infinity to the interior of Σ. The simplicity of this procedure is somewhat lost in the SU(2) framework because of the necessity of doing a 3 + 1 decomposition.[11] The reason why we work with the SU(2) spinors in the main part of the discussion is that they fit in more naturally with the phase-space formulation.

(ii) Fix a spin dyad at infinity. Equation (2.6)′ then provides us with a spin dyad everywhere on Σ. Alternatively, in terms of vectors, given a tetrad at infinity, we obtain a tetrad field everywhere on Σ. If we rotate the tetrad at infinity, the entire field rotates rigidly by the same amount; the freedom is that of global rather than local Lorentz transformations. It is in this sense that we have a "gauge fixing procedure." Note, however, that the preferred tetrad fields depend on the choice of the variables $(q_{ab},P_m{}^m)$ on Σ: if we change the metric g_{ab} near Σ, the tetrad fields change. [The fields are insensitive to the trace-free part of P^{ab} because only the trace enters the Witten equations (2.6) and (2.6)′.]

(iii) Since Eq. (2.6) is linear and since the solution λ^A is determined everywhere by the boundary value $\mathring{\lambda}^A$, we have an isomorphism[10] Ψ_p from the vector space of constant spinors at infinity to the vector space of spinors at any point p of Σ. Denote by $\mathring{\epsilon}_{AB}$, the natural two-form on the asymptotic spin space. Its image $\Psi_p(\mathring{\epsilon}_{AB})$ is a two-form in the spin space at p, and hence, proportional to $\epsilon_{AB|p}$; say $\epsilon_{AB|p} = R\Psi_p(\mathring{\epsilon}_{AB})$. Thus, Eq. (2.6)′ defines a complex function R on Σ such that for any two solutions λ^A and μ^A with boundary values $\mathring{\lambda}^A$ and $\mathring{\mu}^A$ at infinity, $\epsilon_{AB}\lambda^A\mu^B = R\mathring{\epsilon}_{AB}\mathring{\lambda}^A\mathring{\mu}^B$. Next, note that the isomorphism Ψ_p extends uniquely to tangent vectors: Given any asymptotic translation on Σ (or, a vector in the tangent space of the point i^0 at spatial infinity), say T^a_0, Ψ_p provides us a vector T^a in the tangent space of p. The relation between $\epsilon_{AB|p}$ and $\Psi_p(\mathring{\epsilon}_{AB})$ now implies that, for any two of such vector fields, T^a and V^a with asymptotic values $\mathring{T}^a$ and $\mathring{V}^a$, we have $V\cdot T = |R|^2\,\mathring{V}\cdot\mathring{T}$. Thus, if $\mathring{T}^a$ is timelike (respectively, null, spacelike) at infinity, T^a is timelike (respectively, null, spacelike) everywhere.

(iv) Let (M, g_{ab}) be Minkowski space. Then, since the constant spinor fields in (M, g_{ab}) automatically satisfy (2.6)′ for any choice of Σ, the transport of translations at infinity yields the translational Killing fields everywhere on M. In a generic space-time, however, the transport is tied to the choice of Σ. Thus, it is only when we are given a foliation of a generic space-time that we can obtain four vector fields everywhere on the space-time, and they depend on the choice of the foliation. The transport is well suited to the canonical framework, however, because in this framework one deals only with three-surfaces.

C. Hamiltonian corresponding to preferred lapses and shifts

Let us now combine the results of Secs. 2A and 2B. For simplicity, let us consider the lapse-shift pair (T,T^a) corresponding to the "pure time translation at infinity," i.e., satisfying the boundary conditions $\mathring{T}=1$ and $T^a=0$. As we saw above, (T,T^a) is guaranteed to correspond to a future-directed timelike four-vector, i.e., to satisfy $T>0$, $T^2>T^aT_a$. [In general, $T^a\neq 0$ except at infinity. It is not difficult to show that $T^a=0$ everywhere iff (q_{ab},P^{cd}) satisfies $P^{ab}q_{ab}=0$.] This pair (T, T^a) will be referred to as the *Witten-time translation*. Let us now substitute for the lapse and the shift in Eq. (2.3) the Witten-time translation. The resulting Hamiltonian $H_W(q,P)$ is

$$H_W(q,P) = \frac{-1}{16\pi}\int_\Sigma (T(q,P)C(q,P) + 2T_a(q,P)C^a(q,P))dV_q + E_{\rm ADM}, \qquad (2.7)$$

where

$$E_{\rm ADM} = \frac{1}{16\pi}\oint_S (\partial_a q_{bc} - \partial_b q_{ac})f^{ac}\,dS^b$$

is the ADM energy. Note that, even though the lapse and the shift now themselves depend on q_{ab} and $P^m{}_m$, this Hamil-

tonian generates the correct dynamics on $\bar{\Gamma}$: The Hamiltonian vector field evaluated at a point (q,P) of $\bar{\Gamma}$ yields the infinitesimal change in that (q,P), dictated by Einstein's equation, caused by a diffeomorphism corresponding to $(T(q,P), T^a(q,P))$. This comes about because, in the expression of the Hamiltonian, $T(q,P)$ and $T^a(q,P)$ are multiplied by quantities which vanish on $\bar{\Gamma}$.

Expression (2.7) of the Hamiltonian simplifies considerably if we substitute for $T(q,P)$ and $T^a(q,P)$ in terms of Witten spinors, i.e., solutions to (2.6). Note, first, that any pair $(\mathring{T}, \mathring{T}^a)$ can be expressed as $(\mathring{T}, \mathring{T}^a) = (\mathring{\alpha}^{+A}\,\mathring{\alpha}_A + r\mathring{\beta}^{+A}\mathring{\beta}_A, -\sqrt{2}\,i(\mathring{\alpha}^{+(A}\mathring{\alpha}^{B)} + r\mathring{\beta}^{+(A}\mathring{\beta}^{B)})$ for some constant spinors $\mathring{\alpha}^A$ and $\mathring{\beta}^A$ and real number r. For $T=1$, $T^a=0$, it suffices to choose for $\mathring{\alpha}^A$ any spinor normalized so that $\mathring{\alpha}^{+A}\mathring{\alpha}_A = \frac{1}{2}$, and so set $\mathring{\beta}^A = \mathring{\alpha}^{+A}$, and $r=1$. Let us make this choice. Denote, as before, by α^A and β^A the Witten spinors—i.e., the solutions to (2.6)—which asymptotically tend to $\mathring{\alpha}^A$ and $\mathring{\beta}^A$. (Note that, in general, $\beta^A \neq \alpha^{+A}$ except at infinity.) Then, the Hamiltonian generating the Witten-time translation becomes[6]

$$H_W(q,P) = \frac{-1}{16\pi}\int_\Sigma [(\alpha^{+A}\alpha_A + \beta^{+A}\beta_A)C - 2\sqrt{2}\,i(\alpha^{+A}\alpha^B + \beta^{+A}\beta^B)C_{AB}] \times dV_q + E_{\rm ADM}\,. \tag{2.8}$$

Now, *any* spinor field λ^A on Σ satisfies the identity[4]

$$\begin{aligned}\frac{1}{4\pi}\int_\Sigma (D^{AB}\lambda^C)^+(D_{AB}\lambda_C)dV_q &= \frac{-1}{16\pi}\int_\Sigma (\lambda^{+A}\lambda_A)\,C - (2\sqrt{2}\,i\lambda^{+A}\lambda^B)C_{AB}dV_q \\ &+ \frac{1}{2\pi}\int_\Sigma \lambda^{+A} D_{AB}^{+} D^B{}_C\lambda^C dV_q \\ &+ \frac{1}{4\pi}\oint_S \lambda^{+B}(D_a\lambda_B)\,dS^a\end{aligned} \tag{2.9}$$

provided all the integrals exist. If we substitute α^A and β^A for λ^A in this identity, the second term on the right vanishes by the Witten equation and the sum of the surface terms yields[4]

$$\frac{1}{4\pi}\oint_S (\alpha^{+B}D_a\alpha_B + \beta^{+B}D_a\beta_B)\,dS^a = E_{\rm ADM}\,. \tag{2.10}$$

Hence, we have

$$H_W(q,P) = \frac{1}{4\pi}\int_\Sigma [(D^{AB}\alpha^C)^+(D_{AB}\alpha_C) + (D^{AB}\beta^C)^+(D_{AB}\beta_C)]\,dV_q \tag{2.11}$$

which is the desired result. To summarize, (at least) in a neighborhood of the constraint submanifold $\bar{\Gamma}$ of the phase space, one can introduce Witten spinors to select preferred lapses and shifts. The Hamiltonian generating the Witten-time translation—which is asymptotically a pure time translation, with $\mathring{T}=1$ and $\mathring{T}=0$—has an especially simple form in terms of the (spatial) derivatives of the Witten spinors.

Remarks: (i) It is only for concreteness that we have restricted ourselves to the Witten-time translation. Given any asymptotic null translation with $T = \lambda^{+A}\lambda_A$, $T_a = -\sqrt{2}\,i\lambda^{+(A}\lambda^{B)}$, the corresponding Hamiltonian is

$$H_{(T,T^a)}(q,P) = \frac{1}{4\pi}\int_\Sigma (D^{AB}\lambda^C)^+(D_{AB}\lambda_C)dV_q\,. \tag{2.12}$$

The Hamiltonian corresponding to any preferred lapse shift can be obtained by superposing these generators of null translations. Note that right side of (2.12) is manifestly non-negative, and vanishes iff $D_{AB}\lambda_C = 0$, which can occur only if (q_{ab}, P^{cd}) corresponds to Cauchy datum for Minkowski space.

(ii) The mathematics leading from (2.7) to (2.11) is exactly the same as that in the positive energy theorems where one shows the positivity of the surface term in (2.7), given that C and C^a satisfy the dominant energy condition. Our motivation here is however, different: we are interested in the structure of the phase space and the Hamiltonians. Also, note that the Hamiltonian has the form (2.11) as long as asymptotically constant solutions to the Witten equation exist, we need not tie ourselves down to the constraint surface. In fact, we will see in Sec. 3 that the required solutions can exist even when the energy condition is violated: even when the energy is negative, H_W can be positive. Thus, positivity of H_W is a more general result than the positivity of energy.

(iii) The form of $H_W(q,P)$ given in Eq. (2.11) is similar to the form $H = Z_1\bar{Z}_1 + \cdots + Z_n\bar{Z}_n$ of the Hamiltonian of the n-dimensional oscillator, with $Z_k = P_k + iq_k$. This suggests a quantization method using a Bargmann-type representation. This can be carried out in the weak-field limit in a straightforward manner. The situation is more complicated in the exact theory. This is because, whereas the Poisson bracket between Z_i and Z_j is a constant, the analogous Poisson bracket in the gravitational case is a constant function on Γ only in the weak field limit. In otherwords, (2.11) is not a true diagonalization in the exact theory. It does, nonetheless, suggest an approach to quantization which will be discussed in a subsequent paper.

3. EXISTENCE AND UNIQUENESS OF WITTEN SPINORS

In this section we shall obtain the results on the existence and the uniqueness of solutions to Eq. (2.6) on which the discussion of Sec. 2B and 2C was based. The main idea is to combine the techniques introduced by Reula,[5] to show the existence and uniqueness of the Witten spinors in the case when the dominant energy condition is satisfied, and the conformal invariance of the Witten equation to obtain a generalization of Reula's results.

We shall first show that Reula's results hold even when the energy condition is mildly violated. Although the proof requires only minor modifications of Reula's work, we present it in some detail for completeness as well as because, as was pointed out to us by Reula, his published proof[5] is incomplete because it uses an incorrect Sobolev inequality.

Let us begin with some preliminaries. Consider a neighborhood N_1 of the constraint submanifold $\bar{\Gamma}$ of Γ such that for all (q_{ab}, P^{cd}) in N_1, the following conditions hold[12]:

(i) Given (q_{ab}, P^{cd}), there exists a flat metric f_{ab} on the complement $\Sigma - K$ of a compact set K of Σ such that $(\Sigma - K, f_{ab})$ is isometric with the Euclidean space minus a ball and such that there exists a $C_0 > 0$ for which $C_0^{-1} f_{ab}v^av^b \leqslant q_{ab}v^av^b \leqslant C_0 f_{ab}v^av^b$ for all v^a;

(ii) $|\partial_a q_{bc}|^2$, $q^{-1}P^{ab}P_{ab}$, $|C^a|$, and C [of Eqs. (2.1) and (2.2)] are all integrable on (Σ, q_{ab}), where, as before, ∂_a is the derivative operator compatible with f_{ab}. One expects that any (q_{ab}, P^{cd}) in Γ will, by virtue of asymptotic conditions, automatically satisfy all conditions except possibly the integrability of $|C^a|$ and C; it is to ensure this integrability that we introduce the subspace N_1. Next, fix a pair (q_{ab}, P^{ab}) in N_1 and consider the SU(2) spinors corresponding to q_{ab}. Then, for all C^∞-spinors λ^A with compact support ($\lambda_A \in C_0^\infty$) the following inequality holds[13]:

$$\int_\Sigma (D^{AB}\lambda^C)^+(D_{AB}\lambda_C)\, dV_q \geq k \int_\Sigma \frac{\lambda^{+A}\lambda_A}{1+d^2}\, dV_q \tag{3.1}$$

for some (positive) constant k, where d is the distance function on (Σ, q_{ab}) with respect to an arbitrarily chosen origin. Note that the constant k depends on the choice of the pair (q_{ab}, P^{cd}). In terms of this k, we can define a subset in N_2 of N_1, which includes $\overline{\Gamma}$, as follows (q_{ab}, P^{cd}) in N_1 will be in N_2 if and only if C^a and C of Eqs. (2.1) and (2.2) satisfy, for some $\epsilon > 0$,

$$C - 2|C^a| \geq -(4k-\epsilon)/(1+d^2) \tag{3.2}$$

with some choice of origin in Σ. Thus, N_2 includes all points (q_{ab}, P^{cd}) at which the dominant energy condition holds as well as points at which it is "mildly" violated.[14] Finally we introduce an Hermitian inner product on the C_0^∞-spinor fields on Σ associated with any pair (q_{ab}, P^{cd}) in N_2:

$$\langle \alpha, \beta \rangle := \int_\Sigma (D^A{}_B \alpha^B)^+ (D_{AC}\beta^C)\, dV_q\,. \tag{3.3}$$

The linearity in the second member of the inner product as well as the properties $\langle \alpha, \beta \rangle = \overline{\langle \beta, \alpha \rangle}$ and $\langle \alpha, \alpha \rangle \geq 0$ follow from inspection. Next, combining the identity (2.9) with (3.1) and (3.2), one obtains, for C_0^∞-spinor fields α^A,

$$\langle \alpha, \alpha \rangle \geq \frac{\epsilon}{8} \int_\Sigma \frac{\alpha^{A+}\alpha_A}{1+d^2}\, dV_q; \tag{3.4}$$

whence it follows that $\langle \alpha, \alpha \rangle = 0$ if and only if $\alpha = 0$. Thus, (3.3) is indeed an Hermitian inner product on C_0^∞-spinor fields on Σ. The Cauchy completion of this pre-Hilbert space will be denoted by H.

We are now ready to prove the first result:

Lemma 1: Fix a C^∞-pair (q_{ab}, P^{cd}) in the neighborhood N_2 of $\overline{\Gamma}$ and a C^∞, asymptotically constant spinor field $\mathring{\alpha}^A$ on Σ. Then, there exists a unique element β^A of H such that $\alpha^A = \mathring{\alpha}^A + \beta^A$ is a solution to the second-order equation $D_A^{+B} D_{BC}\alpha^C = 0$. Furthermore, α^A is C^∞.

Proof: Set $\rho^A = -D^A{}_C \mathring{\alpha}^C$. ρ^A is square integrable on Σ.[15] Now, $\alpha^A = \mathring{\alpha}^A + \beta^A$ satisfies the second-order equation iff β^A satisfies $D_A^{+B} D_{BC}\beta^C = D^{+}{}_A{}^B \rho_B$. To find a solution in H to this equation, consider the linear functional $f(\lambda)$ defined on C_0^∞-spinor fields λ^A:

$$f(\lambda) = \int_\Sigma (D^A{}_B \lambda^B)^+ \rho_A\, dV_q\,. \tag{3.5}$$

Since ρ is in $L^2(\Sigma, dV_q)$, $f(\lambda)$ is continuous wrt λ^A in the H-topology. Since the space of C_0^∞-spinor fields is dense in H, by the Reisz lemma there exists an element β in H such that $f(\lambda) = \langle \lambda, \beta \rangle$, $\lambda \in C_0^\infty(\Sigma)$. Now, Eq. (3.4) implies that elements of H are distributions which are square integrable on $(\Sigma, (1+d^2)^{-1}\, dV_q)$. Hence,

$$f(\lambda) = \int_\Sigma (D^A{}_B \lambda^B)^+ (D_{AC}\beta^C)\, dV_q \tag{3.6}$$

for some distribution β^C in $L^2(\Sigma, (1+d^2)^{-1} dV_q)$. Subtracting (3.6) from (3.5) and integrating by parts, we obtain

$$\int_\Sigma (D^{+}_{AC} D^A{}_B \lambda^B)^+ (\beta^C + \mathring{\alpha}^C) dV_q = 0 \tag{3.7}$$

for all $\lambda \in C_0^\infty(\Sigma)$, i.e., that $\alpha^A = \beta^A + \mathring{\alpha}^A$ is a distributional (or, weak) solution to the second-order equation. Now, from standard theorems on elliptic equations, it follows that α^A is necessarily C^∞. Next, we establish uniqueness. Let β^A and $\hat{\beta}^A$ be two elements of H such that $\mathring{\alpha}^A + \beta^A$ and $\mathring{\alpha}^A + \hat{\beta}^A$ satisfy the second-order equation. Then, $\mu^A = \beta^A - \hat{\beta}^A$ satisfies $D^{+}{}_A{}^B D_{BC}\mu^C = 0$. Hence, for any C_0^∞-field λ^A, we have

$$\int_\Sigma (D^{+}{}_{AC} D^A{}_B \lambda^B)^+ \mu^C\, dV_q = 0\,,$$

whence $\langle \lambda, \mu \rangle = 0$. Since the space of C_0^∞-fields is dense in H, we have $\mu = 0$.

Remarks: (i) Note that, since α^A is asymptotically constant and since H is obtained by Cauchy completion of $C_0^\infty(\Sigma)$, α^A does not belong to H. This is compatible with the proof of the uniqueness result which also implies that there are no nonzero solutions to the second-order equation in the Hilbert space H.

(ii) Note that the solution α^A does not have to tend to the fixed constant spinor field $\mathring{\alpha}^A$ pointwise. That is, even though β^A is a smooth spinor field in H and even though Eq. (3.1) implies that β^A has to be square integrable [wrt $(1+d^2)^{-1}\, dV_q$], *a priori* there is no guarantee that it will tend to zero pointwise at infinity. α^A tends to $\mathring{\alpha}^A$ only in the sense that $\alpha^A - \mathring{\alpha}^A$ belongs to H [and hence to $L^2(\Sigma, (1+d^2)^{-1} dV_q)$].

Next, we consider the (first-order) Witten equation (2.6). We have:

Lemma 2: The spinor field α^A of Lemma 1 necessarily satisfies the first-order equation $D_{AB}\alpha^B = 0$.

Proof: Set $D_{AB}\alpha^B = \mu_A$. Then, it follows from the proof of Lemma 1 that μ_A is C^∞ and square integrable. Consider a sequence $\mu^A_{(n)}$ of C^∞-spinor fields with compact support on Σ which (together with their first derivative) tend to μ^A pointwise. Applying (2.9) and (3.1) to $\mu^A_{(n)}$ and taking the limit as $n \to \infty$, one obtains $\mu^A = 0$. Hence α^A satisfies the first-order Witten equation.

We now wish to use the behavior of Witten spinors under conformal rescalings. Let $(\hat{q}_{ab}, \hat{P}^{cd}) = (\varphi^4 q_{ab}, P^{cd})$ and $\hat{\lambda}^A = \varphi^{-3}\lambda^A$. Then, $\hat{D}_{AB}\hat{\lambda}^B = \varphi^{-3} D_{AB}\lambda^B$. Hence, it is natural to extend the neighborhood N_2 of $\overline{\Gamma}$ to N_3 as follows: An element (q_{ab}, P^{cd}) of N_1 will be said to belong to N_3 iff there exists a C^∞-conformal factor φ such that $(\varphi^4 q_{ab}, P^{cd})$ belongs to N_2.[16] Clearly, $N_2 \subset N_3$. Finally, since the Witten equation does not refer to the trace-free part of P^{ab}, we can further enlarge N_3 to N as follows: (q_{ab}, P^{cd}) belongs to N iff N_3 admits an element $(q_{ab}, \tilde{P}^{cd})$ with $P^{ab}q_{ab} = \tilde{P}^{ab}q_{ab}$. Thus, in essence, N is the subset of Γ consisting of points (q_{ab}, P^{ab}) for

which $(q_{ab}, P^{cd} q_{cd})$ is conformally related to an element of Γ which either satisfies the dominant energy condition on Σ or violates it only slightly. It is straightforward to construct the Hilbert space H for any (q_{ab}, P^{cd}) in N. In terms of H we can now establish the following result:

Theorem: Fix a C^∞ -pair (q_{ab}, P^{cd}) in the neighborhood N of $\bar{\Gamma}$ and an asymptotically constant C^∞ -spinor field $\mathring{\alpha}^A$ on Σ. Then, there exists a unique element β^A in H such that $\alpha^A = \mathring{\alpha}^A + \beta^A$ is a solution to $D_{AB}\alpha^B = 0$. Furthermore, α^A is C^∞.

Proof: Since the Witten equation does not refer to the trace-free part of P^{ab}, without loss of generality, we can assume that (q_{ab}, P^{cd}) is in N_3, i.e., that there exists a conformal factor φ such that $(\hat{q}_{ab}, \hat{P}^{cd}) = (\varphi^4 q_{ab}, P^{cd})$ is in N_2. Now, from Lemmas 1 and 2, given $\mathring{\alpha}^A$, there exists a unique $\hat{\beta}^A$ in $\hat{H}$ such that $\hat{\alpha}^A = \mathring{\alpha}^A + \hat{\beta}^A$ solves $\hat{D}_{AB}\hat{\alpha}^B = 0$.[17] Set $\alpha^A = \varphi^3 \hat{\alpha}^A$. Then α^A satisfies the Witten equation. Furthermore, since $\beta^A := \alpha^A - \mathring{\alpha}^A = \varphi^3 \hat{\alpha}^A - \mathring{\alpha}^A$ and since φ is a smooth and bounded function[16] on Σ, it follows that β^A is in H. This establishes the existence part of the theorem. The uniqueness of β^A in H follows from that of $\hat{\beta}^A$ in $\hat{H}$.

Remarks: The theorem has been proved by combining in a straightforward way the techniques introduced by Reula, the fact that the Witten equation involves only the trace of P^{ab}, and the fact that the equation is conformally invariant. The resulting subset N of Γ for which the result holds is, however, surprisingly large: Not only does N contain points (q_{ab}, P^{cd}) at which the dominant energy condition can be violated—in fact N_2 contains such points already—but, thanks to the conformal invariance, it also contains some points at which the ADM energy is negative. [Example: consider the point $(f_{ab}, 0)$ in the constraint surface, f_{ab} being a flat metric. Choose a C^∞ -conformal factor φ such that, outside, say, $r = 17$, $\varphi^4 f_{ab}$ is a negative mass spatial Schwarzschild metric. Then, $(\varphi^4 f_{ab}, 0)$ is in N and has negative ADM energy.] Even at these points of N, the preferred lapses and shifts exist and the Hamiltonian generating the Witten-time translation is positive. How big is N? Clearly, it includes all points of the phase space from which solutions to the constraint equations can be constructed using York's[18] procedure. Furthermore, since N includes all pairs (q_{ab}, P^{cd}) which are conformally related to pairs which violate the dominant energy condition mildly, it may well be that N is essentially all of Γ: it may be possible to obtain arbitrarily large violations of the energy condition by conformally rescaling the mild violations.[19] However, as we shall see in subsequent papers, the fact that N constitutes a finite neighborhood of $\bar{\Gamma}$ is already sufficient for the quantization program based on Witten spinors.

ACKNOWLEDGMENTS

It is a pleasure to thank Oscar Reula for discussions pertaining to Sec. 3.

APPENDIX

Although it has never been stated in the literature, one often hears the statement that Witten's proof of positivity of energy should go through as long as the dominant energy condition holds "on the average" [i.e., as long as $\int_\Sigma (C - 2|C^a|)\, dV_q \geq 0$], even though it may not hold pointwise. The reason behind such an expectation is that C and C^a appear only under an integral in the Witten identity [the first term on the right-hand side of Eq. (2.9)] and the ADM energy is positive if that integral is positive. In this Appendix, we shall present a counterexample to this statement. This discussion is not relevant to the main body of the paper; it is included here only as a general observation.

Let Σ be R^3, q_{ab} be flat and $P^{ab} = \frac{1}{3} P^m{}_m\, q^{ab}$, where $-\frac{1}{2} q^{-1/2} P^m{}_m = \pi$ is a smooth function satisfying $\pi = \pi_0$, a constant, for $r < r_1$ and $\pi = 0$ for $r > r_2$ $(r_2 > r_1)$. The pair (q_{ab}, P^{ab}) is obviously asymptotically flat and its ADM energy is obviously zero. Since it does not correspond to flat space, it is clear from the positive energy theorem that it must violate the dominant energy condition. A straightforward substitution in Eqs. (2.1) and (2.2) yields

$$C^a = -\tfrac{2}{3} D^a \pi \quad \text{and} \quad C = \tfrac{2}{3} \pi^2. \tag{A1}$$

Hence the violation can occur only in the transition region $r_1 \leqslant r \leqslant r_2$. Since C is positive in the range $r < r_2$ while C^a is nonzero only in the range $r_1 < r < r_2$, given a π_0, we can always choose r_1, r_2, and $\pi(r)$ in the range $r_1 < r < r_2$ such that the dominant energy condition is satisfied on the average, i.e., such that

$$\int_0^{r_2} \pi^2 r^2\, dr - \int_{r_1}^{r_2} 2 \left| \frac{\partial \pi}{\partial r} \right| r^2\, dr \geqslant 0. \tag{A2}$$

Thus, satisfaction of the energy condition on an average cannot be sufficient for Witten's proof to go through. What goes wrong? One's first reaction may be that the Witten spinors may not exist on the background (q_{ab}, P^{cd}) under consideration. Note, however, that if π_0 is chosen to be sufficiently small, the Witten spinors will exist by Lemmas 1 and 2 of Sec. 3.[20] Hence, the identity (2.9) still holds and implies that any Witten spinor λ^A must satisfy

$$\frac{-1}{4\pi} \int_\Sigma (D^{AB}\lambda^C)^+ (D_{AB}\lambda_C)\, dV_q = \frac{1}{16\pi} \int_\Sigma (TC + 2T^a C_a)\, dV_q, \tag{A3}$$

where (T, T^a) is the lapse-shift pair (null translation) defined by λ^A (with $T > 0$.) Thus, the Witten argument for positivity of energy is inapplicable because, with the rigidly transported (T, T^a) the matter integral has the wrong sign [i.e., $\int_\Sigma (TC + 2T^a C_a)\, dV_q$ is negative] even though the dominant energy condition is satisfied on the average [i.e., $\int_\Sigma (C - 2|C_a|)\, dV_q$ is positive].

Using the fact that the spatial metric of the negative mass Schwarzschild solution is conformally flat, it is straightforward to modify the example to obtain a (q_{ab}, P^{cd}) pair which satisfies the dominant energy condition on the average and yet has negative ADM energy.

[1] K. Kuchař, in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U.P., Oxford, 1981).
[2] See, e.g., M. Ko, M. Ludvigsen, and E. T. Newman, Phys. Rep. **71**, 51–139 (1981).
[3] R. Penrose, Gen. Rel. Grav. **7**, 31 (1976).
[4] E. Witten, Commun. Math. Phys. **80**, 381 (1981).

[5]O. Reula, J. Math. Phys. **23**, 810 (1982).

[6]dV_q is the volume element on Σ defined by the metric q_{ab}.

[7]That is, $T = \tilde{T} + O(1/r)$ and $T^a = \tilde{T}^a + O(1/r)$.

[8]N, $\tilde{N}$, N^a, and $\tilde{N}^a$ may be functions of (q_{ab}, P^{ab}). However, each of the two integrals as well as $g(q,P)$ have to admit bounded derivatives on $\bar{\Gamma}$ in order to get the correct dynamics.

[9]Note, however, that $\mathring{\lambda}^A$ itself is defined using the spin structure of q_{ab}, and not of f_{ab}.

[10]Since the transport equation is elliptic, the solutions cannot vanish on an open region of Σ. It is of interest to know if they can have isolated zeros. In the intuitive discussion, we shall often assume that such zeros do not exist, although our main results are insensitive to this assumption.

[11]For the relation between the SL(2,C) and the SU(2) spinors, see, e.g., A. Sen, J. Math. Phys. **22**, 1781 (1981) and Int. J. Theor. Phys. **21**, 1 (1982). (Note, however, that these references use the signature $+ - - -$.) Associated with the solution λ^A of (2.6)′ is a null vector field $K^a = \lambda^A \lambda^{A'}$ defined on Σ. The lapse and the shift defined by this vector field are, in the SU(2) notation, given by $T = \lambda^{+A} \lambda_A$ and $T^a = -\sqrt{2}\, i\lambda^{+(A} \lambda^{B)}$.

[12]For simplicity, we shall assume that all fields on Σ are C^∞.

[13]O. Reula (private communication). The generalized Sobolev inequality given on p. 811 of Ref. 5 is incorrect. In effect, Lemma 1 rectifies (and extends slightly) the proof of the main theorem in Ref. 1 by using Eq. (3.1) in place of the incorrect inequality.

[14]In the present notation, the dominant energy condition is $C - 2|C^a| \geqslant 0$. C and C^a are related to the more familiar μ and J^a by $C = 16\pi\mu$ and $C^a = 8\pi J^a$, whence the usual form of the energy condition $\mu - |J^a| \geqslant 0$ follows.

[15]This follows from the square integrability of $|\partial_a q_{bc}|$ and the fact that, by definition $\mathring{\alpha}^A$ satisfies $\partial_a \mathring{\alpha}^B = 0$, where ∂_a is the derivative operator compatible with a flat metric f_{ab} to which q_{ab} approaches.

[16]Since q_{ab} and $\hat{q}_{ab}$ approach f_{ab} at infinity, φ tends to 1 at spatial infinity. Smoothness of φ and the fact that the complement of a compact set of Σ is homeomorphic to the complement of a ball of R^3 now imply that φ is bounded on Σ.

[17]Since q_{ab} and $\hat{q}_{ab}$ are conformally related, they define the same spin structure. Hence, α^A is a spinor field wrt q_{ab} as well as $\hat{q}_{ab}$.

[18]See, e.g., J. W. York, in *The Proceedings of the 1982 Les Houche Summer School*, edited by N. Deruelle and S. Piran (North-Holland, Amsterdam, 1983).

[19]Indeed, given any point (q_{ab}, P^{cd}) in Γ, one can define a nonnegative functional

$$S[\lambda] = \int_\Sigma (D^A{}_B \lambda^B)^+ (D_{AC} \lambda^C)\, dV_q$$

on the space of spinor fields λ^A which asymptotically approach a constant $\mathring{\lambda}^A$. The minimum of this functional (under perturbations of compact support) satisfies the second-order equation. Since $S[\lambda]$ is nonnegative, one expects the minimum to exist.

[20]It is apparently unknown if the Witten spinors continue to exist for arbitrarily large values of π_0.

Spherical charged fluid distributions in general relativity

Ramesh Tikekar
Department of Mathematics, Sardar Patel University, Vallabh Vidyanagar, 388 120, India

(Received 12 May 1983; accepted for publication 2 December 1983)

Formal features of Einstein–Maxwell equations for spherically symmetric distributions of a charged perfect fluid in equilibrium are discussed. An exact solution of the system of equations for a specified choice of matter density and fluid pressure, representing a charged perfect gas is presented.

PACS numbers: 04.20.Jb

1. INTRODUCTION

The problem of determination of exact solutions of Einstein–Maxwell equations in general relativity, representing static, spherically symmetric distributions of matter in the form of a charged perfect fluid has attracted wide attention. Such solutions will be convenient to describe the equilibrium states of collapsing spherical distributions of matter whose collapse to a point singularity is countered by the repulsive electrical force due to the presence of charge. The gravitational field in the exterior region of such charged fluid spheres is described by the Reissner–Nordstrom metric.

Spherically symmetric distributions of charged incoherent matter were studied by Bonnor,[1,2] who showed that equilibrium of such distributions could only be maintained when charge density bears a constant ratio with matter density throughout the distribution. De and Raychaudhari[3] have shown that this result is a consequence of Einstein–Maxwell equations for distributions which are singularity-free. Cooperstock and De la Cruz[4] have shown that, for charged perfect fluid spheres with nonzero pressure, in equilibrium $m^2 > q^2$, where m denotes the gravitating mass and q denotes the total charge contained within the sphere.

In this paper some formal features of the coupled Einstein–Maxwell equations of general relativity for static, spherical distributions of matter in the form of a charged perfect fluid are discussed, and a new class of their analytic solutions is presented. The physical plausibility of the solutions is discussed, and it is shown that Pant and Sah's[5] solution is a member of this class.

2. EINSTEIN–MAXWELL EQUATIONS

The coupled Einstein–Maxwell equations in the geometric units with $c = G = 1$ read

$$R_i^k - \tfrac{1}{2} R\delta_i^k = -8\pi T_i^k. \tag{1}$$

For a perfect fluid with charge, the energy–momentum tensor

$$T_i^k = (\rho + p)u_i u^k - p\delta_i^k + (1/4\pi)(-F_{ij}F^{kj} + \tfrac{1}{4}F_{mn}F^{mn}\delta_i^k). \tag{2}$$

Here, ρ, p, and u^i, respectively, denote the matter density, the fluid pressure, and the unit, timelike 4-velocity field of the fluid. The components of the Maxwell stress tensor F_{ij} satisfy

$$F_{ij,k} + F_{jk,i} + F_{ki,j} = 0, \tag{3}$$

$$\frac{\partial}{\partial x^j}(F^{ij}\sqrt{-g}) = 4\pi\sqrt{-g}J^i. \tag{4}$$

For nonconducting fluids, the 4-current density

$$J^i = \sigma u^i, \tag{5}$$

where σ denotes the charge density. We consider the spherically symmetric space-time metric in the Schwarzschild coordinates

$$ds^2 = -e^{\lambda(r)}dr^2 - r^2 d\theta^2 - r^2\sin^2\theta\, d\delta\phi^2 + e^{\nu(r)}dt^2, \tag{6}$$

as representing the geometry of the spherical charged perfect fluid distribution in an equilibrium configuration. The 4-velocity of the fluid will be

$$u^i = (0, 0, 0, e^{-\nu/2}) \tag{7a}$$

and, consequently,

$$J^i = (0, 0, 0, \sigma e^{-\nu/2}). \tag{7b}$$

In view of spherical symmetry, the Maxwell stress tensor has only one nonzero independent component, which is determined by Eq. (4) as

$$F_{41} = -\frac{e^{(\lambda+\nu)/2}}{r^2}\int_0^r 4\pi\sigma r^2 e^{\lambda/2}\,dr. \tag{8a}$$

We introduce the electric field intensity E as

$$E^2 = -F_{41}F^{41}, \tag{8b}$$

so that the total charge contained within the sphere of coordinate radius r is given by

$$q(r) = 4\pi\int_0^r \sigma r^2 e^{\lambda/2}\,dr = r^2 E. \tag{8c}$$

The coupled Einstein–Maxwell equations stated explicitly read

$$8\pi T_1^1 = -8\pi p + E^2 = -e^{-\lambda}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) + \frac{1}{r^2}, \tag{9}$$

$$8\pi T_2^2 = -8\pi p - E^2 = -e^{-\lambda}\left(\frac{\nu''}{2} + \frac{\nu'^2}{4} - \frac{\lambda'\nu'}{4} + \frac{\nu' - \lambda'}{2r}\right), \tag{10}$$

$$8\pi T_3^3 = 8\pi T_2^2,$$

$$8\pi T_4^4 = 8\pi\rho + E^2 = e^{-\lambda}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2}. \tag{11}$$

Here and in what follows a prime indicates a differentiation with respect to r.

The system of equations (9), (10), and (11) can formally be solved to give

$$8\pi\rho = -\frac{e^{-\lambda}}{2}\left(\frac{\nu''}{2} + \frac{\nu'^2}{4} - \frac{\lambda'\nu'}{4} - \frac{\nu' + 5\lambda'}{2r}\right)$$

 0022-2488/84/051481-03$02.50

$$+\frac{1-e^{-\lambda}}{2r^2}, \tag{12a}$$

$$8\pi p=\frac{e^{-\lambda}}{2}\left(\frac{\nu''}{2}+\frac{\nu'^2}{4}-\frac{\lambda'\nu'}{4}+\frac{3\nu'\lambda'}{2r}\right)+\frac{e^{-\lambda}-1}{2r^2}, \tag{12b}$$

$$E^2=\frac{e^{-\lambda}}{2}\left(\frac{\nu''}{2}+\frac{\nu'^2}{4}-\frac{\lambda'\nu'}{4}-\frac{\nu'+\lambda'}{2r}\right)+\frac{1-e^{-\lambda}}{2r^2}. \tag{12c}$$

For charged fluid distributions with specified matter density $\rho(r)$ and known equation of state $p=p(\rho)$, one would like to integrate this system of equations for λ and ν. Equations (12a) and (12b) give

$$8\pi(\rho+p)=(e^{-\lambda}/r)(\lambda'+\nu'). \tag{13}$$

If we use this relation to eliminate ν' and ν'' from (12a), we get

$$e^{-\lambda}\left(\frac{\lambda''}{2}-\frac{\lambda^2}{2}+\frac{2\lambda'}{r}-\frac{1}{r^2}\right)+2\pi(\rho+p)\lambda'-16\pi^2(\rho+p)^2r^2e^{\lambda}-16\pi\rho-4\pi(\rho'+p')r+\frac{1}{r^2}=0, \tag{14}$$

the equation for determining λ. Subsequently, ν is given by a formal integration of (13) as

$$\nu=8\pi\int(\rho+p)e^{\lambda}r\,dr-\lambda+\text{const.} \tag{15}$$

The electric field required to maintain the equilibrium of the configuration is determined by (12c) as

$$E^2=e^{-\lambda}\left(\frac{\lambda'}{r}-\frac{1}{r^2}\right)+\frac{1}{r^2}-8\pi\rho. \tag{16}$$

3. A SOLUTION OF FIELD EQUATIONS

We assume the following expressions for the matter density and the fluid pressure:

$$4\pi\rho=(a/r^2)e^{-\lambda}, \tag{17a}$$

$$4\pi p=(b/r^2)e^{-\lambda}, \tag{17b}$$

where a and b are nonnegative constants. These expressions imply that the matter distribution consists of a perfect gas with equation of state $p\alpha\rho$. Substitution for ρ and p in (14) gives

$$r^2\psi''+2r\psi'\{\tfrac{3}{2}(a+b)+2\}+2\{(a+b+1)^2-4b\}\psi=2, \tag{18}$$

wherein we have put $e^{-\lambda}=\psi$. Equation (18) admits the general solution in the form

$$e^{-\lambda}=\psi=\alpha r^{-p_1}+\beta r^{-p_2}+\{(a+b+1)^2-4b\}^{-1}, \tag{19}$$

where

$$p_1=\tfrac{1}{2}[3(a+b+1)+\{(a+b+1)^2+32b\}^{1/2}], \tag{20a}$$

$$p_2=\tfrac{1}{2}[3(a+b+1)-\{(a+b+1)^2+32b\}^{1/2}]. \tag{20b}$$

Equation (15) then determines

$$e^{\nu}=A^2r^{2(a+b)}e^{-\lambda}, \tag{21}$$

A^2 being the arbitrary constant of integration. The space-time metric of the solution reads

$$ds^2=-\left[\frac{\alpha}{r^{p_1}}+\frac{\beta}{r^{p_2}}+\frac{1}{(a+b+1)^2-4b}\right]^{-1}dr^2-r^2\,d\theta^2-r^2\sin^2\theta\,d\phi^2+A^2r^{2(a+b)}\left[\frac{\alpha}{r^{p_1}}+\frac{\beta}{r^{p_2}}+\frac{1}{(a+b+1)^2-4b}\right]dt^2. \tag{22}$$

The square of the electric field intensity E^2 and the proper charge density $\sigma^*=\sigma e^{\lambda/2}$ follow from (16), (4), (5), and (8b), respectively, as

$$E^2=\frac{1}{r^2}\left[\frac{(a+b)^2-2b}{(a+b+1)^2-4b}+\frac{\alpha(p_1-\overline{2a+1})}{r^{p_1}}+\frac{\beta(p_2-\overline{2a+1})}{r^{p_2}}\right], \tag{23}$$

$$4\pi\sigma^*=\frac{\pm 1}{2r^3E}\left[\frac{2(a+b)^2-4b}{(a+b+1)^2-4b}+\frac{\alpha(2-p_1)(p_1-\overline{2a+1})}{r^{p_1}}+\frac{\beta(2-p_2)(p_2-\overline{2a+1})}{r^{p_2}}\right]. \tag{24}$$

4. PHYSICAL PLAUSIBILITY

It is necessary that expression (19) assigns positive values for $e^{-\lambda}$ throughout the region of validity of the distribution represented by (22). The physical requirements

$$\rho>0,\quad p\geqslant 0,\quad \rho-3p\geqslant 0 \tag{25}$$

imply the following conditions on a and b:

$$a>0,\quad b\geqslant 0,\quad a-3b\geqslant 0. \tag{26}$$

The requirement $e^{-\lambda}>0$ throughout the distribution implies $\alpha>0, \beta>0$. However, if $a>0$ and $\beta>0$, (20a) and (20b) imply that Eq. (23) gives $E^2<0$ in the central region of the distribution. If R denotes the radius of this region obtained by solving $E^2(R)=0$, the space-time metric (22) provides an analytic solution of Einstein–Maxwell equations in the region $r>R$. If we choose $b=0$, the distribution will be in the form of charged incoherent matter, and, if we choose $a=3b$, the distribution will exist in the form of a disordered radiation with charge.

5. A PARTICULAR CASE

The gravitational field in the exterior region of a static charged fluid sphere of radius r_0 is uniquely described by the Reissner–Nordstrom metric

$$ds^2=-\left(1-\frac{2m}{r}+\frac{q^2}{r^2}\right)^{-1}dr^2-r^2\,d\theta^2-r^2\sin^2\theta\,d\phi^2$$

$$+\left(1-\frac{2m}{r}+\frac{q^2}{r^2}\right)dt^2. \tag{27}$$

Accordingly, the metric (22) will describe the field in the interior of a charged fluid sphere of radius r_0, if and only if the boundary conditions

$$e^{\nu(r_0)} = e^{-\lambda(r_0)} = (1 - 2m/r_0 + q^2/r_0^2) \tag{28a}$$

and

$$p(r_0) = 0 \tag{28b}$$

are satisfied across the boundary $r = r_0$.

Equation (17b) implies that it is not possible to find a finite radius $r = r_0$ such that (28b) can be satisfied unless $b = 0$. Accordingly, the metric (22) with $b \neq 0$ represents a distribution of infinite extent and so cannot be matched with the Reissner–Nordstrom metric.

In particular, when $\alpha = \beta = 0$, the metric (22) describes a physically plausible distribution of a charged perfect gas with matter density and the fluid pressure given by

$$4\pi\rho = a/[(a + b + 1)^2 - 4b\,]r^2, \tag{29a}$$

$$4\pi p = b/[(a + b + 1)^2 - 4b\,]r^2, \tag{29b}$$

respectively. The total charge contained within a spherical region of radius r_0 is found from (23) and (8c) as

$$q = \left[\frac{(a+b)^2 - 2b}{(a+b+1)^2 - 4b}\right]^{1/2} r_0. \tag{30}$$

When the constants a and b are identified with

$$a = [1 - c(n-1)^2]/4c, \tag{31a}$$

$$b = [c(n+1)^2 - 1]/4c, \tag{31b}$$

the metric (22) with $\alpha = 0, \beta = 0$ assumes the form of the solution given by Pant and Sah,[5] which reads

$$ds^2 = -(1/c)\,dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\phi^2 + A^2 r^{2n}\,dt^2. \tag{32}$$

The form presented here has the advantage that the physical requirements (25) leading to the choice of a and b as conditioned by (26) are suitably handled.

Pant and Sah have matched the metric (32) with (27) across a finite boundary $r = r_0$ by imposing the conditions (28a) only. Since it is not possible to find $r = r_0$ such that $p(r_0) = 0$, it should be noted that this solution describes an infinite distribution and so cannot be matched with (27).

However, the metric (22) with $\alpha = \beta = 0$, $b = 0$, which represents a distribution of charged incoherent matter can be matched with Reissner–Nordstrom metric (27) across a finite boundary $r = r_0$. The boundary conditions (28a) determine the constants A^2 and m as

$$A^2 = (1/r_0)^{2a}, \tag{33}$$

$$m = [a/(a+1)]r_0. \tag{34}$$

It is observed that the De–Raychaudhari requirement $\sigma = \pm\rho$ is fulfilled in this case, and accordingly, the equilibrium of the static charge sphere can be maintained.

ACKNOWLEDGMENT

The author wishes to express his sincere thanks to the referee for his invaluable suggestions in the revision of the manuscript.

[1] W. B. Bonnor, Z. Phys. 160, **59** (1960).
[2] W. B. Bonnor, Mon. Not. R. Astron. Soc. **129**, 443 (1965).
[3] U. K. De and A. K. Raychaudhari, Proc. R. Soc. (London), Ser. A **303**, 97 (1968).
[4] F. I. Cooperstock and V. De la Cruz, Gen. Rel. Grav. **9**, 835 (1978).
[5] D. N. Pant and A. Sah, J. Math. Phys. **20**, 2537 (1979).

A new class of exact pp-wave solutions in simple supergravity[a)]

Franz Embacher
Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria

(Received 7 June 1983; accepted for publication 5 August 1983)

A class of exact solutions of the O(1) supergravity theory is presented. It is obtained from an ansatz motivated by a previously known exact plane-wave solution together with the additional assumption that the connection forms equal their torsion-free and zero-order Grassmann part. The demonstrated solutions are nontrivial and admit a covariantly constant spinor field.

PACS numbers: 04.20.Jb, 04.30. + x, 04.50. + h

1. INTRODUCTION

There are only few known exact nontrivial solutions of the O(1) supergravity field equations.[1] As starting point, we use one of them given by Aichelburg and Dereli, describing a plane-fronted wave with parallel rays[2] (pp-wave, Ref. 3). The Rarita–Schwinger field $\psi = \psi_\mu\, dx^\mu$ in the Weyl representation of the γ matrices[4] takes the form

$$\psi = F(u)\begin{pmatrix} \alpha\, d\zeta \\ \alpha\, d\zeta \\ \alpha^*\, d\zeta^* \\ -\alpha^*\, d\zeta^* \end{pmatrix}, \tag{1}$$

where $(x^0, x^1, x^2, x^3) = (t,x,y,z)$, $\zeta := y + iz$, $u := t - x$, and α a constant (complex) Grassmann parameter (* denotes complex conjugation). ψ is associated with the orthonormal tetrad $e^a = e^a_\mu dx^\mu$,

$$\begin{aligned} e^0 &= dt - H(u,y,z)du, \\ e^1 &= dx - H(u,y,z)du, \\ e^2 &= dy, \quad e^3 = dz, \end{aligned} \tag{2}$$

where H has to satisfy the two-dimensional Laplace equation

$$\frac{\partial^2 H}{\partial y^2} + \frac{\partial^2 H}{\partial z^2} = 0,$$

and F is an arbitrary real function.

The energy-momentum tensor of ψ vanishes as a consequence of the anticommutativity $\alpha^*\alpha = -\alpha\alpha^*$,

$$*\mathcal{T}_a := \tfrac{1}{2}\, i\bar{\psi} \wedge \gamma_5\gamma_a D \wedge \psi = 0$$

($D \wedge \psi = d\psi + \frac{1}{2}\omega^{ab}\ \sigma_{ab} \wedge \psi$, the exterior covariant derivative of ψ, * = Hodge star operator). The two parts of the Rarita–Schwinger equation are separately satisfied:

$$\gamma \wedge d\psi = \tfrac{1}{2}\gamma \wedge \omega^{ab}\sigma_{ab} \wedge \psi = 0$$

($\gamma := \gamma_a e^a$). Since α is constant, the torsion part of the Einstein tensor vanishes although the torsion itself is nonzero. The Einstein equations just express the fact that the line element $ds^2 = e^a_\mu e^b_\nu \eta_{ab}\, dx^\mu dx^\nu$ is a vacuum metric. The solution (ψ, e^a) cannot be generated from $(\psi = 0, e^a)$ by a finite global (or local infinitesimal) supersymmetry transformation, hence it is nontrival.[2,5]

In this paper, a new class of exact solutions is presented. It does not include the mentioned one. One possible aim of an alternative ansatz to (1), (2) is to obtain a dependence of ψ_μ on the coordinates y and z in order to describe gravitino waves localized in a finite region of space. If we only allow α to depend on u, y, z without changing the vierbein e^a, the field equations[6] demand $d\alpha = 0$.[2] Thus we also have to add a Grassmann-valued part to the tetrad.

2. THE ANSATZ

We define the one-forms (2) to be the zero-order Grassmann part ("body") of the tetrad,

$$\begin{aligned} \tilde{e}^0 &:= dt - H\, du, \\ \tilde{e}^1 &:= dx - H\, du, \\ \tilde{e}^2 &:= dy, \quad \tilde{e}^3 := dz, \end{aligned} \tag{3}$$

where $H = H(u,y,z)$,

$$\delta^{AB}\ \partial_{AB} H = 0; \tag{4}$$

capital indices $A, B, C, \ldots$ run from 2 to 3, $\partial_A = \partial/\partial x^A$. The gravitino field is now taken to be

$$\psi = \begin{pmatrix} \beta\, d\zeta \\ \beta\, d\zeta \\ \beta^*\, d\zeta^* \\ -\beta^*\, d\zeta^* \end{pmatrix}, \tag{5}$$

where $\beta(u,\zeta)$ is an anticommuting (Grassmann-valued) complex function. It shall be analytic in ζ, which can be expressed by the usual Cauchy–Riemann differential equations

$$\frac{\partial \beta}{\partial y} + \frac{i\,\partial \beta}{\partial z} = 0. \tag{6}$$

As a consequence, $d\beta = (\partial\beta/\partial u)\, du + (\partial\beta/\partial\zeta)\, d\zeta$; thus $d\psi \sim du \wedge d\zeta$.

Now we make an ansatz for the orthonormal tetrad,

$$e^a = \tilde{e}^a + \mathscr{E}^a, \tag{7}$$

where $\mathscr{E}^a$ is the Grassmann-valued part ("soul"). In general, it may consist of a second-order and a fourth-order part. To evaluate the connection forms, we define $\tilde{\omega}^a_b$ to be the connection of the "background" $\tilde{e}^a$ only,

$$d\tilde{e}^a = -\tilde{\omega}^a_b \wedge \tilde{e}^b, \quad \tilde{\omega}^{ab} = -\tilde{\omega}^{ba}, \tag{8a}$$

and $\hat{\omega}^a_b$ to be the metric connection of e^a,

$$de^a = -\hat{\omega}^a_b \wedge e^b, \quad \hat{\omega}^{ab} = -\hat{\omega}^{ba}. \tag{8b}$$

The $\tilde{\omega}^a_b$ are well known: $\tilde{\omega}^{01} = \tilde{\omega}^{23} = 0$, $\tilde{\omega}^{0A} = \tilde{\omega}^{1A}$

[a)] Work supported in part by the Einstein Memorial Foundation.

 0022-2488/84/051484-05$02.50

$= (\partial_A H)du$. Now we decompose the $\hat{\omega}^a_b$ into real part $\overset{(\circ)}{\omega}{}^a_b$ and Grassmann-valued part ξ^a_b,

$$\hat{\omega}^a_b = \overset{(\circ)}{\omega}{}^a_b + \xi^a_b, \tag{9}$$

and insert the decompositions (7) and (9) into (8b), obtaining

$$d\tilde{e}^a + d\mathscr{C}^a = -\overset{(\circ)}{\omega}{}^a_b \wedge \tilde{e}^b - \overset{(\circ)}{\omega}{}^a_b \wedge \mathscr{C}^b - \xi^a_b \wedge \tilde{e}^b - \xi^a_b \wedge \mathscr{C}^b.$$

Only the first terms on the left- and right-hand sides are non-Grassmannian, thus we read off, comparing with (8b),

$$\overset{(\circ)}{\omega}{}^a_b = \tilde{\omega}^a_b.$$

The first step to specify our ansatz is that we demand the tetrad to satisfy

$$\xi^a_b \wedge \mathscr{C}^b = 0. \tag{10}$$

$\xi^a_b \wedge \mathscr{C}^b$ are fourth-order Grassmann-valued two-forms. Equation (10) is surely satisfied, if the tetrad (and the vector-spinor field ψ which is to be related to the tetrad by the field equations) involves only *three* real Grassmann parameters. As a consequence of (10) we find

$$d\mathscr{C}^a + \tilde{\omega}^a_b \wedge \mathscr{C}^b = -\xi^a_b \wedge \tilde{e}^b. \tag{11}$$

The sense of condition (10) will be clear after the next step, the evaluation of the torsion.

The torsion of ψ can be calculated directly from (5), giving

$$\begin{aligned} &S^a := \tfrac{1}{4} i\bar{\psi}\gamma^a \wedge \psi, \\ &S^0 = S^1 = -2\beta\, {}^*\beta\, dy \wedge dz = :\kappa(u,y,z) dy \wedge dz, \\ &S^1 = S^3 = 0. \end{aligned} \tag{12}$$

Expanded with respect to the zero-order background forms $\tilde{e}^a$, the torsion takes the form

$$S^a = \lambda^a_b \wedge \tilde{e}^b, \quad \lambda^{ab} = -\lambda^{ba} \tag{13}$$

with

$$\begin{aligned} &\lambda^{01} = 0, \quad \lambda^{12} = \lambda^{02} = \tfrac{1}{2}\kappa\tilde{e}^3, \\ &\lambda^{13} = \lambda^{03} = -\tfrac{1}{2}\kappa\tilde{e}^2, \quad \lambda^{23} = \tfrac{1}{2}\kappa\, du. \end{aligned}$$

In general, the one-forms λ^a_b consist of a second-order and a fourth-order Grassmann part. The complete connection forms ω^a_b are now given by

$$\begin{aligned} &S^a = de^a + \omega^a_b \wedge e^b, \quad \omega^{ab} = -\omega^{ba}, \\ &\omega^a_b = \tilde{\omega}^a_b + \xi^a_b + \lambda^a_b. \end{aligned} \tag{14}$$

They are split up here in the $\tilde{e}$-part, the $\mathscr{C}$-part, and the torsion part.

In order to specify a class of solutions that allows the field equations to be easily integrated, we have to impose an additional restriction. We demand that the torison part of the connection forms compensates their $\mathscr{C}$-induced part,

$$\xi^a_b + \lambda^a_b = 0, \tag{15a}$$

or, in other words, that the total connection (14) consists only of its non-Grassmann ("body") part

$$\omega^a_b = \tilde{\omega}^a_b. \tag{15b}$$

From (11) one sees that this is equivalent to

$$d\mathscr{C}^a + \tilde{\omega}^a_b \wedge \mathscr{C}^b = S^a, \tag{15c}$$

which is a differential equation for the Grassmann-valued one-forms $\mathscr{C}^a$. Our ansatz is now characterized [besides the obvious equations (4) and (6)] by the conditions (10) and (15). Condition (10) is, e.g., easily satisfied by restricting the ansatz to only three anticommuting variables, and (15) leads to a set of differential equations.

3. THE FIELD EQUATIONS

The Rarita–Schwinger equation for the ansatz (ψ, e^a) reads

$$\begin{aligned} 0 &= \gamma \wedge (D \wedge \psi) \\ &= \gamma_a \tilde{e}^a \wedge d\psi + \tfrac{1}{2}\gamma_a \tilde{e}^a \wedge \tilde{\omega}^{bc}\sigma_{bc} \wedge \psi + \gamma_a \mathscr{C}^a \wedge D \wedge \psi. \end{aligned}$$

It is easily checked that the first two terms vanish separately; thus we are left with the equation

$$\gamma_a \mathscr{C}^a \wedge D\psi = 0 \tag{16}$$

(which is trivially solved if one restricts the ansatz to only *two* real Grassmann variables).

Now we turn to the differential equations (15c). Imposing $\mathscr{C}^0 = \mathscr{C}^1 =: \mathscr{F}$ [which is strongly suggested by the component version of (15c)], we see that $d\mathscr{C}^A = 0$. We take $\mathscr{C}^2$ and $\mathscr{C}^3$ to be exact forms:

$$\mathscr{C}^A = d\mathscr{G}^A, \quad \mathscr{G}^A = \mathscr{G}^A(u,y,z). \tag{17}$$

The equation for $\mathscr{F}$ becomes

$$d\mathscr{F} - (\partial_A H)du \wedge d\mathscr{G}^A = \kappa\, dy \wedge dz. \tag{18}$$

Taking the exterior derivative, we find that $\mathscr{F}$ exists locally if and only if (integrability condition)

$$\epsilon^{BC}(\partial_{AB}H).\ \partial_C \mathscr{G}^A = -\frac{\partial \kappa}{\partial u} \tag{19}$$

($\epsilon^{23} = -\epsilon^{32} = 1$). Expanding $\mathscr{F}$, we set

$$\mathscr{F} = \mathscr{F}_u(u,y,z)du + \mathscr{F}_A(u,y,z)dx^A,$$

which gives Eq. (18) the form

$$\partial_A \mathscr{F}_u - \frac{\partial \mathscr{F}_A}{\partial u} = -(\partial_C H)\partial_A \mathscr{G}^C, \tag{20}$$

$$\epsilon^{AB}\partial_A \mathscr{F}_B = \kappa. \tag{21}$$

Clearly, (19) is a consequence of these two equations. If κ splits into $\kappa(u,\zeta) = \kappa_1(u)\kappa_2(\zeta)$, (21) is a consequence of (19) and (20).

Our solution is now determined by (4), (6), (16), (20), and (21) which have to be inserted into the Einstein equations. Since the curvature forms are given by

$$\mathscr{R}^{ab} = d\tilde{\omega}^{ab}$$

(remember $\tilde{\omega}^{ab} \sim du$, thus $\tilde{\omega}^a_b \wedge \tilde{\omega}^b_c = 0$), the Einstein equations reduce to

$$\begin{aligned} \mathscr{R}^{bc} \wedge {}^*e_{abc} &\equiv \epsilon_{abcd}\mathscr{R}^{bc} \wedge \tilde{e}^d + \epsilon_{abcd}\mathscr{R}^{bc} \wedge \mathscr{C}^d \\ &= -2{}^*\mathscr{T}_a. \end{aligned}$$

The first term on the left-hand side vanishes because the $\tilde{e}^a$

form a vacuum field as in ordinary general relativity. The energy-momentum tensor is easily calculated from (5) to be zero,

$$*\mathscr{T}_a := \tfrac{1}{2} i\psi \wedge \gamma_5 \gamma_a D \wedge \psi = 0.$$

Thus we are left with

$$\epsilon_{abcd} \mathscr{R}^{bc} \wedge \mathscr{E}^d = 0,$$

which reduces to

$$\delta^{AB} (\partial_{BC} H) \partial_A \mathscr{G}^C = 0. \tag{22}$$

In the next step we satisfy (10) and (16) by reducing the number of independent Grassmann parameters.

4. SPECIALIZATION TO ONE COMPLEX GRASSMANN PARAMETER

We now impose the condition that the vierbein and the gravitino field depend at most on *one* complex parameter. This specialization leads to a drastic simplification of the field equations without reducing the solution to be trivial. The general ansatz for this restriction is

$$\beta(u,\zeta) = \alpha A(u,\zeta) + \alpha^* B(u,\zeta),$$

where α is a constant Grassmann parameter and A and B are analytic in ζ. For simplicity, we restrict the ansatz to

$$\beta(u,\zeta) = \alpha A(u,\zeta). \tag{23}$$

The $\alpha^* B$-term gives only a trivial generalization. As A is analytic in ζ,

$$\frac{\partial A}{\partial y} + i \frac{\partial A}{\partial z} = 0. \tag{24}$$

As a consequence of (23), the Grassmann-valued functions and forms may be expressed as follows:

$$\mathscr{E}^a = \alpha^* \alpha \hat{e}^a, \quad \mathscr{F} = \alpha^* \alpha f, \quad \mathscr{F}_A = \alpha^* \alpha f_A,$$
$$\mathscr{F}_u = \alpha^* \alpha f_u, \quad \mathscr{G}^A = \alpha^* \alpha g^A,$$
$$\xi^a_{\ b} = \alpha^* \alpha \nu^a_b, \quad \lambda^a_{\ b} = \alpha^* \alpha \mu^a_b,$$
$$\kappa = -2\alpha^* \alpha A^* A = \alpha^* \alpha k.$$

The Grassmann parts split off explicitly and we are left to deal with the ordinary one-forms $\hat{e}^a, f, \nu^a_b, \mu^a_b$ and the ordinary functions f_A, f_u, g^A, and k. Equations (10) and (16) are automatically satisfied, and the remaining equations are

$$\mu^a_b + \nu^a_b = 0 \quad \text{[cf. (15a)]}, \tag{25}$$

$$\epsilon^{BC} (\partial_{AB} H)\, \partial_C g^A = -\frac{\partial k}{\partial u} \quad \text{[cf. (19)]}, \tag{26}$$

$$\partial_A f_u - \frac{\partial f_A}{\partial u} = (\partial_C H) \partial_A g^C \quad \text{[cf. (20)]}, \tag{27}$$

$$\epsilon^{AB} \partial_A f_B = k \quad \text{[cf. (21)]}, \tag{28}$$

$$\delta^{AB} (\partial_{BC} H)\, \partial_A g^C = 0 \quad \text{[cf. (22)]}. \tag{29}$$

Now our relevant equations are (4), (24), (27), (28), and (29). Equation (26) is the integrability condition for the local existence of the one-form f and follows from (27) and (28).

The generalized ansatz $\beta = \alpha A + \alpha^* B$ would give exactly the same equations (26)–(29) with $k = -4(A^* A - B^* B)$ instead of $k = -4A^* A$.

5. SEPARATION OF THE u-DEPENDENCE

If the u dependence is split off via

$$A(u,\zeta) = F(u) b(\zeta),$$
$$H(u,y,z) = F_1(u) K(y,z),$$
$$g^A(u,y,z) = F_2(u) h^A(y,z),$$
$$f_u(u,y,z) = F'(u) F(u) m(y,z),$$
$$f_A(u,y,z) = \tfrac{1}{2} F^2(u) n_A(y,z),$$
$$k(u,y,z) = -2F^2(u)\, b^*(\zeta) b(\zeta),$$

such that

$$F_1(u) F_2(u) = F'(u) F(u),$$

(4), (24), and (26)–(29) may be rewritten, now involving only the two coordinates y and z, as

$$\delta^{AB} \partial_{AB} K \quad (K \text{ harmonic}), \tag{30}$$

$$\frac{\partial b}{\partial y} + i \frac{\partial b}{\partial z} = 0 \quad (b \text{ analytic}), \tag{31}$$

$$\epsilon^{BC} (\partial_{AB} K) \partial_C h^A = 4b^* b, \tag{32}$$

$$\partial_A m - n_A = (\partial_C K) \partial_A h^C, \tag{33}$$

$$\epsilon^{AB} \partial_A n_B = -4b^* b, \tag{34}$$

$$\delta^{AB} (\partial_{BC} K) \partial_A h^C = 0. \tag{35}$$

[The generalized ansatz

$$\beta = \alpha A + \alpha^* B,$$
$$A(u,\zeta) = F(u) b_1(\zeta), \quad B(u,\zeta) = F(u) b_2(\zeta),$$

would give the same equations (32)–(35) with $b^* b$ replaced by $b_1^* b_1 - b_2^* b_2$.]

The integrability condition (32) follows from (33) and (34). Conversely, (34) is a consequence of (32) and (33). Hence the equations determining our solutions are either [(30) to (35) without (32)] or [(30) to (35) without (34)]. If one chooses the second possibility, (33) may be regarded as the definition of n_A. Then $f = f_u\, du + f_A\, dx^A$ may be calculated explicitly,

$$f = \tfrac{1}{2} d[F^2(u)\, m(y,z)] + \tfrac{1}{2} F^2(u) (\partial_C K)(\partial_A h^C) dx^A,$$

and n_A has disappeared. If K is given, two functions h^A satisfying (35) are easily found. The only problem that remains is, that the function $+\tfrac{1}{4} \epsilon^{BC} (\partial_{AB} K) \partial_C h^A$ has to be the squared absolute value of an analytic function b. If b is given, K and h^A have to be found from (32) and (35) which turn out to be the two fundamental equations of our problem.

Simple choices of the u-dependence are $F_1(u) = F(u)$ and $F_1(u) = F'(u)$. In both cases, the g^A are linear in F, f is quadratic. This is a consequence of the nonlinearity of the Einstein equations; the F^2-terms arise from the torsion which is quadratic in ψ.

Equations (30) and (32)–(35) may be expressed in a covariant manner on the two-dimensional flat manifold given by the line element

$$dl^2 = dy^2 + dz^2 = \delta_{AB}\, dx^A\, dx^B.$$

In an arbitrary coordinate system $\bar{x}^A(y,z)$, the line element takes the form $dl^2 = \tau_{AB}\, d\bar{x}^A\, d\bar{x}^B$. On this manifold, the one-form

$$n := n_A dx^A = \bar{n}_A d\bar{x}^A$$

and the vector field

$$h := h^A \partial_A = \bar{h}^A \bar{\partial}_A$$

are covariant objects; K, m, and $b\,{}^*b$ are scalar functions. The above equations (30), (32)–(35) take their covariant form if ∂_A is replaced by the covariant derivative ∇_A with respect to the metric τ_{AB}, δ^{AB} by τ^{AB}, and ϵ^{AB} by $(\det \tau)^{-1/2}\,\epsilon^{AB}$. After the calculation, the components of n and h have to be retransformed into the coordinate system (y,z).

The mathematical structure of Eqs. (32) and (35) is rather simple. Let b and K be given. Raise and lower indices with δ^{AB} (the same thing can be done in an arbitrary coordinate system $\bar{x}^A$). Then (32) and (35) can be brought into the form

$$\partial_2 h_3 + \partial_3 h_2 = -4b\,{}^*b\Delta^{-1}\partial_{22}K = \text{given},$$

where Δ is the determinant of the matrix $(\partial_{AB}K)$, and

$$\partial_2 h_2 - \partial_3 h_3 = 4b\,{}^*b\Delta^{-1}\partial_{23}K = \text{given}.$$

Note that the equations are identically satisfied if $\Delta \equiv 0$ (in general $\Delta \leqslant 0$). Defining the two-dimensional vector $q := (h_2, -h_3)$, we may write, using the two-dimensional differential operators div and rot,

$$\operatorname{div} q = \text{given},$$

$$\operatorname{rot} q = \text{given}.$$

These equations can be integrated (e.g., by the method of Fourier transforms if the functions $b\,{}^*b\Delta^{-1}\partial_{AB}K$ do not behave too wildly). In principle, there is an explicit integral formula for h_A in terms of b and K. In this sense, our solution (with split-off u-dependence) is characterized by (a) an arbitrary analytic function $b(\zeta)$ [or two functions $b_1(\zeta)$, $b_2(\zeta)$ in the generalized ansatz], (b) an arbitrary harmonic function $K(y,z)$, (c) functions $F(u)$, $F_1(u)$, and $F_2(u)$ satisfying $F_1(u)F_2(u) = F'(u)$, $F(u)$, (d) an arbitrary function $m(y,z)$, and (e) an arbitrary constant complex Grassmann parameter α.

Since b is (locally) analytic, it is unbounded (if not a constant) and posesses signularities either in the complex plane or at infinity. The same thing occurs with the harmonic function K (which can be regarded as the real part of an analytic function). If one restricts b to $\lim_{\rho\to\infty}\beta = 0$ and the absence of essential singularities, b has to be a rational function.

The spinor field ψ and the "body" vierbein $\tilde{e}_a$ may be chosen completely independent of each other. This fact is rather strange but corresponds to the remarkable freedom in constructing solutions of the supergravity field equations.

6. AN EXAMPLE

As an example of finding an explicit solution we use polar coordinates $(\bar{x}^2, \bar{x}^3) = (\rho, \phi)$,

$$y = \rho\cos\phi \qquad \tau_{AB} = \begin{pmatrix} 1 & 0 \\ 0 & \rho^2 \end{pmatrix},$$
$$z = \rho\sin\phi$$

and demand that

$\partial\bar{n}_A/\partial\phi = \partial\bar{h}^A/\partial\phi = \partial K/\partial\phi = \partial m/\partial\phi = 0$. The Laplace equation (30) gives essentially

$$K(\rho) = \ln\rho.$$

Equation (35) leads to $\bar{h}^2(\rho) = C = \text{const}$, and Eq. (32) gives $\bar{h}^3(\rho) = 4b\,{}^*b\rho^2$. We define $b(\zeta) := c\zeta^p$ for integer p. Then $b\,{}^*b$ depends only on ρ, and $\bar{h}^3(\rho) = 4c^2\rho^{2p+2}$. Since $\bar{h}^2$ is constant, we get $\bar{n}_A = \bar{\partial}_A m$; thus $n_A = \partial_A m$ and

$$f = \tfrac{1}{2}d\,[F^2(u)\,m(\rho)].$$

Transformation into (y,z)-coordinates gives

$$h = \left(\frac{Cy}{\rho} - 4c^2 z\rho^{2p+2}\right)dy + \left(\frac{Cz}{\rho} + 4c^2 y\rho^{2p+2}\right)dz.$$

Thus the final solution is

$$\begin{aligned} e^0 &= dt - F_1(u)\ln\rho\,du + \tfrac{1}{2}\alpha^*\alpha\,d\,[F^2(u)\,m(\rho)], \\ e^1 &= dx - F_1(u)\ln\rho\,du + \tfrac{1}{2}\alpha^*\alpha\,d\,[F^2(u)\,m(\rho)], \\ e^2 &= dy + \alpha^*\alpha\,d\,[F_2(u)\{Cy/\rho - 4c^2 z\rho^{2p+2}\}], \\ e^3 &= dz + \alpha^*\alpha\,d\,[F_2(u)\{Cz/\rho + 4c^2 y\rho^{2p+2}\}], \end{aligned} \tag{36}$$

$$\psi = \begin{pmatrix} \alpha A\,d\zeta \\ \alpha A\,d\zeta \\ \alpha^* A^*\,d\zeta^* \\ -\alpha^* A^*\,d\zeta^* \end{pmatrix}, \quad A(u,\zeta) = cF(u)\,\zeta^p,$$

with $F_1(u)F_2(u) = F'(u)F(u)$, integer p, m an arbitrary function of $\rho = (y^2+z^2)^{1/2}$, c and C constants. Finally, we calculate the metric $ds^2 = e^a_\mu e^b_\nu \eta_{ab}\,dx^\mu\,dx^\nu$ for $m \equiv 0$, $C = 0$, and $p = -1$:

$$\begin{aligned} ds^2 = {} & dt^2 - dx^2 - 2F_1(u)\ln\rho\,du^2 - d\rho^2 \\ & - \rho^2\,d\phi\,(d\phi - 8c^2\alpha^*\alpha F_2'(u)du). \end{aligned} \tag{37}$$

7. DISCUSSION

Each solution that is subject to the restrictions made in Sec. 2 admits a covariantly constant spinor ϵ, $D\epsilon = 0$,

$$\epsilon = \begin{pmatrix} \alpha \\ \alpha \\ \alpha^* \\ -\alpha^* \end{pmatrix}, \quad d\alpha = 0,\ \alpha \text{ arbitrary},$$

and a corresponding covariantly constant one-form $l = l_a e^a$, namely $l = e^0 - e^1 = du$. l is the gradient of the null coordinate u, thus a null field, $l_a l_b \eta^{ab} = 0$. In this sense, the obtained solutions are pp-waves.[3]

Moreover, the presented solution (5), (7) is nontrivial and different from the previously known one (1), (2) in the following sense. Suppose that the linearized version (infinitesimal β) of (ψ, e^a) is obtained by an infinitesimal local supergauge transformation from $(\psi = 0, \tilde{e}^a)$:

$$\psi = \tilde{D}\epsilon \quad (\tilde{D} = d + \tfrac{1}{2}\tilde{\omega}^{ab}\sigma_{ab} = D),$$

$$e^a = \tilde{e}^a.$$

If $d\epsilon \neq 0$, an immediate consequence of the first equation is that ϵ has the form

$$\epsilon = \begin{pmatrix} \alpha \\ \alpha + c \\ \alpha^* + c^* \\ -\alpha^* \end{pmatrix}, \quad dc = 0$$

(otherwise $\beta\, d\zeta \neq \beta\, d\zeta$). α exists locally, if

$$\frac{\partial \beta}{\partial u} = \frac{ic}{2}\left(\frac{\partial^2 H}{\partial y^2} - i\frac{\partial^2 H}{\partial y\, \partial z}\right). \tag{38}$$

Together with (4), (6), and (19) this leads to $\partial_A H = 0$, $b = \text{const}$. Only these solutions may be generated by a local infinitesimal supersymmetry transformation. If $d\epsilon = 0$, let

$$\epsilon = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_2^* \\ -\alpha_1^* \end{pmatrix}$$

be an arbitrary constant Majorana spinor. It is easily checked that $\psi = \tilde{D}\epsilon$ if and only if

$$\beta(u,\zeta)d\zeta = \tfrac{1}{2}[(\partial H/\partial z) + i(\partial H/\partial y)](\alpha_2 - \alpha_1)du$$

which can only be valid if $\psi = 0$. Thus, the *linearized* solution $(\psi,\tilde{e}^a)$ cannot be generated by a *local infinitesimal* supergauge transformation [except those satisfying (38)]. For the same reasons, ψ can be produced from (1), (2) by an infinitesimal local super-rotation only if $\partial_A b = 0$. But then the linearized version of (ψ,e^a) is identical with that of (1), (2).

Now suppose that a full solution (ψ,e^a) is generated by a finite global supersymmetry transformation from $(\psi = 0,\tilde{e}^a)$ with parameter ϵ $(d\epsilon = 0)$. Ref. 5 tells us

$$\psi = \tilde{\omega}^{bc}\sigma_{bc}\epsilon$$

which implies $\beta\, d\zeta \sim du$, hence $\psi = 0$, i.e., the *full* solution (ψ,e^a) cannot be produced by a *global finite* supergauge transformation. In the same way we check that the solution (1), (2), denoted by $(\psi^{(\circ)},\tilde{e}^a)$, is not related with (ψ,e^a) by a global finite super-rotation if β is of first order in the Grassman algebra. The transformation law for the gravitino fields becomes

$$\psi = \psi^{(\circ)} + \tfrac{1}{2}\tilde{\omega}^{ab}\sigma_{ab}\epsilon + \text{third-order Grassmann terms},$$

giving the same contradiction $(\beta(u,\zeta) - F(u)\alpha)d\zeta \sim du$. Only solutions satisfying $\partial_A \beta = \partial_A H = 0$ remain. Except for these, $(\psi^{(\circ)},\tilde{e}^a)$ and (ψ,e^a) are never related by a finite global supergauge transformation.

The solution (1), (2) is not included in the class presented here, because condition (15) is not valid. Furthermore, if $\partial_A \beta \neq 0$, (ψ,e^a) cannot be obtained by a pure tetrad rotation applied to $\psi^{(\circ)}$. Suppose $e^a = \Lambda^a_{\ b}(x^\mu)\tilde{e}^b$, $\Lambda^a_{\ b}$ a Grassmann-valued Lorentz matrix (i.e., $\Lambda^a_{\ b}\Lambda^c_{\ d}\eta_{ac} = \eta_{bd}$) for each space-time point. The spinor components $\psi^{(\circ)} = \psi^{(\circ)}_\mu\, dx^\mu$ transform into $\hat{\psi} = \hat{\psi}_\mu dx^\mu$ according to

$$\begin{aligned} &\hat{\psi}_\mu = S\psi^{(\circ)}_\mu, \\ &S^{-1}\gamma^a S = \Lambda^a_{\ b}\gamma^b, \quad \det S = 1. \end{aligned} \tag{39}$$

If $\partial_A \beta \neq 0$, β has a zero either in the complex plane or at infinity. Hence our solution (5) satisfies

$$\lim_{\substack{\zeta \to \zeta_\circ \\ u = \text{const}}} \psi_\mu = 0,$$

where $\zeta_\circ$ denotes either a complex number or a certain manner for ζ to go to infinity. If $\hat{\psi}$ is such a solution, we note that $\lim_{\zeta\to\zeta_\circ} \psi^{(\circ)}_\mu \neq 0$, thus [from (39)] $\lim_{\zeta\to\zeta_\circ} S = 0$ which is a contradiction to $\det S = 1$. $\psi^{(\circ)}$ and ψ are not related by a tetrad rotation. Moreover, a pure coordinate transformation $x^\mu \to x^\mu$ cannot produce Grassmann-valued terms in the metric coefficients $g_{\mu\nu}$.

As a result we conclude that most of our new solutions are nontrivial and truly different from (1), (2).

Finally, we discuss the plane-wave solution found by Urrutia (Ref. 1). His solution is the general one obtained by the ansatz $\psi \sim du$, $e^a = \tilde{e}^a$. Obviously, the torsion is zero $(du \wedge du = 0)$. This solution cannot be related to the one presented here by a global finite supergauge transformation if β is of first order in the Grassmann algebra. If it is denoted by $\psi^{(1)}$, the relation

$$\psi = \psi^{(1)} + \tfrac{1}{2}\tilde{\omega}^{ab}\sigma_{ab}\epsilon + \text{third-order Grassmann terms}$$

reads

$$\beta d\zeta \sim (\text{first order})du + \text{third-order terms}.$$

Since β is taken to be of first order, we conclude $\beta = 0$. In contrast, an infinitesimal relationship may be established. If we apply a local infinitesimal supergauge transformation with parameter

$$\epsilon = \begin{pmatrix} \alpha(u,\zeta) \\ \alpha(u,\zeta) + \tau(u) \\ \alpha^*(u,\zeta) + \tau^*(u) \\ -\alpha^*(u,\zeta) \end{pmatrix}$$

to produce $\Phi := \psi + D\epsilon$, we find

$$\psi = \partial_A \epsilon\, dx^A \quad \text{is of type (5)},$$

$$\Phi = (\partial\epsilon/\partial u)du + \tfrac{1}{2}\tilde{\omega}^{ab}\sigma_{ab}\epsilon \sim du.$$

The linearized version of one of Urrutia's solutions is Φ. The infinitesimal solutions of our class differ from those of Urrutia only by infinitesimal local super-rotations. The finite solutions are truly different from each other.

We are left with a new class of solutions satisfying the O(1) supergravity field equations showing the following properties.

(i) The connection forms are real valued although metric and vierbein are Grassmann valued and the torsion is nonzero.

(ii) The gravitino-energy-momentum tensor vanishes.

(iii) There exists a covariantly constant spinor (and a corresponding covariantly constant null vector field, the propagation vector of the pp-wave).

ACKNOWLEDGMENT

I wish to thank Professor P. C. Aichelburg for ideas and patience concerning my work.

[1] L. F. Urrutia, Phys. Lett. B **102**, 393 (1981); T. Dereli and R. W. Tucker, Phys. Lett. B **97**, 396 (1980); see also R. J. Finkelstein and J. Kim, J. Math. Phys. **22**, 2228 (1981), and Ref. 2.

[2] P. C. Aichelburg and T. Dereli, Phys. Rev. D **18**, 1754 (1978).

[3] D. Kramer, H. Stephani, E. Herlt, and M. MacCallum, *Exact Solutions of Einstein's Field Equations* (Cambridge University, Cambridge, 1980), p. 233.

[4] P. C. Aichelburg and H. K. Urbantke, Gen. Relativ. Gravit. **13**, 817 (1981).

[5] N. S. Baaklini, S. Ferrara, and P. van Nieuwenhuizen, Lett. Nuovo Cimento **20**, 113 (1977).

[6] P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).

Symmetric space property and an inverse scattering formulation of the SAS Einstein–Maxwell field equations

Ahmet Eriş,
Department of Physics, Middle East Technical University, Ankara, Turkey

Metin Gürses
Department of Applied Mathematics, Marmara Scientific and Industrial Research Institute, Gebze-Kocaeli, Turkey

Atalay Karasu
Department of Physics, Middle East Technical University, Ankara, Turkey

(Received 7 April 1983; accepted for publication 29 July 1983)

We formulate stationary axially symmetric (SAS) Einstein–Maxwell fields in the framework of harmonic mappings of Riemannian manifolds and show that the configuration space of the fields is a symmetric space. This result enables us to embed the configuration space into an eight-dimensional flat manifold and formulate SAS Einstein–Maxwell fields as a σ-model. We then give, in a coordinate free way, a Belinskii–Zakharov type of an inverse scattering transform technique for the field equations supplemented by a reduction scheme similar to that of Zakharov–Mikhailov and Mikhailov–Yarimchuk.

PACS numbers: 04.20.Jb, 11.10.Qr, 11.10.Lm, 11.15.Kc

1. INTRODUCTION

Completely integrable systems and, in connection with these, Bäcklund transformations have attracted much attention in recent years. As a result we now have a better understanding of the nature of certain nonlinear partial differential equations of mathematical physics, and this enables us to devise methods for systematic generation of exact solutions at least in two dimensions. One of these methods is the inverse scattering transform technique of Belinskii and Zakharov.[1,2] It consists of (1) representation of the nonlinear system in the form of compatibility conditions of a more general overdetermined system of linear matrix equations depending on a complex spectral parameter; (2) explicit integration of a Bäcklund transformation for these equations, thus generating new solutions from the known ones.

The Belinskii–Zakharov integration technique was first applied, by the authors, to the Einstein vacuum field equations where the space-time admits two commuting Killing vectors. These authors obtained all multisoliton solutions of Einstein's equations for stationary axially symmetric (SAS) vacuum and colliding plane gravitational wave space-times. The method was later extended and applied by Aleksejev[3] to SAS Einstein–Maxwell equations. In both formulations the parametrization of the problem was such that the relevant linear eigenvalue equation contained the space-time metric functions directly.

In this work we present, in a coordinate-free way, a different formulation of the inverse scattering transform technique of Belinskii and Zakharov for the integration of SAS Einstein–Maxwell field equations written in terms of complex Ernst[4] potentials. In Sec. 2, we first show that the use of Ernst potentials enables one to formulate SAS Einstein–Maxwell field equations as equations determining harmonic mappings[5,6] from a base manifold, which is a three-dimensional flat space, to a four-dimensional Riemannian manifold called the configuration space of the fields. It turns out that the configuration space is a Riemannian symmetric space with the isometry group SU(2,1), in agreement with previous results[7–9] that Ernst equations are invariant under the action of this group. This property enables us to write the kinematical content of the theory in terms of the Maurer–Cartan equations for SU(2,1) while the dynamical content (i.e., the field equations) appear in the form of conservation of Noether currents, both expressed in terms of the set of eight Killing vectors that the configuration space admits. Using 3×3 matrix representation of the generators of SU(2,1) we define a Lie algebra valued flat connection 1-form which can be integrated to give the 3×3 Hermitian matrix, found recently by Gürses and Xanthopoulos,[10] characterizing SAS Einstein–Maxwell fields as a σ-model. The symmetric space property of the configuration space is reflected by the fact that this matrix leaves invariant the metric of the three-dimensional complex vector space on which the group SU(2,1) acts. As a final remark of Sec. 2, we note that the configuration space can be embedded into an eight-dimensional flat space, generalizing the result of Matzner and Misner[11] for SAS Einstein vacuum to the electrovacuum case.

In Sec. 3, we present a Belinskii–Zakharov type of inverse scattering formulation for the integration of SAS Einstein–Maxwell field equations. Using the flat connection 1-form of Sec. 2 and its Hodge dual with respect to a two-dimensional Euclidean space E^2 and introducing a complex spectral parameter, we construct a new connection 1-form defined on $E^2\times C$ whose curvature vanishes modulo the field equations. The associated linear eigenvalue equation follows immediately while gauge transformations of the connection are nothing but Bäcklund transformation for the field equations. Using the technique invented by Belinskii and Zakharov, a particular form of Bäcklund transformation can be integrated explicitly for the 3×3 matrix, characterizing the solution in terms of a known solution. The procedure, however, does not guarantee the symmetric space

 0022-2488/84/051489-07$02.50

property of the 3×3 matrix, which is crucial for the parametrization of SAS Einstein–Maxwell fields. To solve this problem, one has to modify the integration scheme by imposing additional conditions an the Bäcklund transformation. This, the so-called reduction problem, was solved by Zakharov and Mikhailov[12] and Mikhailov and Yarimchuk,[13] and is the subject of Sec. 4. We also include an appendix for the details of Sec. 2.

2. THE CONFIGURATION SPACE FOR SAS EINSTEIN–MAXWELL FIELDS

The equations governing SAS Einstein–Maxwell fields in terms of complex Ernst potentials $\epsilon(\rho,z)$ and $\Phi(\rho,z)$ are

$$(\epsilon+\bar{\epsilon}+2\Phi\bar{\Phi})\nabla^2\epsilon=2(\nabla\epsilon+2\bar{\Phi}\nabla\Phi)\cdot\nabla\epsilon,$$
$$(\epsilon+\bar{\epsilon}+2\Phi\bar{\Phi})\nabla^2\Phi=2(\nabla\epsilon+2\bar{\Phi}\nabla\Phi)\cdot\nabla\Phi. \tag{2.1}$$

Here ∇ and ∇^2 are, respectively, the flat space gradient and Laplace operators in cylindrical coordinates $\{\rho,z,\phi\}$. Equations in (2.1) can also be regarded as equations determining harmonic mappings $f: M\to M'$, where M and M' are two Riemannian manifolds with metrics

$$M: ds^2=d\rho^2+dz^2+\rho^2\,d\phi^2, \tag{2.2}$$
$$M': ds'^2=F^{-2}|d\epsilon+2\bar{\Phi}\,d\Phi|^2-4F^{-1}\,d\Phi\,d\bar{\Phi}, \tag{2.3}$$

where $F=\frac{1}{2}(\epsilon+\bar{\epsilon})+\Phi\bar{\Phi}$. To see this, it is enough to consider the set of basis 1-forms ω^A $(A,B,\ldots=1,2,3,4)$ of the cotangent space $T^*(M')$ of M', which are orthonormal with respect to the metric g' of M'. These satisfy the integrability conditions $d\omega^A+\omega^A{}_B\wedge\omega^B=0$, which enables one to determine the connection 1-forms $\omega_{AB}=-\omega_{BA}$. The basis 1-forms ω^A along with the connection 1-forms $\omega^A{}_B$, when pulled back to M using the map f, $\sigma^A=f^*\circ\omega^A$ and $\Omega^A{}_B=f^*\circ\omega^A{}_B$, satisfy

$$d\sigma^A+\Omega^A{}_B\wedge\sigma^B=0, \tag{2.4}$$

displaying the Riemannian structure of the induced vector bundle $f^{-1}T(M')\to M$ transported from that of the tangent bundle $T(M')\to M'$. Introducing now the Hodge dual operation (*), which is determined by the Riemannian structure of M, the field equations (2.1) can be written as

$$d*\sigma^A+\Omega^A{}_B\wedge*\sigma^B=0, \tag{2.5}$$

while the variational principle for the problem is $\delta I=0$ with

$$I=\frac{1}{2}\int_M\sigma_A\wedge*\sigma^A. \tag{2.6}$$

The Riemannian manifold M' with the metric given in Eq. (2.3) is called the configuration space for SAS Einstein–Maxwell fields and is a Riemannian symmetric space. This property implies that there exists eight Killing vectors generating the isometry group SU(2,1) of M' and that the line element (2.3) can be written as (see the Appendix for details)

$$ds'^2=\tfrac{1}{2}\eta_{\mu\nu}\tau^\mu\otimes\tau^\nu, \tag{2.7}$$

where τ^μ $(\mu,\nu,\cdots=1,2,\ldots,8)$ are the corresponding Killing 1-forms satisfying the Maurer–Cartan equations[14]

$$d\tau^\mu+\tfrac{1}{2}C^\mu{}_{\alpha\beta}\tau^\alpha\wedge\tau^\beta=0, \tag{2.8}$$

with $C^\mu{}_{\alpha\beta}$ being the structure constants of SU(2,1) and

$$\eta_{\mu\nu}=\tfrac{1}{6}C^\alpha{}_{\mu\beta}C^\beta{}_{\nu\alpha} \tag{2.9}$$

being the constant group metric. (See the Appendix for details.) The invariance of the action integral (2.6) under SU(2,1) implies the existence of eight Noether currents $f^*\circ\tau^\mu$, which are conserved, i.e.,

$$d*(f^*\circ\tau^\mu)=0. \tag{2.10}$$

As it is for the case of a symmetric space, these conservation laws are in one-to-one correspondence with the field equations (2.5) or equivalently (2.1).

Using the 3×3 matrix representation of the generators X_μ of SU(2,1) we can now define Lie algebra valued connection 1-form

$$W=X_\mu\tau^\mu, \tag{2.11}$$

which, because of the Maurer–Cartan equations (2.8), satisfies

$$dW+W\wedge W=0. \tag{2.12}$$

This implies that the curvature of W vanishes identically; hence

$$W=-(dP)P^{-1}, \tag{2.13}$$

where P is a 3×3 Hermitian matrix with unit determinant. Using Eqs. (2.11) and (2.13) the matrix P can be determined, up to a constant gauge transformation, as

$$P=F^{-1}\begin{pmatrix}1 & \sqrt{2}\Phi & i(\bar{\epsilon}-\epsilon+2\Phi\bar{\Phi})/2\\ \sqrt{2}\bar{\Phi} & -(\epsilon+\bar{\epsilon}-2\Phi\bar{\Phi})/2 & -i\sqrt{2}\epsilon\bar{\Phi}\\ i(\bar{\epsilon}-\epsilon-2\Phi\bar{\Phi})/2 & i\sqrt{2}\bar{\epsilon}\Phi & \epsilon\bar{\epsilon}\end{pmatrix}, \tag{2.14}$$

where

$$F=\tfrac{1}{2}(\epsilon+\bar{\epsilon}+2\Phi\bar{\Phi}),$$

in agreement with the results of Gürses and Xanthopoulos.[10] Because of Eqs. (2.10) and (2.13) the field equations (2.1) can be written in terms of the matrix P as

$$d[(*dP)P^{-1}]=0. \tag{2.15}$$

This form for the SAS Einstein–Maxwell field equations is particularly suitable for the application of inverse scattering transform techniques for generation of soliton solutions.

The matrix P given in (2.14) leaves the metric of the complex vector space on which the group SU(2,1) acts invariant, i.e.,

$$P\gamma P=\gamma, \tag{2.16}$$

where

$$\gamma=\begin{pmatrix}0 & 0 & i\\ 0 & -1 & 0\\ -i & 0 & 0\end{pmatrix},\quad \gamma^2=I. \tag{2.17}$$

By defining $Z=\gamma P$, the configuration space can be considered as the four-dimensional hypersurface $Z^2=I$, embedded in an eight-dimensional flat space with the metric

$$ds'^2 = -\tfrac{1}{2}\,\mathrm{Tr}(dZ \otimes dZ), \tag{2.18}$$

where

$$Z = i\begin{pmatrix} Z_8 - \frac{1}{3}iZ_2 & \sqrt{2}(Z_7 + iZ_6) & Z_3 \\ \sqrt{2}(-Z_4 + iZ_5) & i + \frac{2}{3}iZ_2 & \sqrt{2}(Z_6 + iZ_7) \\ -Z_1 & \sqrt{2}(-Z_5 + iZ_4) & -Z_8 - \frac{1}{3}iZ_2 \end{pmatrix}. \tag{2.19}$$

The embedding equations are

$$\begin{aligned} &Z_1 = F^{-1}, \quad Z_2 = -3\Phi\bar{\Phi}/F, \\ &Z_3 = \epsilon\bar{\epsilon}/F, \quad Z_4 = \tfrac{1}{2}i(\Phi - \bar{\Phi})/F, \\ &Z_5 = \tfrac{1}{2}(\Phi + \bar{\Phi})/F, \quad Z_6 = \tfrac{1}{2}(\bar{\epsilon}\Phi + \epsilon\bar{\Phi})/F, \\ &Z_7 = \tfrac{1}{2}i(\bar{\epsilon}\Phi - \epsilon\bar{\Phi})/F, \quad Z_8 = i(\bar{\epsilon} - \epsilon)/2F, \end{aligned} \tag{2.20}$$

and the line element (2.18) becomes

$$\begin{aligned} -ds'^2 = {} & (dZ^1 \otimes dZ^3 + \tfrac{1}{3}\,dZ^2 \otimes dZ^2 + 4\,dZ^4 \otimes dZ^7 \\ & + 4\,dZ^5 \otimes dZ^6 - dZ^8 \otimes dZ^8), \end{aligned} \tag{2.21}$$

which has zero signature. This generalizes the results of Matzner and Misner[11] for the SAS Einstein vacuum to the SAS Einstein–Maxwell case.

3. THE INVERSE SCATTERING TRANSFORM TECHNIQUE

Even though the base manifold M, whose metric is given by Eq. (2.2), is a three-dimensional flat space, the fact that the map f is independent of the azimuthal coordinate ϕ makes M effectively two-dimensional. For a coordinate-free formulation of the inverse scattering transform technique, we shall from now on consider the two-dimensional Euclidean space E^2 as the base manifold M. Therefore, in what follows the Hodge dual operation (*) should be understood as the one defined with respect to the Riemannian structure on E^2. The field equation given by (2.15) will then read

$$d\,[\Delta\,(*dP)P^{-1}] = 0, \tag{3.1}$$

where Δ is a scalar function on E^2, satisfying

$$d(\Delta^{-1} * d\Delta) = 0. \tag{3.2}$$

The explicit functional form of Δ depends on the particular choice of local coordinates by

$$\Delta = |\det g|^{1/2}, \tag{3.3}$$

where g is the metric of the three-dimensional base manifold M of the previous section.

The flat connection 1-form W defined in Eqs. (2.13) and (2.14) is not suitable for the application of an inverse scattering transform technique to generate soliton solutions of the field equations. This is because its curvature vanishes identically without any references to the field equations. What is needed is a connection whose curvature vanishes on the solution submanifold (i.e., modulo the field equations) and, furthermore, contains a complex parameter in such a way that the connection defined in Eq. (2.13) is obtained in the limit that the value of the parameter goes to zero. For this purpose we introduce a generalized exterior derivative operator D, satisfying $D^2 = 0$, by

$$D \equiv d - \left(\frac{\partial\theta}{\partial\lambda}\right)^{-1} d\theta\,\frac{\partial}{\partial\lambda}, \tag{3.4}$$

where λ is a complex parameter independent of the coordinate on E^2 and $O(\lambda,E^2)$ is any scalar function with the property

$$\lim_{\lambda\to 0} D = d. \tag{3.5}$$

Next we consider a linear eigenvalue problem for a 3×3 matrix $\Psi(\lambda,E^2)$ written as

$$D\Psi = -\Omega\Psi, \tag{3.6}$$

where

$$\Omega = aW + b\Delta * W, \tag{3.7}$$

and $a(\lambda,E^2)$ and $b(\lambda,E^2)$ are complex functions defined on $E^2\times C$ satisfying

$$\lim_{\lambda\to 0} a(\lambda,E^2) = 1, \quad \lim_{\lambda\to 0} b(\lambda,E^2) = 0. \tag{3.8}$$

Integrability of Eq. (3.6) requires

$$D\Omega + \Omega\wedge\Omega = 0 \tag{3.9}$$

on the solution submanifold which restricts further the functions a and b to satisfy

$$a^2 + \Delta^2 b^2 - a = 0 \tag{3.10}$$

and

$$Da = \Delta * Db. \tag{3.11}$$

With this choice we now have a connection 1-form which is integrable because of the field equations (3.1) and which satisfies

$$\lim_{\lambda\to 0} \Omega = W. \tag{3.12}$$

Using Eqs. (3.6) and (3.12), we see that the matrix P can be identified as

$$P(E^2) = \lim_{\lambda\to 0} \Psi(\lambda,E^2). \tag{3.13}$$

Except for the condition given by Eq. (3.5) the function $\theta(\lambda,E^2)$ is arbitrary, with different choices leading to different functions a and b and hence to a different linear eigenvalue problem. But, as will be seen later, its choice is crucial for the explicit integration of a Bäcklund transformation using a solution of the "matrix Riemann problem." In theory, given a proper function $\theta(\lambda,E^2)$ Eqs. (3.10) and (3.11) can be solved for a and b satisfying the conditions given in Eq. (3.8). These functions together with the definition of the connection 1-form Ω, given in Eq. (3.7), determine the linear eigenvalue Eq. (3.6) completely.

We are now at a position in which we can apply the inverse scattering transform technique of Belinskii and Zakharov to the problem under consideration. For this purpose we assume the knowledge of a particular solution P_0 of (3.1) in terms of which we construct the corresponding connection 1-forms $W_0 = -(dP_0)P_0^{-1}$, $\Omega_0 = aW_0 + b\Delta * W_0$, and hence the function $\Psi_0(\lambda,E^2)$, by solving Eq. (3.6). A transformation of the form

$$\Psi = \chi \Psi_0 \tag{3.14}$$

defines a new matrix Ψ leading to a new solution P [using Eq. (3.13)] of (3.1), where, by Eq. (3.6), the matrix χ is restricted to solutions of

$$D\chi = \chi \Omega_0 - \Omega \chi. \tag{3.15}$$

Even though the above procedure guarantees that P is a solution to Eq. (3.1), it does not yet provide solutions to SAS Einstein–Maxwell equations. To represent SAS Einstein–Maxwell fields, the new matrix P must also satisfy the symmetric space property (2.16) and must be Hermitian. These additional conditions, which can be demonstrated easily to be consistent with the field equation (3.1), put additional restrictions on the matrices Ψ and χ. This, the so-called reduction problem, will be the subject of the next section.

4. THE REDUCTION AND INTEGRATION

In order for a solution P of (3.1) to represent SAS Einstein–Maxwell fields, it must be consistent with the parametrization given by Eq. (2.14). This means that the matrix P should be Hermitian ($P = P^\dagger$) and, furthermore, must satisfy the symmetric space property

$$(\gamma P)^2 = I, \tag{4.1}$$

where γ is given in Eq. (2.17). It can easily be shown that these conditions restrict the connection 1-form W to satisfy

$$WP - PW^\dagger = 0, \quad \gamma W + W^\dagger \gamma = 0, \tag{4.2}$$

which in turn imply that we should have

$$\Omega(\lambda)P - P\Omega^\dagger(\bar{\lambda}) = 0, \quad \gamma\Omega(\lambda) + \Omega^\dagger(\bar{\lambda})\gamma = 0. \tag{4.3}$$

With these in mind, if we now reconsider the linear eigenvalue equation (3.6) and Eq. (3.15) for χ, we find that

$$P\gamma\Psi(\tau) = \Psi(\lambda)J, \quad \Psi^\dagger(\bar{\lambda})\gamma = \gamma\Psi^{-1}(\lambda), \tag{4.4}$$

and that

$$\chi^{-1}(\lambda) = \gamma\chi^\dagger(\bar{\lambda})\gamma, \quad P = \chi(\bar{\lambda})P_0\chi^\dagger(\tau), \tag{4.5}$$

where J is a 3×3 matrix satisfying $DJ = 0$, $J^2 = I$, and τ: $\lambda \to \tau(\lambda, E^2)$ is a fractional linear transformation on the complex λ plane with $\tau^2 = I$ leaving the function $\theta(\lambda, E^2)$ invariant. The functions $a(\lambda, E^2)$ and $b(\lambda, E^2)$ transform under τ as

$$\begin{aligned} a(\tau, E^2) &= 1 - a(\lambda, E^2), \\ b(\tau, E^2) &= -b(\lambda, E^2), \end{aligned} \tag{4.6}$$

leaving conditions (3.10) and (3.11) invariant.

The remaining problem is explicit construction of the matrix χ, satisfying the requirements in (4.5) with a given P_0. This can be carried out as done by Belinskii and Zakharov,[1,2] Zakharov and Mikhailov,[12] Mikhailov and Yarimchuk,[13] and Eriş and Gürses[15] once the local coordinates on E^2 and the set of functions $\{\theta, a, b\}$ are fixed. For the integration of SAS Einstein–Maxwell equations a convenient choice is the one given by Belinskii and Zakharov, namely,

$$\begin{aligned} &\Delta = \rho, \quad \theta(\lambda, \rho, z) = \rho^2/2\lambda - \lambda/2 - z, \\ &a(\lambda, \rho) = \rho^2/(\lambda^2 + \rho^2), \quad b(\lambda, \rho) = \lambda/(\lambda^2 + \rho^2), \\ &\tau = -\rho^2/\lambda. \end{aligned} \tag{4.7}$$

Whether this set is unique or not, or whether a different choice satisfying all the requirements mentioned above leads to different solutions to the field equations is still under investigation.

For the N-soliton configuration it is assumed that the matrix $\chi(\lambda, \rho, z)$ is of the form

$$\chi = \mathrm{I} + \sum_{i=1}^{2N} \frac{R_i}{\lambda - \mu_i}, \tag{4.8}$$

where the scalar functions μ_i are the roots of the equation

$$\theta(\mu_i, \rho, z) = -w_i \quad (\text{consts}) \tag{4.9}$$

and the matrices R_i are independent of the complex spectral parameter λ. The above form for χ together with Eq. (4.5) implies that

$$P = \chi(0)P_0. \tag{4.10}$$

For reasons that will become clear later we shall choose the $2N$ poles of χ to be related pairwise as

$$\mu_{N+k} = \tau(\mu_k), \quad k = 1,2,...,N. \tag{4.11}$$

The unknown matrices R_i will be determined using Eqs. (3.15) and (4.5). Since R_i are independent of the spectral parameter λ, it suffices to consider these equations at the poles $\lambda = \mu_k$ and look at the residues at these points. Considering the relation $\chi\chi^{-1} = I$ at the poles $\lambda = \mu_k$, we get

$$R_k\chi^{-1}(\mu_k) = 0, \quad k = 1,2,...,2N, \tag{4.12}$$

displaying the fact that the matrices R_k are degenerate. Equation (3.15) evaluated at $\lambda = \mu_k$ gives

$$(DR_k - R_k\Omega_0)\chi^{-1}|_{\lambda=\mu_k} = 0, \quad k = 1,2,...,2N, \tag{4.13}$$

which, using the fact that the matrix Ψ_0^{-1} satisfies

$$D\Psi_0^{-1} - \Psi_0^{-1}\Omega_0 = 0, \tag{4.14}$$

shows that we have

$$R_k = M_k\Psi_0^{-1}(\mu_k), \quad k = 1,2,...,2N, \tag{4.15}$$

where, for the moment, the matrices M_k appear to be arbitrary except that they have to be degenerate because of (4.12). To determine M_k, we consider the two reduction conditions given by Eq. (4.5). The first condition, when evaluated at $\lambda = \mu_k$, requires that χ^{-1} should have poles at $\lambda = \bar{\mu}_k$ and that

$$R_k + \sum_{i=1}^{2N} \frac{R_k\gamma R_i^\dagger\gamma}{\mu_k - \bar{\mu}_i} = 0, \quad k = 1,2,...,2N. \tag{4.16}$$

On the other hand, the second condition of Eq. (4.5) requires

$$R_kP_0 + \sum_{i=1}^{2N} \frac{R_kP_0R_i^\dagger}{\tau(\mu_k) - \bar{\mu}_i} = 0, \quad k = 1,2,...,2N. \tag{4.17}$$

Since the 3×3 matrices M_k are degenerate, writing

$$M_k = m_kn_k^\dagger + p_kq_k^\dagger, \quad k = 1,2,...,2N, \tag{4.18}$$

Eqs. (4.16) and (4.17) reduce to two systems of linear algebraic equations which must consistently be solved for the column matrices m_k, n_k, p_k, and q_k. The consistency of the two systems, apart from the particular ordering of the poles μ_i as given in Eq. (4.11), requires that we should have

$$J^\dagger n_k = n_{N+k}, \quad J^\dagger q_k = q_{N+k}, \quad k = 1,2,...,N, \tag{4.19}$$

which can equivalently be written as

$$P_0\gamma\psi_0(\bar{\mu}_k)\gamma n_k = \psi_0(\bar{\mu}_{k+N})\gamma n_{k+N},$$
$$P_0\gamma\psi_0(\bar{\mu}_k)\gamma q_k = \psi_0(\bar{\mu}_{k+N})\gamma q_{k+N}. \tag{4.20}$$

With these conditions, Eqs. (4.16) and (4.17) become identical. Hence, using any one of these equations, we obtain the following set of linear algebraic equation:

$$\sum_{i=1}^{2N}\left(\frac{1}{\mu_i-\bar{\mu}_k}\right)[(n_i^\dagger S_{ik}n_k)m_i + (q_i^\dagger S_{ik}n_k)P_i] = \gamma(\psi_0^\dagger(\mu_k))^{-1}n_k,$$
$$\sum_{i=1}^{2N}\frac{1}{\mu_i-\bar{\mu}_k}[(n_i^\dagger S_{ik}q_k)m_i + (q_i^\dagger S_{ik}q_k)P_i] = \gamma(\psi_0^\dagger(\mu_k))^{-1}q_k, \tag{4.21}$$

where

$$S_{ik} = \gamma\psi_0^\dagger(\bar{\mu}_i)\gamma\psi_0(\bar{\mu}_k)\gamma. \tag{4.22}$$

By use of these equations we determine the column vectors m_k and p_k in terms of the column vectors n_k and q_k.

So far the vectors n_k and q_k which have been choosen as constant vectors by Belinskii and Zakharov as a particular solution of Eq. (4.13) appeared to be arbitrary. We shall show that their choice is indeed the unique solution. Using Eqs. (4.15), (4.18), and (4.14) in Eq. (4.13), we obtain

$$(Dn_k^\dagger)\psi_0^{-1}(\mu_k)\chi^{-1}|_{\lambda=\mu_k} = 0. \tag{4.23}$$

Knowing the fact

$$n_k^\dagger\psi_0^{-1}(\mu_k)\chi^{-1}|_{\lambda=\mu_k} = 0, \tag{4.24}$$

then (4.23) can be written as

$$Dn_k^\dagger = H_k n_k^\dagger, \quad \text{at } \lambda=\mu_k, \tag{4.25}$$

where H_k's are 1-forms (not matrix-valued). Applying D operator ($\lambda=\mu_k$) to (4.25), we get

$$DH_k = 0, \quad \lambda=\mu_k. \tag{4.26}$$

Since $D\wedge D=0$, then $H_k = D\gamma_k|_{\lambda=\mu_k}$, where γ_k [$=\gamma_k(E^2)$] are arbitrary functions. Hence $n_k^\dagger$ in Eq. (4.25) can be solved exactly as

$$n_k^\dagger = e^{-\gamma_k}n_{0k}^\dagger, \tag{4.27}$$

where n_{0k}'s are constant vectors. On the other hand, it can be shown from the linear set (4.21) that the vectors m_k's are also scaled by the factors e^{γ_k}. Hence the degenerate matrix M_k given in (4.18) and the new solution P does not contain these functions. The same results are also valid for the vectors q_k and p_k. Therefore we do not lose any generality by taking n_k and q_k as constant vectors.

5. CONCLUDING REMARKS

The symmetric space property and hence the σ-model formulation of the SAS Einstein–Maxwell problem may lead to some other new results. One of them has recently been given by Mazur.[16] He, independently, using such a formulation, proved the uniqueness of the Kerr–Newman black-hole solution. Einstein–Maxwell field equations for space-times admitting only one Killing vector formally look like the field equations given in Eq. (2.1). The crucial difference is that the differential operators are functions of the metric variables for the former case. Nevertheless, generalizing the differential operator D in (3.4), there is a hope of linearizing the E–M field equations for space-times admitting a non-null Killing vector. Work along this direction is in progress.

Although in this work we concentrate on soliton solutions for SAS Einstein–Maxwell fields, the procedure can be applied with minor modifications to every field theory which can be formulated in the framework of harmonic mappings of Riemannian manifolds provided that the configuration space is a Riemannian symmetric space and the base manifold is effectively two-dimensional. Recently,[10] it was shown that there is a close relationship between static axially symmetric self-dual SU(3) Yang–Mills and SAS Einstein–Maxwell fields. Thus, the solution generation technique presented here may also be used to obtain the monopole or the instanton solutions for SU(3). Furthermore, since the dimension of the matrices may be arbitrary, static axially symmetric self-dual SU(N) Yang–Mills fields can be treated as well.

ACKNOWLEDGMENTS

Two of the authors (AE and AK) would like to thank the Turkish Scientific and Technical Research Council (TBTAK) for financial support.

APPENDIX

Choosing the complex tetrad 1-forms as

$$\omega^1 = F^{-1}(d\epsilon + 2\bar{\Phi}\,d\Phi), \quad \omega^2 = F^{-1}(d\bar{\epsilon} + 2\Phi\,d\bar{\Phi}),$$
$$\omega^3 = 2F^{-1/2}\,d\Phi, \quad \omega^4 = 2F^{-1/2}\,d\bar{\Phi}. \tag{A1}$$

The line element (2.3) for the configuration space can be written as

$$d's^2 = \omega^1\otimes\omega^2 - \omega^3\otimes\omega^4. \tag{A2}$$

The connection 1-forms are

$$\omega^1{}_1 = -\omega^2{}_2 = -\tfrac{1}{2}\omega^1 + \tfrac{1}{2}\omega^2,$$
$$\omega^1{}_3 = \omega^4{}_2 = -\tfrac{1}{2}\omega^4,$$
$$\omega^2{}_4 = \omega^3{}_1 = -\tfrac{1}{2}\omega^3,$$
$$\omega^3{}_3 = \omega^4{}_4 = -\tfrac{1}{4}\omega^1 + \tfrac{1}{4}\omega^2. \tag{A3}$$

Using these, the tetrad components of the curvature tensor are found as

$$-R'^1{}_{112} = R'^2{}_{212} = R'^3{}_{334} = -R'^4{}_{434} = \tfrac{1}{2},$$
$$R'^1{}_{134} = R'^1{}_{314} = -R'^2{}_{234} = R'^2{}_{423} = R'^3{}_{123} = -R'^3{}_{312} = R'^4{}_{214} = R'^4{}_{412} = \tfrac{1}{4}, \tag{A4}$$

while the nonzero components of the Ricci tensor are

$$R'_{12} = -\tfrac{3}{4}, \quad R'_{34} = \tfrac{3}{4}. \tag{A5}$$

Hence, $R' = -6$ is the curvature scalar. Hence we have

$$R'_{AB} = -\tfrac{3}{2}g'_{AB} \quad (A,B,\ldots = 1,2,3,4), \tag{A6}$$

where

$$g'_{AB} = \tfrac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}; \tag{A7}$$

therefore (M',g'_{AB}), is an Einstein space. In addition to this property we have

$$R'_{GBCD}R'^{G}{}_{AEF} + R'_{AGCD}R'^{G}{}_{BEF} + R'_{ABGD}R'^{G}{}_{CEF} + R'_{ABCG}R'^{G}{}_{DEF} = 0, \tag{A8}$$

meaning that the Riemann curvature tensor of M' is covariantly constant. This in turn means that M' is a Riemannian symmetric space. For this case the two properties together mean that M' is a harmonic space in the sense of Walker.[17] Thus M' has eight Killing vectors, $\tau_\mu = h_\mu{}^A\,\partial/\partial y^A$ ($\mu,\nu,\ldots, = 1,2,\ldots,8$). With $\{y^A\}$ denoting the set $\{\epsilon,\bar\epsilon,\Phi,\bar\Phi\}$ these are

$$\begin{aligned}
\tau_1 &= i\frac{\partial}{\partial\epsilon} - i\frac{\partial}{\partial\bar\epsilon}, \quad \tau_2 = i\Phi\frac{\partial}{\partial\Phi} - i\bar\Phi\frac{\partial}{\partial\bar\Phi},\\
\tau_3 &= -i\epsilon^2\frac{\partial}{\partial\epsilon} + i\bar\epsilon^2\frac{\partial}{\partial\bar\epsilon} - i\Phi\epsilon\frac{\partial}{\partial\Phi} + i\bar\Phi\bar\epsilon\frac{\partial}{\partial\bar\Phi},\\
\tau_4 &= -2\Phi\frac{\partial}{\partial\epsilon} - 2\bar\Phi\frac{\partial}{\partial\bar\epsilon} + \frac{\partial}{\partial\Phi} + \frac{\partial}{\partial\bar\Phi},\\
\tau_5 &= 2i\Phi\frac{\partial}{\partial\epsilon} - 2i\bar\Phi\frac{\partial}{\partial\bar\epsilon} + i\frac{\partial}{\partial\Phi} - i\frac{\partial}{\partial\bar\Phi},\\
\tau_6 &= -2i\Phi\epsilon\frac{\partial}{\partial\epsilon} + 2i\bar\Phi\bar\epsilon\frac{\partial}{\partial\bar\epsilon} + i(\epsilon - 2\Phi^2)\frac{\partial}{\partial\Phi} + i(2\bar\Phi^2 - \bar\epsilon)\frac{\partial}{\partial\bar\Phi},\\
\tau_7 &= 2\Phi\epsilon\frac{\partial}{\partial\epsilon} + 2\bar\Phi\bar\epsilon\frac{\partial}{\partial\bar\epsilon} + (\epsilon + 2\Phi^2)\frac{\partial}{\partial\Phi} + (\bar\epsilon + 2\bar\Phi^2)\frac{\partial}{\partial\bar\Phi},\\
\tau_8 &= 2\epsilon\frac{\partial}{\partial\epsilon} + 2\bar\epsilon\frac{\partial}{\partial\bar\epsilon} + \Phi\frac{\partial}{\partial\Phi} + \bar\Phi\frac{\partial}{\partial\bar\Phi}.
\end{aligned} \tag{A9}$$

These Killing vectors satisfy the SU(2,1) algebra

$$[\tau_\mu,\tau_\nu] = C^\alpha{}_{\mu\nu}\tau_\alpha, \tag{A10}$$

where

$$\begin{aligned}
&-C^4{}_{16} = C^4{}_{25} = C^4{}_{48} = C^5{}_{17} = -C^5{}_{24} = C^5{}_{58} = -C^6{}_{27},\\
&C^6{}_{34} = -C^6{}_{68} = C^7{}_{26} = -C^7{}_{35} = -C^7{}_{78} = C^8{}_{13} = 1,\\
&C^1{}_{18} = -C^3{}_{38} = C^8{}_{47} = C^8{}_{56} = 2,\\
&C^1{}_{45} = -C^3{}_{67} = 4, \quad -C^2{}_{46} = C^2{}_{57} = 6.
\end{aligned} \tag{A11}$$

The corresponding Killing 1-forms $\tau^\mu = h^\mu{}_A\,dy^A$ are given as

$$\begin{aligned}
\tau^1 &= -iF^{-2}\{\tfrac{1}{2}\bar\epsilon^2 + \Phi\bar\Phi\bar\epsilon\}\,d\epsilon + iF^{-2}\{\tfrac{1}{2}\epsilon^2 + \Phi\bar\Phi\epsilon\}\,d\bar\epsilon + iF^{-2}\bar\Phi\epsilon\bar\epsilon\,d\Phi - iF^{-2}\Phi\epsilon\bar\epsilon\,d\bar\Phi,\\
\tau^2 &= \tfrac{3}{2}iF^{-2}\Phi\bar\Phi\,d\epsilon - \tfrac{3}{2}iF^{-2}\Phi\bar\Phi\,d\bar\epsilon - \tfrac{3}{2}iF^{-2}\bar\Phi(\epsilon+\bar\epsilon)\,d\Phi + \tfrac{3}{2}iF^{-2}\Phi(\epsilon+\bar\epsilon)\,d\bar\Phi,\\
\tau^3 &= \tfrac{1}{2}iF^{-2}\,d\epsilon - \tfrac{1}{2}iF^{-2}\,d\bar\epsilon + iF^{-2}\bar\Phi\,d\Phi - iF^{-2}\Phi\,d\bar\Phi,\\
\tau^4 &= -\tfrac{1}{4}F^{-2}\{\bar\epsilon(\Phi+\bar\Phi) + 2\Phi\bar\Phi^2\}\,d\epsilon - \tfrac{1}{4}F^{-2}\{\epsilon(\Phi+\bar\Phi) + 2\Phi^2\bar\Phi\}\,d\bar\epsilon,\\
&\quad + \tfrac{1}{4}F^{-2}\{\bar\epsilon^2 + \epsilon\bar\epsilon + 2\epsilon\bar\Phi^2\}\,d\Phi + \tfrac{1}{4}F^{-2}\{\epsilon^2 + \epsilon\bar\epsilon + 2\bar\epsilon\Phi^2\}\,d\bar\Phi,\\
\tau^5 &= -\tfrac{1}{4}iF^{-2}\{\bar\epsilon(\bar\Phi-\Phi) + 2\Phi\bar\Phi^2\}\,d\epsilon + \tfrac{1}{4}iF^{-2}\{\epsilon(\Phi-\bar\Phi) + 2\Phi^2\bar\Phi\}\,d\bar\epsilon,\\
&\quad - \tfrac{1}{4}iF^{-2}\{\bar\epsilon^2 + \epsilon\bar\epsilon - 2\epsilon\bar\Phi^2\}\,d\Phi + \tfrac{1}{4}iF^{-2}\{\epsilon^2 + \epsilon\bar\epsilon - 2\bar\epsilon\Phi^2\}\,d\bar\Phi,\\
\tau^6 &= \tfrac{1}{4}iF^{-2}(\Phi+\bar\Phi)\,d\epsilon - \tfrac{1}{4}iF^{-2}(\Phi+\bar\Phi)d\bar\epsilon + \tfrac{1}{2}iF^{-2}\{\bar\Phi^2 - \tfrac{1}{2}(\epsilon+\bar\epsilon)\}\,d\Phi - \tfrac{1}{2}iF^{-2}\{\Phi^2 - \tfrac{1}{2}(\epsilon+\bar\epsilon)\}\,d\bar\Phi,\\
\tau^7 &= -\tfrac{1}{4}F^{-2}(\Phi-\bar\Phi)\,d\epsilon + \tfrac{1}{4}F^{-2}(\Phi-\bar\Phi)\,d\bar\epsilon + \tfrac{1}{2}F^{-2}\{\bar\Phi^2 + \tfrac{1}{2}(\epsilon+\bar\epsilon)\}\,d\Phi + \tfrac{1}{2}F^{-2}\{\Phi^2 + \tfrac{1}{2}(\epsilon+\bar\epsilon)\}\,d\bar\Phi,\\
\tau^8 &= \tfrac{1}{2}F^{-2}\{\bar\epsilon + \Phi\bar\Phi\}\,d\epsilon + \tfrac{1}{2}F^{-2}\{\epsilon + \Phi\bar\Phi\}\,d\bar\epsilon + \tfrac{1}{2}F^{-2}\bar\Phi\{\bar\epsilon - \epsilon\}\,d\Phi + \tfrac{1}{2}F^{-2}\Phi\{\epsilon - \bar\epsilon\}\,d\bar\Phi.
\end{aligned} \tag{A12}$$

They satisfy the Maurer–Cartan equations

$$d\tau^\mu + \tfrac{1}{2}C^\mu{}_{\alpha\beta}\tau^\alpha\wedge\tau^\beta = 0. \tag{A13}$$

Using these, the line element of M' can be written as

$$\begin{aligned}
d's^2 &= \tfrac{1}{2}\eta_{\mu\nu}\tau^\mu\otimes\tau^\nu,\\
d's^2 &= -\tau^1\otimes\tau^3 - \tfrac{1}{2}\tau^2\otimes\tau^2 - 4\tau^4\otimes\tau^7\\
&\quad - 4\tau^5\otimes\tau^6 + \tau^8\otimes\tau^8.
\end{aligned} \tag{A14}$$

We now define a Lie algebra valued connection 1-form W as

$$W = X_\mu\tau^\mu,$$

where X_μ is the 3×3 matrix representation of the generators of the group SU(2,1) satisfying

$$[X_\mu,X_\nu] = C^\alpha{}_{\mu\nu}X_\alpha.$$

Since the group metric $\eta_{\mu\nu}$ can also be written as

$\eta_{\mu\nu} = \frac{1}{2}\mathrm{tr}(X_\mu X_\nu + X_\nu X_\mu)$. The line element (A2), or, equivalently, (A14), can also be written as

$$ds'^2 = \tfrac{1}{2}\,\mathrm{tr}(W \otimes W), \tag{A15}$$

where, written out explicitly in terms of the Killing 1-forms,

$$W = \begin{pmatrix} -\frac{1}{3}i\tau^2 + \tau^8 & \sqrt{2}(i\tau^6 + \tau^7) & \tau^3 \\ \sqrt{2}(-\tau^4 + i\tau^5) & \frac{2}{3}i\tau^2 & 2(\tau^6 + i\tau^7) \\ -\tau^1 & \sqrt{2}(i\tau^4 - \tau^5) & -\frac{1}{3}i\tau^2 - \tau^8 \end{pmatrix},$$

because of the Maurer–Cartan equations (A13). The curvature 2-form of this connection vanishes identically, i.e., $dW + W \wedge W = 0$.

[1]V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 935 (1978).
[2]V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **50**, 1 (1979).
[3]G. A. Aleksejev, Abstracts **GR9**, 1 (1980).
[4]F. J. Ernst, Phys. Rev. **168**, 1776 (1968).
[5]J. Eells and J. H. Sampson, Am. J. Math. **86**, 109 (1964).
[6]Y. Nutku, Ann. Inst. H. Poincaré A **21**, 175 (1974).
[7]C. W. Misner, Phys. Rev. D **18**, 4510 (1978); G. Neugebauer and D. Kramer, Ann. Phys. **24**, 62 (1969); D. Kramer *et al.*, *Exact Solutions of Einstein Field Equations*, Cambridge Monographs on Mathematical Physics **6**, edited by E. Schmutzer (Cambridge U. P., Cambridge, 1980).
[8]A. Eriş, M. Gürses, and A. Karasu, "The Symmetric Space Property and the Embedding Problem For the Stationary Axi-Symmetric Einstein–Maxwell Fields," in *Proceedings of the XI International Colloquium on Group Theoretical Methods in Physics, Istanbul, Turkey, 1982*, edited by M. Serdaroğlu and E. İnönü (Springer-Verlag, Berlin, 1982).
[9]W. Kinnersley, J. Math. Phys. **14**, 651 (1973); **18**, 1529 (1977).
[10]M. Gürses and B. C. Xanthopoulos, Phys. Rev. D **26**, 1912 (1982).
[11]R. Matzner and C. W. Misner, Phys. Rev. **154**, 1229 (1967).
[12]V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[13]A. V. Mikhailov and A. I. Yarimchuk, "Cylindrically Symmetric Solutions of the Non-Linear Chiral Model (σ-model)" CERN Preprint TH-3150, 1981.
[14]K. Yano, *Curvature and Betti Numbers*, Annals of Mathematical Studies No. 32 (Princeton U. P., Princeton, NJ, 1953).
[15]A. Eriş and M. Gürses, "An Inverse Scattering Transform Technique For Stationary Axi-Symmetric Einstein–Maxwell Fields," in *Proceedings of the XI International Colloquium on Group Theoretical Methods in Physics, Istanbul, Turkey, 1982*, edited by M. Serdaroğlu and E. İnönü (Springer-Verlag, Berlin, 1982).
[16]Pawel O. Mazur, J. Phys. A: Math. Gen. **15**, 3173 (1982).
[17]A. G. Walker, J. London Math. Soc. **21**, 47 (1946).

Generalized vector products, duality, and octonionic identities in $D = 8$ geometry

Resit Dündarer, Feza Gürsey, and Chia-Hsiung Tze[a)]
Gibbs Laboratory and Physics Department, Yale University, New Haven, Connecticut 06511

(Received 27 September 1983; accepted for publication 18 November 1983)

In an explicit, unified, and covariant formulation, we study and generalize the exceptional vector products in R^8 of Zvengrowski, Gray, and Kleinfeld. We derive the associated general quadratic, cubic, and quartic G_2 invariant algebraic identities and uncover an octonionic counterpart to the $d = 4$ quaternionic duality. When restricted to seven dimensions the latter is an algebraic statement of absolute parallelism on S^7. We further link up with the Ogievetski–Tzeitlin vector product and obtain explicit tensor forms of the SO(7) and G_2 structure constants. SO(8) transformations related by Triality are characterized by means of invariant tensors. Possible physical applications are discussed.

PACS numbers: 04.50. + h, 11.30.Pb, 02.40.Ky

I. INTRODUCTION

In the past few years spurred by the experimental successes of QCD on the one hand and of the minimal grand unified theory, the standard electroweak model, on the other, there have been renewed and vigorous searches[1] for a geometric, truly unified, theory of all basic interactions. The favored approach has been a Kaluza–Klein type reduction of $d = 11$, $N = 1$ Supergravity.[1,2] While a realistic theory is still out of reach,[3] these attempts have revealed novel, exotic structures. They free particle theorists from the confines of four-dimensional space-time. Indeed, the peculiarities[4] of higher dimensional geometries (specifically, octonionic geometries) may suggest some unusual physical mechanisms and generate new insights into the possible geometries of dynamics of some generalized space-times. From the viewpoint of mathematical physics, the above endeavor has made most conspicuous the possibly central role played by octonions[1,5] and their attending exceptional groups and geometries.[6]

The search for physical relevance of nonassociativity is not new.[7,8] Yet the current emergence of octonions in Kaluza–Klein supergravity is singularly remarkable as the octonionic structure is rooted in the very fabrics of the geometry of extended space-time and its dynamics. A telling illustration of this phenomenon is the geometrical Higgs mechanism due to Englert[9] and others.[10] There spontaneous compactification of $d = 11$ supergravity into a Riemannian product of an anti-DeSitter space-time and S^7, an internal seven-sphere of the imaginary unit octonions, is achieved via the rank 4 Kalb–Ramond gauge field. The latter's potential fields are identified as the left (or the right) fully antisymmetric parallelizable torsions on S^7. The existence of these torsions was attributed by Cartan and Schouten[7] to the nonassociativity of the octonions. In a recent note,[11] we have made this connection[12] to octonions most explicit by giving a purely algebraic form of Englert's solutions. Namely at the poles of S^7, the Cartan–Schouten[13] torsions are simply given by the Cayley structure constants. The rank 4 gauge field strength, which is notably dual to the torsion, is given by the associator of the imaginary octonionic units. Thus the seven-dimensional duality and the resulting geometric Higgs mechanism[1,10] are manifestly linked to the lack of associativity and the alternativity of the octonions.[14]

Moreover from Refs. 9 and 11, we realized that the discovered duality[9] between the potential and the field strength of the rank 4 tensor abelian gauge field must be the residue of a more perfect duality in eight dimensions. We recall that for a $d = 4$ Riemannian manifold M, it is well known[15] that the nonsimple nature of their holonomy group, the norm group of the quaternions, $O(4) = Sp(1) \times Sp(1) = S^3 \times S^3$, giving any $d = 4$ space a quaternionic structure, is the *raison d'être* of the duality structure of Euclidean gravity and gauge fields over M. The existence of (anti-) self-duality is connected to that on R^3, the tangent space to S^3, of a vector cross product, the Grassman–Gibbs product, alias Hamilton's quaternion multiplication. The situation is more subtle in R^8 (Ref. 16) as $O(8) = S^7 \times S^7 \times G_2$ which suggests a more subtle and exotic form of duality mediated in part by the exceptional group G_2. In trying to uncover this exceptional duality we realized, as E. Calabi[17] did long ago, that the full algebraic and geometric implications of the Cayley–Graves octonions is most appropriately seen in terms of invariants of the Cayley space, the R^7-space of the pure imaginary octonions, and of the eight-space with its unique Triality Principle. In other words the invariant tensor identities and the duality we seek should be described by generalized vector cross products,[8] the $d = 7,8$ space analogs of Grassmann–Gibbs vector product in R^3,[16] the space of the imaginary quaternions.

In mathematics, vector cross products on manifolds have been an active research topic.[18] The motivations are at least threefold. Firstly, vector products generalize the fundamental notion of almost complex structure. Secondly, the existence of a vector product on a manifold induces unusual, almost complex, structure on its submanifolds.[17] Thirdly, they provide an approach to the exploration of that *terra incognita* which are the Riemannian spaces with exceptional holonomy groups G_2 and Spin (7).[19,20] In our view the above reasons go hand in hand with the current focus by physicists (and mathematicians alike) on the holomorphic interpretation[15] of instantons and on the various exotic geometries[1,4]

[a)] Alfred P. Sloan Foundation Fellow.

 0022-2488/84/051496-11$02.50

begotten and suggested by Kaluza–Klein compactification of $d = 11$, $N = 1$ Supergravity. This perspective certainly provides a strong inducement to seek the implications of vector cross products on exceptional manifolds. Our work is a step in this direction.

Our paper is organized as follows. Along with introducing our notation, Sec. II lists the properties of octonions relevant to our work. The links between octonions and the groups SO(8), SO(7), G_2, and SU(3) are merely stated with some explicit new connections to be subsequently worked out. We then illustrate the workings of the principle of Triality in R^8 by giving a new derivation of the three Moufang identities. Sections III, IV, and V present our main results. Making primary use of a fundamental G_2 invariant self-dual numerical tensor, we give a compact, unified, covariant, and explicit formulation of various allowed r-fold vector cross products in R^8. Central to our formulation is a novel octonionic duality structure paralleling the self- (anti-self-) duality in four dimensions. This connection between vector cross product, i.e., generalized almost complex structure[8] and duality parallels that between complex structure and four-dimensional duality.[5] Specializing to the seven-dimensional submanifold this duality induces the dual algebraic structure on S^7 observed in Ref. 9 and accounted for explicitly in Ref. 11 in terms of octonions. After defining various vector cross products and their generalizations, we compute the most general corresponding quadratic, cubic, then quartic G_2 invariant tensorial identities in R^8. A linkup is achieved with the cross products of Zvengrowski,[21] Gray,[18] and Kleinfeld.[22] In particular the latter product is given an eight-dimensional covariant extension. We then make contact with and explicate the Ogievetski–Tzeitlin vector product[23] and as a by-product new explicit formulas for the G_2 and SO(7) transformations and structure constants are obtained. Section V applies our formalism to SO(8) transformations related to Triality; they are characterized by invariant tensors. Section IV concludes our study.

II. OCTONIONS

To define our notations we list the features of the octonion algebra as far as they are of relevance to our work. We refer for proofs to Freudenthal *et al.*[6,14] We recall that a division algebra is a linear algebra A, $x, y \in A$ with a multiplicative norm $N \in R$, the real number field, such that

$$N(x,y) = N(x)N(y), \quad N > 0,$$
$$N(x) = 0 \rightarrow x = 0. \tag{2.1}$$

According to the celebrated Hurwitz's[14] theorem, there are only four such division algebras, the real, the complex, and quaternion numbers fields, R, C, H, and Ω, the ring of the Cayley–Graves numbers or octonions. The importance of the first three number fields and function theories over them particularly in supersymmetric theories of matter coupled gravity is quite clear.[24] The relevance of octonions to supergravity is under intensive study.[11,12] The division ring Ω is said[6,25] to be a R-module with an additive canonical basis $\{e_a\} = \{e_0, e_\mu\}$, $\mu = 1,2,...,7$. We will henceforth let Greek indices run from 1 to 7 while Latin ones run from 0 to 7. The ring structure is given by

$$e_0 \text{ the unit element in } \Omega, \tag{2.2}$$

$$e_0^2 = e_0, \quad e_\alpha e_0 = e_0 e_\alpha = e_\alpha, \tag{2.3}$$

$$[e_\alpha, e_\beta] = 2\psi_{\alpha\beta\gamma} e_\gamma, \tag{2.4}$$

$$\{e_\alpha, e_\beta\} = -2\delta_{\alpha\beta}, \quad \alpha, \beta, \gamma = 1,2,...,7. \tag{2.5}$$

The Cayley structure constants $\psi_{\alpha\beta\gamma}$ in the multiplication law $e_\alpha e_\beta = -\delta_{\alpha\beta} + \psi_{\alpha\beta\gamma} e_\gamma$, i.e., Eqs. (2.4) are antisymmetric in (α, β, γ) and nonzero and equal to unity for the seven combinations (or cycles)

(123), (246), (435), (367), (651), (572), (714).

The $\psi_{\alpha\beta\gamma}$ are read off by cyclic permutation from the mnemonic triangle (123) on a circle with the seven points [8,26] labeled in the order (1243657). In fact there are 480 ways of exhibiting the octonionic multiplication table!

An octonion may be written in the standard form $x = x^a e_a$, $x \in \Omega$. The involution is the conjugation $x \rightarrow \bar{x}$, the conjugate $\bar{x} = x^a \bar{e}_a$ with $\bar{e}_a = (e_0, -\bar{e})$. We define, respectively, the scalar and vector part of x as $\mathrm{Sc}(x) = \frac{1}{2}(x + \bar{x}) = x_0$ and $\mathrm{Vec}(x) = \frac{1}{2}(x - \bar{x})$. The inner product of x and y is $(x,y) = x_a y^a$ and the norm of x is $|x| = (x\bar{x})^{1/2}$. Nonassociativity is clear from the multiplication table Eqs. (2.3)–(2.5). Ω is alternative in that if $x, y \in \Omega$ they obey the left and right alternativity conditions:

$$xy^2 - (xy)y = 0 \tag{2.6a}$$

and

$$x^2 y - x(xy) = 0, \tag{2.6b}$$

respectively. Since power associativity holds for any alternative algebra $x^n x^m = x^{n+m}$ for any $x \in \Omega$. In our work three other completely antisymmetric objects are constantly used. If $x, y, u, v \in \Omega$ they are in addition to the commutator

$$[x, y] = (xy - yx) = -[\bar{x}, y] \tag{2.7}$$

and the associator

$$[x, y, z] = (xy)z - x(yz) = -[\bar{x}, y, z], \tag{2.8}$$

since only the purely vectorial parts of these octonions contribute. $[x, y, z]$ is fully antisymmetric in (x, y, z) because of alternativity. We also have the Kleinfeld product[22]:

$$K(x, y, u, v) = [xy, u, v] - y[x, u, v] - [y, u, v]x, \tag{2.9}$$

which, despite appearance, is totally antisymmetric in x, y, u, and v. We refer the interested reader to a list of handy octonions formulas in a paper of Yokota.[27]

Leaving details to be filled in later, we recall that closely connected with octonions are the orthogonal groups SO(8) and SO(7), the norm groups of the octonions and the pure imaginary octonions, respectively, and the simply connected compact exceptional Lie group G_2. The latter plays the central role in our work, being the automorphism group of Ω. Namely if $x, y \in \Omega$ and $\lambda \in R$ and if $g \in G_2$ and is nonsingular, then g satisfies the following defining relations $g(x + y) = gx + gy$, $g(x\lambda) = (gx)\lambda$, and $g(xy) = g(x)g(y)$. We then deduce that $g(e_0) = e_0$ for any $g \in G_2$. Hence the automorphisms of Ω leave each real number fixed while preserving the norm and the multiplication table; they can be viewed as a topological transformation group of the six-sphere $S^6 = \mathrm{SO}(7)/\mathrm{SO}(6) = G_2/\mathrm{SU}(3)$, the identity component being the group G_2. SU(3), a maximal subgroup of G_2, is con-

nected to Ω in that it consists of all elements $g \in G_2$ such that $g(e_7) = e_7$ and acts transitively on $S^6 = G_2/\mathrm{SU}(3) = \mathrm{SO}(7)/\mathrm{SO}(6)$.

If we now let $\tilde{\Omega}$ be a R-submodule of elements generated additively by the pure imaginary units e_α ($\alpha = 1,2,...,7$), then $\tilde{\Omega}$ is also a G_2-R-module. We then see that Ω is decomposable into a direct sum of two G_2-R-modules, $\Omega = R \oplus \tilde{\Omega}$, R being the real number field, the subspace spanned by the identity element. Thus eight-space has a natural $7+1$ decomposition. $\tilde{\Omega}$, the seven-space of pure imaginary octonions, also bears the name of Cayley space.[17]

With a more explicit treatment to be given in Sec. III, we only mention that the Lie algebra of SO(7) of $\tilde{\Omega}$ consists of all R-homomorphisms A of Ω obeying

$$A(e_0) = 0, \tag{2.10}$$

and

$$(A(x), y) + (x, A(y)) = 0 \quad \forall x, y \in \Omega. \tag{2.11}$$

If we denote by $G_{\alpha\beta}$ ($\alpha, \beta = 1,2,...,7$, $\alpha \neq \beta$) the R-homomorphisms are given by

$$G_{\alpha\beta}(e_\beta) = e_\alpha, \quad G_{\alpha\beta}(e_\alpha) = -e_\beta,$$
$$G_{\alpha\beta}(e_\gamma) = 0 \text{ for } \gamma \neq \alpha, \beta, \quad 0 \leqslant \gamma \leqslant 7, \tag{2.12}$$

and the 21 elements of $G_{\alpha\beta}$ ($1 \leqslant \alpha < \beta \leqslant 7$) are the additive base in SO(7).

The G_2 subalgebra of SO(7) is made up of all $A \in \Omega$ such that

$$A(x)y + xA(y) = A(xy) \quad \text{for any } x, y \in \Omega. \tag{2.13}$$

The G_2 algebra g_2 is a G_2-R-module with the group operation

$$(xA)(y) = x(A(x^{-1}y)), \tag{2.14}$$

where $x \in G_2$, $A \in g_2$, and $y \in \Omega$.

Finally we consider SO(8), the norm group of Ω. The workings of its Triality Principle could be nicely exhibited by an embedding of SO(8) within the group F_4. So we consider the 26-dimensional representation of F_4 through the 3×3 matrix[8]

$$J = \begin{pmatrix} \alpha & c & \bar{b} \\ \bar{c} & \beta & a \\ b & \bar{a} & \gamma \end{pmatrix}, \tag{2.15}$$

where α, β, and γ are scalar and $a,b,c \in \Omega$ with $\alpha + \beta + \gamma = \mathrm{Tr}(J) = 0$.
If we now define[8] the Jordan associator by

$$[A,B,C] \equiv (A \cdot B) \cdot C - A \cdot (B \cdot C), \tag{2.16}$$

where $A \cdot B = \frac{1}{2}\{A,B\}$ and A, B, and C are octonionic Hermitian 3×3 matrices, the infinitesimal action of F_4 on J is given by[14]

$$\delta J = [H_1, J, H_2], \tag{2.17}$$

with $\mathrm{Tr}\, H_1 = \mathrm{Tr}\, H_2 = 0$. Therefore F_4 has $2 \times 26 = 52$ parameters and has O(8) as a subgroup.

Under the F_4 action, we have the invariants

$$\mathrm{Tr}\, J = \alpha + \beta + \gamma,$$
$$\tfrac{1}{2}\mathrm{Tr}(J^2) = \tfrac{1}{2}(\alpha^2 + \beta^2 + \gamma^2) + |a|^2 + |b|^2 + |c|^2,$$
$$\mathrm{Det}(J) = \alpha\beta\gamma - \alpha|a|^2 - \beta|b|^2 - \gamma|c|^2 + 2\,\mathrm{Sc}(abc).$$

The O(8) invariants are $\alpha, \beta, \gamma, |a|^2, |b|^2, |c|^2$, and $\mathrm{Sc}(abc)$.

Now let $J \times J = J^{-1}\,\mathrm{Det}(J)$ so that $J \cdot (J \times J) = I\,\mathrm{Det}(J)$. We find

$$J \times J = \begin{pmatrix} \beta\gamma - |a|^2 & \bar{b}\bar{a} - \gamma c & ca - \beta\bar{b} \\ ab - \gamma\bar{c} & \gamma\alpha - |b|^2 & \bar{c}\bar{b} - \alpha a \\ \bar{a}\bar{c} - \beta b & bc - \alpha\bar{a} & \alpha\beta - |c|^2 \end{pmatrix}, \tag{2.18}$$

which also transforms like the 26-dimensional representation of F_4. Thus $J \times J = 0$ is an F_4 invariant constraint. Let K be such that $K \times K = 0$, $\mathrm{Tr}(K) = 3$ then

$$K = \begin{pmatrix} \gamma^{-1}|b|^2 & \gamma^{-1}\bar{b}\bar{a} & \bar{b} \\ \gamma^{-1}ab & \gamma^{-1}|a|^2 & a \\ b & \bar{a} & \gamma \end{pmatrix} = \lambda\bar{\lambda}^T, \tag{2.19}$$

where $\lambda^T = (\gamma^{-1/2}\bar{b}, \gamma^{-1/2}a, \gamma^{1/2})$, $|a|^2 + |b|^2 = \gamma(3 - \gamma)$.

With γ being an O(8) invariant, we can take $|a| = |b| = 1$, $\gamma = 1$, and $\lambda^T = (\bar{b}, a, 1)$ is covariant under the O(8) subgroup. We have for $\gamma = 1$,

$$K = \begin{pmatrix} 1 & b\bar{a} & \bar{b} \\ ab & 1 & a \\ b & \bar{a} & 1 \end{pmatrix}. \tag{2.20}$$

The $\mathrm{SO}(8)/G_2$ transformations can be written as a combination of $K' = LKL^{-1}$ and $K'' = MKM^{-1}$ with

$$L = \begin{pmatrix} \bar{l} & & 0 \\ & l & \\ 0 & & l \end{pmatrix}, \quad M = \begin{pmatrix} m & & 0 \\ & m & \\ 0 & & \bar{m}^2 \end{pmatrix}, \tag{2.21}$$

where $|m| = |l| = 1$ so that L and M each has seven parameters. With $c = \bar{b}\bar{a}$ we can expand K, K', and K'' about the identity I:

$$K - I = \begin{pmatrix} 0 & c & \bar{b} \\ \bar{c} & 0 & a \\ b & \bar{a} & 0 \end{pmatrix},$$

$$K' - I = \begin{pmatrix} 0 & c' & \bar{b}' \\ \bar{c}' & 0 & a' \\ b' & \bar{a}' & 0 \end{pmatrix} = \begin{pmatrix} 0 & \bar{l}c\bar{l} & \bar{l}\bar{b} \\ l\bar{c}l & 0 & la \\ bl & \bar{a}\bar{l} & 0 \end{pmatrix}, \tag{2.22}$$

$$K'' - I = \begin{pmatrix} 0 & c'' & \bar{b}'' \\ \bar{c}'' & 0 & a'' \\ b'' & \bar{a}'' & 0 \end{pmatrix} = \begin{pmatrix} 0 & mc\bar{m} & m\bar{b}m^2 \\ m\bar{c}\bar{m} & 0 & mam^2 \\ \bar{m}^2 b\bar{m} & \bar{m}^2\bar{a}\bar{m} & 0 \end{pmatrix},$$

or we have

$$a' = la, \quad b' = bl, \quad c' = \bar{b}'\bar{a}' = \bar{l}c\bar{l},$$
$$a'' = mam^2, \quad b'' = \bar{m}^2 b\bar{m}, \quad c'' = \bar{b}''\bar{a}'' = mc\bar{m}, \tag{2.23}$$

from which it is apparent that $c'' = mc\bar{m}$ is a $\mathrm{SO}(7)/G_2$ transformation, $c' = lc\bar{l}$ is a SO(8)/SO(7) transformation with the others being all different $\mathrm{SO}(8)/G_2$ transformations. Now let us define the operations

$$T_1(x) = lx, \quad T_2(x) = xl,$$
$$T_3(x) = \bar{l}x\bar{l}, \tag{2.24}$$

and

$$R_1(x) = mxm^2, \quad R_2(x) = \bar{m}^2 x\bar{m},$$
$$R_3(x) = mx\bar{m}. \tag{2.25}$$

We verify that

$$\overline{T_2(b)}\,\overline{T_1(a)} = T_3(\bar{b}\bar{a}) = T_3(\overline{ab}) \text{ or } T_1(a)T_2(b) = \overline{T_3(\overline{ab})}, \tag{2.26}$$

$$\overline{R_2(b)}\,\overline{R_1(a)} = R_3(\bar{b}\bar{a}) = R_3(\overline{ab}) \text{ or } R_1(a)R_2(b) = \overline{R_3(\overline{ab})}, \tag{2.27}$$

which is a statement of the Principle of Triality,[14] namely given an action d_3 of SO(8) on Ω with $a,b \in \Omega$, then there exist two unique (up to a sign) others d_1 and $d_2 \in$ SO(8) such that

$$(d_1a)(d_2b) = \overline{d_3(\overline{ab})} \tag{2.28}$$

along with a cyclic permutation symmetry of the action of d_1, d_2, and d_3. Thus consider the transformation Eq. (2.26); it is equivalent to

$$(la)(bl) = \overline{l(\overline{ab})\bar{l}} = l(ab)l, \tag{2.29}$$

which is just one of the three Moufang identities.[14] Now since the F_4 invariants are also invariant under the discrete symmetric S^3 group of permutations of three objects a, b, and c we have

$$T_3(a)T_1(b) = \overline{T_2(\overline{ab})} \text{ and } T_2(a)T_3(b) = \overline{T_1(\overline{ab})}, \tag{2.30}$$

which translate into

$$(\bar{l}a\bar{l})(lb) = \overline{(\overline{ab})l} = \bar{l}(ab), \quad (al)(\bar{l}b\bar{l}) = (\overline{l\,\overline{ab}}) = (ab)\bar{l}. \tag{2.31}$$

These are recognized as the remaining two Moufang identities: $(yay)z = y[a(yz)]$ and $z(yay) = [(zy)a]y$, a, y, and $z \in \Omega$.[14] Therefore the latter are seen in a new light as consequences of the Triality Principle.

We will not perform a similar exercise for the R_i, $i = 1,2,3$ defined above. It suffices to mention that the Triality formulation given above provides a compact and elegant basis for the detailed study of parallellism on exceptional manifold such as $S^7 = \mathrm{SO}(7)/G_2$. A more detailed account will be the topic of another work. This concludes our brief summary on octonions.

III. VECTOR PRODUCTS, OCTONIONIC DUALITY, AND IDENTITIES

An r-fold vector cross product in R^n is a multilinear map $P_r: R^{nr} = R^n \times R^n \times \cdots \times R^n \to R^n$ such that if $\langle\ ,\ \rangle$ is a nondegenerate bilinear form in R^n, for every set of vectors $(a_1,a_2,\ldots,a_r) \in R^{nr}$ we have the conditions (a) $P_r(a_1,a_2,\ldots,a_r)a_i = 0$, i.e., every vector is orthogonal to P_r, and (b) $P_r(a_1,a_2,\ldots,a_r) = \det(\langle a_i,a_j\rangle)$. It was proved[28] that R^n admits an r-fold vector cross product only in the following cases:

$$\begin{aligned} &(1)\ r = 1 \text{ and } n \text{ even},\\ &(2)\ r = n-1,\\ &(3)\ r = 2 \text{ and } n = 3,7,\\ &(4)\ r = 3 \text{ and } n = 8. \end{aligned} \tag{3.1}$$

Explicit expressions for these vector products are known.[21] With also an eye on possible future physical applications our first task is to put them in a covariant form so as to facilitate subsequent generalizations. Thus the usual "vector product" in R^n corresponds to $r = 2$, $n = 3$ in Cases (2) and (3). This twofold product is obtained by identifying R^3 with the pure quaternions and for any two vectors $a,b \in R^3$ we have $P_2(a,b) = ab - a\cdot b = \mathrm{Vec}(ab)$. Or it suffices to take for a and b two basic imaginary units e_α and e_β, then $P_2(e_\alpha,e_\beta) = \epsilon_{\alpha\beta\gamma}e_\gamma$ is just another expression of the remarkable existence on R^3, the tangent space of three-sphere S^3, of multiplication table of the imaginary unit quaternions $e_\alpha e_\beta = -\delta_{\alpha\beta} + \epsilon_{\alpha\beta\gamma}e_\gamma$ ($\alpha,\beta = 1,2,3$). In the same vein, in the case of $r = 2$, $n = 7$, we identify R^7 with the pure Cayley numbers and the basic twofold product $P_2(e_\alpha,e_\beta) = \psi_{\alpha\beta\gamma}e_\gamma$ is again just a statement of the existence of a multiplication table Eq. (2.4) on R^7, the tangent space of S^7. The kernel of the cross product is encapsuled in the Cayley structure constants, the basic algebraic objects throughout this work. We should mention that the existence of vector cross products in Cases (2) ($r = 2,n = 3$) and (3) is tied in a one-to-one way to the following sequence of celebrated theorems of K-theory[29]: (1) The sphere S^m is parallelizable if and only if $m = 1$, 3, or 7, this property is induced by the multiplication of complex numbers, quaternions, and octonions, (2) R^n admits the structure of a real division algebra if and only if, $n = 1$, 2, 4, or 8, respectively, (3) among the spheres only S^2 and S^6 admit an almost complex structure.

Proceeding to the threefold vector products, we observe that for octonions Case (4) of Eq. (3.1) has a four-dimensional analog in Case (2) for $r = 3,n = 4$. In fact the vector product is given[8] by the Hodge star operators of the associated bilinear form, of signature $(2p,n-2p)$. Their (anti)automorphism groups are $(\mathrm{O}(2p,n-2p))\ \mathrm{SO}(2p,n-2p)$ $(0 < 2p < 4)$. Here for a given bilinear form, say of signature (0,2) or (2,0), the autormorphism group is then O(4), there are precisely two anti-isomorphic threefold vector products. We easily recognize in them the self- and anti-self-duality structure of Riemannian four-space.[15,30] To give a new parallel treatment to the octonionic threefold product, we must first consider in Cayley space not just the G_2 invariants, the structure constants $\psi_{\alpha\beta\gamma}$ but also what we call[11] the "Cayley curvature" constants $\varphi_{\alpha\beta\gamma\rho}$. They are defined by the basic associator of any three imaginary units

$$[e_\alpha,e_\beta,e_\gamma] \equiv 2\varphi_{\alpha\beta\gamma\rho}e_\rho, \tag{3.2}$$

$$= \tfrac{1}{3}\Psi^{\sigma}_{[\alpha\beta}\Psi^{\sigma}_{\gamma]\rho}e_\rho, \tag{3.3}$$

where the G_2 invariant fully antisymmetric four-index object $\varphi_{\alpha\beta\gamma\rho}$ is given through the identity

$$\varphi_{\alpha\beta\gamma\rho} = (\delta_{\beta\gamma}\delta_{\alpha\rho} - \delta_{\alpha\gamma}\delta_{\beta\rho}) + \Psi_{\alpha\beta\sigma}\Psi_{\gamma\rho\sigma} \tag{3.4}$$

and the square bracket antisymmetrization ranges through all the elements of the symmetric group. Moreover φ and ψ are dual to each other in seven-dimensional space as shown by the fundamental relation[11]:

$$\varphi_{\alpha\beta\gamma\rho} = -\tfrac{1}{3!}\epsilon_{\alpha\beta\gamma\rho\sigma\mu\nu}\psi_{\sigma\mu\nu}, \tag{3.5}$$

$$\psi_{\alpha\beta\gamma} = -\tfrac{1}{4!}\epsilon_{\alpha\beta\gamma\rho\sigma\mu\nu}\varphi_{\rho\sigma\mu\nu}, \tag{3.6}$$

$$= -\tfrac{1}{4}\epsilon_{\alpha\beta\gamma\rho\sigma\mu\nu}\psi_{\rho\sigma\lambda}\psi_{\mu\nu\lambda}. \tag{3.7}$$

The proof of Eqs. (3.5) and (3.6) is given by merely reading off the following suitable variant of Cayley's multiplication table:

$$\psi_{\alpha\beta\gamma}\begin{cases} 1 & 2 & 4 & 3 & 6 & 5 & 7\\ 2 & 4 & 3 & 6 & 5 & 7 & 1\\ 3 & 6 & 5 & 7 & 1 & 2 & 4 \end{cases} \tag{3.8}$$

and

$$\varphi_{\alpha\beta\gamma\rho}\begin{cases}4&3&6&5&7&1&2\\5&7&1&2&4&3&6\\6&5&7&1&2&4&3\\7&1&2&4&3&6&5\end{cases}. \tag{3.9}$$

This table reads as follows: Each row is a cyclic permutation of the first row, the columns in the first rows give the structure constants or half the commutators of the imaginary units e_α, e.g., $e_1e_2 = e_3$, $e_2e_4 = e_6$, etc., while the columns of the remaining rows give half the associator of the e_α, e.g., $\frac{1}{2}[e_3,e_7,e_5] = -e_1$. Therefore they determine the $\varphi_{\alpha\beta\gamma\rho}$, e.g., if $\psi_{123} = 1$, $\varphi_{4567} = -1$, etc.

Now we make the crucial observation that from the numerical tensors $\psi_{\alpha\beta\gamma}$ and $\varphi_{\alpha\beta\gamma\rho}$ in Cayley space, we can construct a natural self-dual four-index totally antisymmetrical tensor f_{abcd} $(a,b,c,d = 0,1,...,7)$ in eight dimensions. The basic object is defined by

$$f_{\alpha\beta\gamma o} = -f_{o\alpha\beta\gamma} = \psi_{\alpha\beta\gamma}, \tag{3.10}$$

$$f_{\alpha\beta\gamma\rho} = \varphi_{\alpha\beta\gamma\rho} = -\tfrac{1}{3!}\epsilon_{\alpha\beta\gamma\rho\mu\nu\sigma}\psi_{\mu\nu\sigma} \quad (\alpha,\beta,\gamma,\rho = 1,2,...,7),$$

from which we verify the self-duality property

$$f_{abcd} = \tfrac{1}{4!}\epsilon_{abcdlmnp}f_{lmnp}. \tag{3.11}$$

The nonzero components of f_{abcd} are

$$f_{0123} = f_{0246} = f_{0435} = f_{0367} = f_{0651} = f_{0572} = f_{0714} = 1, \tag{3.12}$$

$$f_{4567} = f_{3571} = f_{6172} = f_{5214} = f_{7423} = f_{1346} = f_{2635} = -1. \tag{3.13}$$

Similarly we can define an anti-self-dual tensor f'_{abcd} by

$$f'_{o\alpha\beta\gamma} = -f_{o\alpha\beta\gamma}, \quad f'_{\alpha\beta\gamma\rho} = f_{\alpha\beta\gamma\rho}. \tag{3.14}$$

In complete analogy to the quaternionic case,[30] we now introduce the skew symmetric tensor octonion units e'_{ab},e_{ab} through

$$e'_{ab} = \tfrac{1}{2}(e_a\bar{e}_b - e_b\bar{e}_a), \quad e_{ab} = \tfrac{1}{2}(\bar{e}_a e_b - \bar{e}_b e_a). \tag{3.15}$$

These tensors satisfy the relations

$$e_{ab} = \tfrac{1}{3!}f_{abcd}e_{cd}, \quad e'_{ab} = -\tfrac{1}{3!}f'_{abcd}e'_{cd}, \tag{3.16}$$

namely, $e_{ab}(e'_{ab})$ is a skew tensor in $d = 8$ space which, with respect to the numerical tensor $f_{abcd}(f'_{abcd})$ is like a four-space tensor which is self-(anti-self-)dual with respect to the Levi–Civita symbol $\epsilon_{\alpha\beta\gamma\delta}$. To emphasize the fact that self-(anti-self-)duality here is with respect to $f_{abcd}(f'_{abcd})$ we call the $e_{ab}(e'_{ab})$ self-(anti-self-)coupled tensor in eight dimensions. We also have

$$f_{abcd} = -\mathrm{Sc}(e_{ab}e_{cd}),$$
$$f'_{abcd} = \mathrm{Sc}(e'_{ab}e'_{cd}). \tag{3.16a}$$

Making use of the Kronecker symbol δ^a_b to raise indices, we derive the following fundamental identity obeyed by the f_{abcd}:

$$\begin{aligned} f^{abcd}f_{ijkd} = {} & (\delta^a_i\delta^b_j - \delta^a_j\delta^b_i)\delta^c_k + (\delta^b_i\delta^c_j - \delta^b_j\delta^c_i)\delta^a_k \\ & + (\delta^c_i\delta^a_j - \delta^c_j\delta^a_i)\delta^b_k \\ & + f^{ab}{}_{ij}\delta^c_k + f^{bc}{}_{ij}\delta^a_k + f^{ca}{}_{ij}\delta^b_k \\ & + f^{ab}{}_{jk}\delta^c_i + f^{bc}{}_{jk}\delta^a_i + f^{ca}{}_{jk}\delta^b_i \\ & + f^{ab}{}_{ki}\delta^c_j + f^{bc}{}_{ki}\delta^a_j + f^{ca}{}_{ki}\delta^b_j. \end{aligned} \tag{3.17}$$

In an obvious abbreviated notation, it also reads

$$f^{abcd}f_{ijkd} = \delta^{[a}_i\delta^b_j\delta^{c]}_k + \tfrac{1}{4}f^{[ab}{}_{[ij}\delta^{c]}{}_{k\,]}. \tag{3.17a}$$

Self-duality tells us that f_{abcd} has 35 components

$$\tfrac{1}{2}(8\times7\times6\times5)\tfrac{1}{4!},$$

an irreducible O(8) tensor. However the existence of the identity Eq. (3.17) cuts down the number of independent components of f_{abcd} to $2\times7 = 14$ nonzero elements [as seen from Eqs. (3.12) and (3.13)], which are G_2 invariant.

By contraction we obtain from Eq. (3.17)

$$f^{abcd}f_{ijcd} = 6(\delta^a_i\delta^b_j - \delta^a_j\delta^b_i) + 4f^{ab}{}_{ij}, \tag{3.18}$$

$$f^{abcd}f_{ibcd} = 42\delta^a_i, \tag{3.19}$$

and

$$f^{abcd}f_{abcd} = 336. \tag{3.20}$$

Recalling the definitions $f_{o\alpha\beta\gamma} = -\psi_{\alpha\beta\gamma}$ and $f_{\alpha\beta\gamma\rho} = \varphi_{\alpha\beta\gamma\rho}$, we derive for the ψ's and φ's the following seven-dimensional algebraic identities:

$$\psi^{\alpha\beta\gamma}\psi_{\mu\nu\gamma} = \delta^\alpha_\mu\delta^\beta_\nu - \delta^\alpha_\nu\delta^\beta_\mu + \varphi^{\alpha\beta}{}_{\mu\nu}, \tag{3.21}$$

$$\varphi^{\alpha\beta\gamma\tau}\varphi_{\mu\nu\sigma\tau} = \delta^{[\alpha}_\mu\delta^\beta_\nu\delta^{\gamma]}_\sigma - \psi^{\alpha\beta\gamma}\psi_{\mu\nu\sigma} + \tfrac{1}{4}\varphi^{[\alpha\beta}{}_{[\mu\nu}\delta^{\gamma]}_{\sigma]}, \tag{3.22}$$

$$\begin{aligned} \psi^{\alpha\beta\gamma}\varphi_{\mu\nu\sigma\gamma} &= \psi^\alpha{}_{\mu\nu}\delta^\beta_\sigma - \psi^\beta{}_{\mu\nu}\delta^\alpha_\sigma + \psi^\alpha{}_{\nu\sigma}\delta^\beta_\mu \\ &\quad - \psi^\beta{}_{\nu\sigma}\delta^\alpha_\mu + \psi^\alpha{}_{\sigma\mu}\delta^\beta_\nu - \psi^\beta{}_{\sigma\mu}\delta^\alpha_\nu \\ &= \tfrac{1}{2}\Psi^{[\alpha}{}_{[\mu\nu}\delta^{\beta]}_{\sigma]}, \end{aligned} \tag{3.23}$$

$$\varphi^{\alpha\beta\gamma\tau}\varphi_{\mu\nu\gamma\tau} = 4(\delta^\alpha_\mu\delta^\beta_\nu - \delta^\alpha_\nu\delta^\beta_\mu) + 2\varphi^{\alpha\beta}{}_{\mu\nu}, \tag{3.24}$$

$$\psi^{\alpha\beta\gamma}\varphi_{\mu\nu\beta\gamma} = 4\psi^\alpha{}_{\mu\nu}, \quad \psi^{\alpha\beta\gamma}\varphi_{\mu\alpha\beta\gamma} = 0, \tag{3.25}$$

$$\psi^{\alpha\beta\gamma}\psi_{\mu\beta\gamma} = 6\delta^\alpha_\mu, \quad \psi^{\alpha\beta\gamma}\psi_{\alpha\beta\gamma} = 42, \tag{3.26}$$

$$\varphi^{\alpha\beta\gamma\tau}\varphi_{\mu\beta\gamma\tau} = 24\delta^\alpha_\mu, \quad \varphi^{\alpha\beta\gamma\tau}\varphi_{\alpha\beta\gamma\tau} = 168. \tag{3.27}$$

Also we recover the key duality connection Eqs. (3.5) and (3.6). From Eq. (3.21), we further derive

$$\begin{aligned} \psi_{\alpha[\beta}{}^\gamma\psi_{\mu\nu]\gamma} &= 2\psi_{\alpha\beta}{}^\gamma\psi_{\mu\nu\gamma} + 2\psi_{\alpha\mu}{}^\gamma\psi_{\nu\beta\gamma} \\ &\quad + 2\psi_{\alpha\nu}{}^\gamma\psi_{\beta\mu\gamma} = 6\varphi_{\alpha\beta\mu\nu}, \end{aligned} \tag{3.28}$$

or

$$\begin{aligned} \varphi_{\alpha\beta\mu\nu} &= \tfrac{1}{3!}\psi_{\alpha[\beta}{}^\gamma\psi_{\mu\nu]\gamma} = \tfrac{1}{3}(\psi_{\alpha\beta}{}^\gamma\psi_{\mu\nu\gamma} \\ &\quad + \psi_{\alpha\mu}{}^\gamma\psi_{\nu\beta\gamma} + \psi_{\alpha\nu}{}^\gamma\psi_{\beta\mu\gamma}), \end{aligned} \tag{3.29}$$

which upon insertion into Eq. (3.21) gives

$$\begin{aligned} \delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\nu}\delta_{\beta\mu} &= \psi_{\alpha\beta}{}^\gamma\psi_{\mu\nu}{}^\gamma - \tfrac{1}{3!}\psi_{\alpha[\beta}{}^\gamma\Psi_{\mu\nu]\gamma}, \qquad (3.30) \\ &= \psi_{\alpha\beta}{}^\gamma\psi_{\mu\nu}{}^\gamma - \tfrac{1}{3}(\psi_{\alpha\beta}{}^\gamma\psi_{\mu\nu\gamma} \\ &\quad + \psi_{\alpha\mu}{}^\gamma\psi_{\nu\beta\gamma} + \psi_{\alpha\nu}{}^\gamma\psi_{\beta\mu\gamma}), \qquad (3.31) \end{aligned}$$

alias Eq. (3.4). We observe that if the ψ are replaced by the structure constants $C_{\alpha\beta\gamma}$ of a Lie algebra, then $C_{\alpha[\beta}{}^\gamma C_{\mu\nu]\gamma} = 0$ is a statement of the Jacobi identity, i.e., of the associativity property and the second term in Eq. (3.30) vanishes. The left-hand side of Eq. (3.30) is modulo a multiplicative constant the Riemann curvature of an Einstein space, the Lie group space in an orthonormal frame. However there is no seven-dimensional Lie group space, Eq. (3.30) only holds true in seven-dimensions[31] for the conformally flat round seven-sphere with its standard Einstein metric. It gives the orthonormal frame components of the Riemannian curvature of S^7.

An alternative expression for the left-hand side of Eq. (3.30) is

$$\delta_{\alpha\mu}\delta_{\beta\nu}-\delta_{\alpha\nu}\delta_{\beta\mu}=\tfrac{1}{4}\varphi_{\alpha\beta}{}^{\gamma\tau}\varphi_{\mu\nu\gamma\tau}-\tfrac{1}{6}(\psi_{\alpha\beta}{}^{\gamma}\psi_{\mu\nu\gamma}+\psi_{\alpha\mu}{}^{\gamma}\psi_{\nu\beta\gamma}+\psi_{\alpha\nu}{}^{\gamma}\psi_{\beta\mu\gamma}). \quad (3.32)$$

The above relations exhaust the quadratic identities. In a previous work[11] some of these seven-dimensional identities have been used to obtain an explicit, purely algebraic solutions to the Englert solutions on S^7.

Moving on to cubic identities satisfied by the f_{abcd}, we shall evaluate the rank 6 tensor $T_{abcdeg}=f^{i}_{abj}f^{j}_{cdk}f^{k}_{egi}$. Making use of Eq. (3.17) twice we find

$$\begin{aligned}T_{abcdeg}=&\,4[\delta_{ac}(\delta_{bg}\delta_{de}-\delta_{be}\delta_{dg})-\delta_{bc}(\delta_{ag}\delta_{de}-\delta_{ae}\delta_{dg})\\&-\delta_{ad}(\delta_{bg}\delta_{ce}-\delta_{be}\delta_{cg})+\delta_{bd}(\delta_{ag}\delta_{ce}-\delta_{ae}\delta_{cg})]\\&-\delta_{ac}f_{bdeg}+\delta_{bc}f_{adeg}+\delta_{ad}f_{bceg}-\delta_{bd}f_{aceg}\\&-\delta_{ce}f_{dgab}+\delta_{de}f_{cgab}+\delta_{cg}f_{deab}-\delta_{dg}f_{ceab}\\&-\delta_{ea}f_{gbcd}+\delta_{ga}f_{ebcd}+\delta_{eb}f_{gacd}-\delta_{gb}f_{eacd}.\end{aligned} \quad (3.33)$$

We also verified that the T_{abcdeg} obey the identity

$$T_{abcdeg}+T_{cdegab}+T_{egabcd}=3T_{abcdeg}=3T_{cdegab}=3T_{egabcd} \quad (3.34)$$

from the cyclic symmetry of ab, cd, and eg in the definition of T. As before we work out the special cases in seven dimensions. They are

$$\psi_{\alpha\beta}{}^{\sigma}\psi_{\sigma\gamma}{}^{\mu}\psi_{\mu\nu}{}^{\alpha}=3\psi_{\beta\gamma\nu}, \quad (3.35)$$

$$\psi_{\alpha\beta}{}^{\sigma}\psi_{\sigma\gamma}{}^{\mu}\varphi_{\mu\nu\tau}{}^{\alpha}=4(\delta_{\beta\tau}\delta_{\gamma\nu}-\delta_{\beta\nu}\delta_{\gamma\tau})-\varphi_{\beta\gamma\nu\tau}, \quad (3.36)$$

$$\begin{aligned}&\gamma^{\alpha}{}_{\beta\gamma}\varphi^{\gamma}{}_{\sigma\tau\lambda}\varphi^{\lambda}{}_{\mu\nu\alpha}\\&\quad=2(\delta_{\beta\mu}\psi_{\nu\sigma\tau}-\delta_{\beta\nu}\psi_{\mu\sigma\tau}+\delta_{\beta\tau}\psi_{\sigma\mu\nu}-\delta_{\beta\sigma}\psi_{\tau\mu\nu}),\end{aligned} \quad (3.37)$$

$$\begin{aligned}&\varphi^{\alpha}{}_{\beta\gamma\sigma}\varphi^{\sigma}{}_{\mu\nu\lambda}\varphi^{\lambda}{}_{\tau\rho\alpha}\\&\quad=\mathscr{A}[\delta_{\beta\mu}\delta_{\gamma\rho}\delta_{\nu\tau}+\delta_{\beta\mu}\varphi_{\gamma\tau\rho\nu}+\delta_{\tau\beta}\varphi_{\rho\mu\nu\gamma}+\delta_{\mu\tau}\varphi_{\nu\beta\gamma\rho}\\&\qquad+\tfrac{1}{4}(\psi_{\beta\mu\nu}\psi_{\gamma\tau\rho}+\psi_{\mu\tau\rho}\psi_{\nu\beta\gamma}+\psi_{\tau\beta\gamma}\psi_{\rho\mu\nu})],\end{aligned} \quad (3.38)$$

where $\mathscr{A}$ stands for antisymmetrization of μ and ν, β and γ, τ and ρ.

IV. GENERALIZED VECTOR PRODUCTS

In terms of the above basic invariant tensors e_{ab} and f_{abcd}, we can now give a unified covariant formulation of the twofold, threefold, and finally new fourfold products in eight dimensions. Given two octonions a and b with components a_l and b_n we can define two antisymmetric double products

$$a\wedge b=a_l b_n e_{ln}=\bar{a}b-(a,b), \quad (4.1)$$

where $(a,b)=\mathrm{Sc}(\bar{a}b)=\mathrm{Sc}(a\bar{b})=\frac{1}{2}(\bar{a}b+\bar{b}a)$ and

$$(a\wedge b)'=a_l b_n e'_{ln}=a\bar{b}-(a,b). \quad (4.2)$$

As to the triple products, let us define the entities $e_{abc}=f_{abcd}e_d$ and $e'_{abc}=-f'_{abcd}e_d$. They are completely antisymmetric. We have the following defining relations:

$$e_{\alpha\beta\gamma}=\psi_{\alpha\beta\gamma}+\varphi_{\alpha\beta\gamma\rho}e_\rho, \quad (4.3)$$

$$e'_{\alpha\beta\gamma}=\psi_{\alpha\beta\gamma}-\varphi_{\alpha\beta\gamma\rho}e_\rho, \quad (4.4)$$

$$e_{o\beta\gamma}=-\psi_{\beta\gamma\rho}e_\rho,\quad e'_{o\beta\gamma}=-\psi_{\beta\gamma\rho}e_\rho. \quad (4.5)$$

On the other hand, recalling Eq. (3.15), let us consider the octonions $E_{abc}=e_a e_{bc}$, where

$$\begin{aligned}E_{\alpha\beta\gamma}&=\psi_{\alpha\beta\gamma}+\psi_{\beta\gamma\rho}\psi_{\alpha\sigma\rho}e_\sigma\\&=\psi_{\alpha\beta\gamma}+\varphi_{\alpha\beta\gamma\rho}e_\rho+\delta_{\alpha\beta}e_\gamma-\delta_{\gamma\alpha}e_\beta,\end{aligned} \quad (4.6)$$

$$E_{o\alpha\beta}=\psi_{\alpha\beta\gamma}e_\gamma,\quad E_{\alpha\beta o}=-\psi_{\alpha\beta\rho}e_\rho,\quad E_{\alpha o\beta}=-\psi_{\beta\alpha\rho}e_\rho. \quad (4.7)$$

Hence we can write

$$e_{\alpha\beta\gamma}=\tfrac{1}{3}(E_{\alpha\beta\gamma}+E_{\beta\gamma\alpha}+E_{\gamma\alpha\beta}), \quad (4.8)$$

$$e_{o\beta\gamma}=\tfrac{1}{3}(E_{o\beta\gamma}+E_{\beta\gamma o}+E_{\gamma o\beta}). \quad (4.9)$$

Also

$$E_{\alpha\beta\gamma}=e_{\alpha\beta\gamma}+\delta_{\alpha\beta}e_\gamma-\delta_{\alpha\gamma}e_\beta. \quad (4.10)$$

Consequently,

$$e_{\alpha\beta\gamma}=e_\alpha(\bar{e}_\beta e_\gamma)-e_\alpha\delta_{\beta\gamma}+e_\beta\delta_{\gamma\alpha}-e_\gamma\delta_{\alpha\beta}, \quad (4.11)$$

$$e_{o\beta\gamma}=e_o(\bar{e}_\beta e_\gamma)-e_o\delta_{\beta\gamma}+e_\beta\delta_{\gamma o}-e_\gamma\delta_{o\beta}. \quad (4.12)$$

Finally in light of the above we can write the following symmetrical formula:

$$e_{abc}=\tfrac{1}{3}(e_a e_{bc}+e_b e_{ca}+e_c e_{ab}), \quad (4.13)$$

$$=e_a(\bar{e}_b e_c)-e_a\delta_{bc}+e_b\delta_{ca}-e_c\delta_{ab}=f_{abcd}e_d. \quad (4.14)$$

Similarly we introduce

$$E'_{abc}=e'_{ab}e_c, \quad (4.15)$$

$$E'_{\alpha\beta\gamma}=\psi_{\alpha\beta\gamma}-\varphi_{\alpha\beta\gamma\rho}e_\rho-\delta_{\gamma\alpha}e_\beta+\delta_{\beta\gamma}e_\alpha, \quad (4.16)$$

$$=(e_\alpha\bar{e}_\beta)e_\gamma-e_\gamma\delta_{\alpha\beta}. \quad (4.17)$$

From which it follows that

$$\begin{aligned}e'_{\alpha\beta\gamma}&=\tfrac{1}{3}(e'_{\alpha\beta}e_\gamma+e'_{\beta\gamma}e_\alpha+e'_{\gamma\alpha}e_\beta)\\&=(e_\alpha\bar{e}_\beta)e_\gamma-e_\alpha\delta_{\beta\gamma}+e_\beta\delta_{\gamma\alpha}-e_\gamma\delta_{\alpha\beta}.\end{aligned} \quad (4.18)$$

On the other hand

$$E'_{o\beta\gamma}=-\psi_{\beta\gamma\rho}e_\rho=(e_o\bar{e}_\beta)e_\gamma-e_o\delta_{\beta\gamma}+e_\beta\delta_{\gamma o}-e_\gamma\delta_{o\beta}. \quad (4.19)$$

Hence

$$\begin{aligned}e'_{abc}&=\tfrac{1}{3}(e'_{ab}e_c+e'_{bc}e_a+e'_{ca}e_b)\\&=(e_a\bar{e}_b)e_c-e_a\delta_{bc}+e_b\delta_{ca}-e_c\delta_{ab}\\&=-f'_{abcd}e_d.\end{aligned} \quad (4.20)$$

With the above constructed fundamental tensors, it is an easy matter to define natural triple products for any three octonions $a=a_l e_l$, $b=b_l e_l$, and $c=c_l e_l$. They are

$$\begin{aligned}P_3(abc)&=a^l b^m c^n e_{lmn}=f_{lmnp}a^l b^m c^n e^p\\&=a(\bar{b}c)-a(b,c)+b(c,a)-c(a,b),\end{aligned} \quad (4.21)$$

$$\begin{aligned}P'_3(abc)&=a^l b^m c^n e'_{lmn}=-f'_{lmnp}a^l b^m c^n e^p\\&=(a\bar{b})c-a(b,c)+b(c,a)-c(a,b),\end{aligned} \quad (4.22)$$

$$P_3(abc)-P'_3(abc)=-[a,\bar{b},c]=[a,b,c]. \quad (4.23)$$

They are precisely the two triple products defined by Zvengrowski[21] and Gray[18] expressed here in terms of the tensors f_{abcd} and f'_{abcd}. That as in four-space there are two such products is connected to the twofold orientability of eight-space.[8]

The above noted relation[8] between vector product and curvature as exemplified by Eq. (3.30) leads us to seek fourfold vector products in eight-space. It is given by

$$E_{abcd}=\bar{e}_a e_{bcd}=f_{bcdl}\bar{e}_a e_l=f_{bcda}+f_{bcdl}e_{al}, \quad (4.24)$$

with

$$\mathrm{Sc}(E_{abcd}) = \tfrac{1}{2}(\bar{e}_a e_{bcd} + \bar{e}_{bcd} e_a) = -f_{abcd}. \tag{4.25}$$

Thus $\mathrm{Sc}(E_{abcd})$ is totally antisymmetric and self-dual. In fact we can define a totally antisymmetric tensor as

$$\begin{aligned} e_{abcd} &= \tfrac{1}{4}(E_{abcd} - \tfrac{1}{2}E_{[bcd]a}) \\ &= \tfrac{1}{4}(E_{abcd} - E_{bcda} - E_{cdba} - E_{dbca}). \end{aligned} \tag{4.26}$$

To see the usefulness of the E_{abcd}, we are led to analyze the perhaps less familiar Kleinfeld function,[22] Eq. (2.9), which is a totally antisymmetric homogeneous function of four octonions x, y, z and w. First we compute $K(x, y, z, w)$ when all four octonions are purely vectorial. There is no loss of generality in taking them as unit vectors. So we consider

$$\begin{aligned} K_{\alpha\beta\gamma\delta} &= K(e_\alpha, e_\beta, e_\gamma, e_\delta) \\ &= [e_\alpha e_\beta, e_\gamma, e_\delta] - e_\beta [e_\alpha, e_\gamma, e_\delta] - [e_\beta, e_\gamma, e_\delta] e_\alpha. \end{aligned} \tag{4.27}$$

By Eqs. (2.4) and (3.2) we get

$$K_{\alpha\beta\gamma\delta} = -2(\psi^{\alpha\beta\rho}\varphi_{\sigma\gamma\delta\rho} + \psi^{\sigma\beta\rho}{}_{\alpha\gamma\delta\rho} + \psi^{\alpha\sigma\rho}\varphi_{\beta\gamma\delta\rho}) e_\sigma. \tag{4.28}$$

By way of the identities Eq. (3.23), we find

$$-\tfrac{1}{4}K_{\alpha\beta\gamma\delta} = \psi_{\alpha\beta\gamma} e_\delta - \psi_{\alpha\beta\delta} e_\gamma + \psi_{\alpha\gamma\delta} e_\beta - \psi_{\beta\gamma\delta} e_\alpha. \tag{4.29}$$

Equation (4.29) is a $d = 7$ tensor since a composition of four purely vectorial octonions is again purely vectorial. Next we turn to the instance when one of the octonion units is $e_0 = 1$. Then

$$K_{\alpha\beta\gamma o} = [e_\alpha e_\beta, e_\gamma, e_o] - e_\beta [e_\alpha, e_\gamma, e_o] - [e_\beta, e_\gamma, e_o] e_\alpha = 0, \tag{4.30}$$

and

$$K_{o\beta\gamma\delta} = K_{\alpha o\gamma\delta} = K_{\alpha\beta o\delta} = 0. \tag{4.31}$$

Since $K_{\alpha\beta\gamma\delta}$ vanishes whenever two of the octonions in its arguments are equal, $K_{\alpha\beta oo} = 0$, $K_{\alpha ooo} = 0$, and $K_{oooo} = 0$. Consequently, in all cases K is a pure octonion.

Having made the above remarks, we can now seek the eight-dimensional form of the Kleinfeld function. We have

$$\tfrac{1}{4}K_{\alpha\beta\gamma\delta} = f_{o\alpha\beta\gamma} e_\delta - f_{o\alpha\beta\delta} e_\gamma + f_{o\alpha\gamma\delta} e_\beta - f_{o\beta\gamma\delta} e_\alpha, \tag{4.32}$$

$$\tfrac{1}{4}K_{\alpha\beta\gamma o} = 0. \tag{4.33}$$

Using the fact that $e_{ab} = \tfrac{1}{2}(\bar{e}_a e_b - \bar{e}_b e_a)$ with $e_o = \bar{e}_o$, $e_\alpha = -\bar{e}_\alpha$ for $\alpha = 1,2,\ldots,7$, then $e_{o\alpha} = +e_\alpha$ and we have

$$\tfrac{1}{4}K_{\alpha\beta\gamma\delta} = f_{o\alpha\beta\gamma} e_{o\delta} - f_{o\alpha\beta\delta} e_{o\gamma} + f_{o\alpha\gamma\delta} e_{o\beta} - f_{o\beta\gamma\delta} e_{o\alpha}. \tag{4.34}$$

Hence an eight-dimensional totally antisymmetric generalization of $K_{\alpha\beta\gamma\delta}$ is

$$H_{abcd} = -f_{mabc} e_{dm} + f_{mabd} e_{cm} + f_{madc} e_{bm} + f_{mdbc} e_{am}, \tag{4.35}$$

or

$$H_{abcd} = -\tfrac{1}{3!} f_{m[abc} e_{d]m}. \tag{4.36}$$

It is readily checked, after some algebra, that indeed

$$H_{\alpha\beta\gamma\delta} = K_{\alpha\beta\gamma\delta}, \tag{4.37}$$

and

$$H_{\alpha\beta\gamma o} = -4\varphi_{\rho\alpha\beta\gamma} e_\rho. \tag{4.38}$$

Note that $H_{abcd} = 0$ whenever $abcd$ correspond to the 14 nonzero values of f_{abcd}. Finally the connection with the covariant fourfold cross product is as follows:

$$H_{klmn} = \mathrm{Vec}(E_{klmn} - E_{lkmn} - E_{mlkn} - E_{nlmk}). \tag{4.39}$$

Equation (4.39) in this symmetrical form is therefore the $d = 8$ covariant generalizations of the Kleinfeld antisymmetric quadruple product. Since we have found[11] the Kleinfeld product and its properties to be most useful in computing nonassociative differential geometry on the seven-sphere, we expect the H_{abcd}, Eq. (4.39), to play at least an equally important role in the elucidation of local structure of $d = 8$ geometries.

As a final topic we wish to make contact with yet another vector cross product due to Ogievetski and Tzeitlin.[23] The latter authors introduced their product in order to represent the automorphisms of the Rosenfeld algebras, automorphisms which appear in the adjoint representations of exceptional groups. Thanks to this vector product, their resulting octonionic 3×3 matrix formulation of exceptional grand unified gauge theories is compact and simplified in an essential way. We begin by defining two operators E_α and $E_{\alpha\beta}$ acting on vectorial octonions by

$$E_\alpha x = \tfrac{1}{2}[e_\alpha, x], \quad x = x_\alpha e_\alpha, \tag{4.40}$$

and

$$E_{\alpha\beta} x = [E_\alpha, E_\beta] x + E_{\frac{1}{2}[e_\alpha, e_\beta]} x, \tag{4.41}$$

where

$$[E_\alpha, E_\beta] x = \tfrac{1}{4}[e_\beta, [e_\beta, x]] - \tfrac{1}{4}[e_\beta, [e_\alpha, x]], \tag{4.42}$$

$$E_{\frac{1}{2}[e_\alpha, e_\beta]} x = \tfrac{1}{2}[\tfrac{1}{2}[e_\alpha, e_\beta], x]. \tag{4.43}$$

We then define a vector product of E_α and E_β:

$$E_\alpha \wedge E_\beta = E_{\alpha\beta} = [E_\alpha, E_\beta] + E_{\frac{1}{2}[e_\alpha, e_\beta]}. \tag{4.44}$$

These operators can be represented by tensors

$$E_\alpha e_\mu = \tfrac{1}{2}[e_\beta, e_\mu] = \psi_{\alpha\mu\nu} e_\nu, \tag{4.45}$$

$$E_{\alpha\beta} e_\mu = L_{\alpha\beta\mu\nu} e_\nu, \tag{4.46}$$

where

$$L_{\alpha\beta\mu\nu} \equiv 2(\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\nu}\delta_{\beta\mu}) - \varphi_{\alpha\beta\mu\nu}, \tag{4.47}$$

which is the tensor analog of $E_\alpha \wedge E_\beta = E_{\alpha\beta}$. Using the identity Eq. (3.4) we can also write

$$L_{\alpha\beta\mu\nu} = 3(\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\alpha\nu}\delta_{\beta\mu}) - \psi_{\alpha\beta\lambda}\psi_{\mu\nu\lambda}, \tag{4.48}$$

which coincides with the Ogievetski–Tzeitlin form.[23] Moreover we can also verify that

$$E_{\alpha\beta} e_\mu = -\tfrac{3}{2}[e_\alpha, e_\beta, e_\mu] + \tfrac{1}{2}[[e_\alpha, e_\beta], e_\mu], \tag{4.49}$$

which is the standard G_2 transformation associated with the octonions e_α and e_β. Hence

$$\delta_{G_2} e_\mu = E_{\alpha\beta} e_\mu = L_{\alpha\beta\mu\nu} e_\nu \tag{4.50}$$

is the infinitesimal G_2 transformation of the octonion unit e_μ; there are 14 independent $E_{\alpha\beta}$ ($\psi_{\alpha\beta\gamma} E_\gamma = 0$).

Since we know how E_α acts on the unit e_μ we can compute the commutator of $E_{\alpha\beta}$ and E_γ. We find

$$\begin{aligned} [E_{\alpha\beta}, E_\gamma] e_\mu = (&2\psi_{\gamma\mu\alpha}\delta_{\beta\lambda} - 2\psi_{\gamma\mu\beta}\delta_{\alpha\lambda} \\ &- 2\psi_{\gamma\beta\lambda}\delta_{\alpha\mu} + 2\psi_{\gamma\alpha\lambda}\delta_{\beta\mu} \\ &+ \varphi_{\alpha\beta\lambda\nu}\psi_{\gamma\mu\nu} - \varphi_{\alpha\beta\mu\nu}\psi_{\gamma\lambda\nu}) e_\lambda. \end{aligned} \tag{4.51}$$

On the other hand

$$L_{\alpha\beta\gamma\nu}E_\nu e_\mu = L_{\alpha\beta\gamma\nu}\psi_{\nu\mu\lambda}e_\lambda, \tag{4.52}$$

which coupled with the identity Eq. (3.23) yields

$$[E_{\alpha\beta},E_\gamma] = L_{\alpha\beta\gamma\nu}E_\nu \tag{4.53}$$

with, as corresponding tensor identity,

$$L_{\alpha\beta\mu\nu}\psi_{\gamma\lambda\nu} - L_{\alpha\beta\lambda\nu}\psi_{\gamma\mu\nu} = L_{\alpha\beta\gamma\nu}\psi_{\mu\lambda\nu}, \tag{4.54}$$

or more compactly,

$$L_{[\mu}{}^{\alpha\beta\nu}\psi_{\gamma\lambda]\nu} = 0. \tag{4.55}$$

Then tensors $L_{\alpha\beta\mu\nu}$ are associated with G_2 while $\psi_{\alpha\beta\gamma}$ are associated with SO(7)/G_2 so that together the $E_{\alpha\beta}$ and E_γ form the 21 generators of SO(7). Now let the octonion units undergo a 21-parameter infinitesimal O(7) rotation. We have

$$\delta_R e_\mu = \Omega_{\alpha\beta}e_\mu = R_{\alpha\beta\mu\nu}e_\nu = \delta_{\alpha\mu}e_\beta - \delta_{\beta\mu}e_\alpha. \tag{4.56}$$

Therefore the tensor associated with the anti-Hermitian operators $\Omega_{\alpha\beta}$ of SO(7) reads $R_{\alpha\beta\mu\nu} = (\delta_{\alpha\mu}\delta_{\beta\nu} - \delta_{\beta\mu}\delta_{\alpha\nu})$. We then find

$$\begin{aligned}[\Omega_{\alpha\beta},\Omega_{\rho\sigma}]e_\mu &= (S_{\alpha\beta\rho\sigma\kappa\lambda}\Omega_{\kappa\lambda})e_\mu \\ &= (\delta_{\alpha\rho}\Omega_{\beta\sigma} - \delta_{\beta\rho}\Omega_{\alpha\sigma} \\ &\quad + \delta_{\beta\sigma}\Omega_{\alpha\rho} - \delta_{\alpha\sigma}\Omega_{\beta\rho})e_\mu,\end{aligned} \tag{4.57}$$

from which the structure constants of O(7) read

$$S_{\alpha\beta\rho\sigma\kappa\lambda} = \tfrac{1}{2}\mathscr{A}(\delta_{\alpha\rho}\delta_{\beta\kappa}\delta_{\sigma\lambda}), \tag{4.58}$$

where $\mathscr{A}$ stands for antisymmetrization of α and β, ρ and σ, and κ and λ.

The 14 generators of the G_2 subgroup of SO(7) are given by

$$E_{\alpha\beta} = \tfrac{1}{2}L_{\alpha\beta\kappa\lambda}\Omega_{\kappa\lambda} = 2\Omega_{\alpha\beta} - \tfrac{1}{2}\varphi_{\alpha\beta\kappa\lambda}\Omega_{\kappa\lambda}. \tag{4.59}$$

The seven operators E_α of SO(7)/G_2 are now extracted from $\Omega_{\kappa\lambda}$ by means of the structure constants $\psi_{\gamma\kappa\lambda}$, i.e.,

$$E_\gamma = \tfrac{1}{2}\psi_{\gamma\kappa\lambda}\Omega_{\kappa\lambda}. \tag{4.60}$$

Indeed with this definition we obtain

$$E_\gamma e_\mu = \psi_{\gamma\mu\nu}e_\nu, \tag{4.61}$$

which coincides with the operators introduced above. Also we find

$$E_{\alpha\beta}e_\mu = L_{\alpha\beta\mu\nu}e_\nu, \tag{4.62}$$

and thus recover our previous definition. We further note that due to Eq. (3.25),

$$\psi_{\gamma\alpha\beta}L_{\alpha\beta\mu\nu} = 0, \tag{4.63}$$

which shows the orthogonality of the SO(7)/G_2 and G_2 generators.

Finally either by evaluating $[E_{\alpha\beta},E_{\rho\sigma}]e_\mu$ directly or by way of the identities involving $\psi_{\alpha\beta\gamma}$ and $\varphi_{\alpha\beta\gamma\delta}$ we can obtain the tensor form of the G_2 structure constants. Since

$$\begin{aligned}[E_{\alpha\beta},E_{\rho\sigma}]e_\mu = [&3(\delta_{\alpha\rho}E_{\beta\sigma} - \delta_{\beta\rho}E_{\alpha\sigma} \\ &+ \delta_{\beta\sigma}E_{\alpha\rho} - \delta_{\alpha\sigma}E_{\beta\rho}) - \psi_{\alpha\beta\kappa}\psi_{\rho\sigma\lambda}E_{\kappa\lambda}]e_\mu,\end{aligned} \tag{4.64}$$

we extract the Lie product

$$E_{\alpha\beta}\wedge E_{\rho\sigma} = C_{\alpha\beta\rho\sigma\kappa\lambda}E_{\kappa\lambda}, \tag{4.65}$$

with

$$C_{\alpha\beta\rho\sigma\kappa\lambda} = \mathscr{A}[\tfrac{3}{2}\delta_{\alpha\rho}\delta_{\beta\kappa}\delta_{\sigma\lambda} - \tfrac{1}{8}\psi_{\alpha\beta\kappa}\psi_{\rho\sigma\lambda}] \tag{4.66}$$

[where $\mathscr{A}$ has the same meaning as before, see Eq. (4.58)] representing the G_2 structure constants in tensor form. We can also write

$$C_{\alpha\beta\rho\sigma\kappa\lambda} = 3S_{\alpha\beta\rho\sigma\kappa\lambda} - \tfrac{1}{2}(\psi_{\alpha\beta\kappa}\psi_{\rho\sigma\lambda} - \psi_{\alpha\beta\lambda}\psi_{\rho\sigma\kappa}), \tag{4.67}$$

where $S_{\alpha\beta\rho\sigma\kappa\lambda}$ are the O(7) structure constants in Eq. (4.58). From the above we can further derive an identity involving $L_{\alpha\beta\mu\nu}$. Thus, on the one hand, we have

$$[E_{\alpha\beta},E_{\rho\sigma}]e_\mu = L_{\rho\sigma\mu\nu}L_{\alpha\beta\nu\gamma}e_\gamma - L_{\alpha\beta\mu\nu}L_{\rho\sigma\nu\gamma}e_\gamma. \tag{4.68}$$

On the other hand, we have Eq. (4.65); therefore we deduce that

$$L_{\rho\sigma\mu\nu}L_{\alpha\beta\nu\gamma} - L_{\alpha\beta\mu\nu}L_{\rho\sigma\nu\gamma} = C_{\alpha\beta\rho\sigma\kappa\lambda}L_{\kappa\lambda\mu\gamma}, \tag{4.69}$$

corresponding to the Lie algebra of G_2. Of course, equipped with the formalism developed above we can pursue the course charted by Ogievetski and Tzeitlin[23] toward the remaining exceptional groups F_4, E_6, E_7, and E_8 of grand unified theories.

V. DECOMPOSITION OF SPIN(8) INTO $S^7_R \times S^7_L \times G_2$ AND TRIALITY

In this section we shall apply the formalism developed above to the decomposition of the group Spin(8) into Spin(8)/Spin(7) = S^7 plus Spin(7)/$G_2 = \Sigma^7$ plus G_2. Moreover, the coset space Spin(8)/G_2 can be further decomposed into generators associated with S^7_L and S^7_R which corresponds, respectively, to the left and right action of a unit octonion on a fixed octonion. We shall also characterize O(8) transformations in Triality by means of the tensors f and f' of Sec. III.

The 28 generators of the Spin(8) Lie algebra obey

$$[J_{ab},J_{cd}] = S_{abcdmn}J_{mn}, \tag{5.1}$$

where

$$\begin{aligned}S_{abcdmn} &= \delta^n_{[d}\delta^{[a}_{c]}\delta^{b]}_m = \delta^a_c\delta^b_m\delta^n_d \\ &\quad - \delta^b_c\delta^a_m\delta^n_d + \delta^b_d\delta^a_m\delta^n_c - \delta^a_d\delta^b_m\delta^n_c.\end{aligned} \tag{5.2}$$

Here the Latin indices range from 0 to 7. The Spin(7) subalgebra is represented by the 21 generators $J_{\alpha\beta}$, where the Greek indices range from 1 to 7:

$$[J_{\alpha\beta},J_{\gamma\delta}] = S_{\alpha\beta\gamma\delta\mu\nu}J_{\mu\nu}, \tag{5.3}$$

with

$$S_{\alpha\beta\gamma\delta\mu\nu} = \delta^\nu_{[\delta}\delta^{[\alpha}_{\gamma]}\delta^{\beta]}_{\mu}. \tag{5.4}$$

The coset Spin(8)/Spin(7) is associated with the generators

$$M_\alpha = J_{\alpha o} = -J_{o\alpha}. \tag{5.5}$$

Indeed we have

$$[M_\alpha,M_\beta] = J_{\alpha\beta}, \quad [J_{\alpha\beta},M_\gamma] = \delta_{\alpha\gamma}M_\beta - \delta_{\beta\gamma}M_\alpha \tag{5.6}$$

which, together with Eq. (5.3) constitute the Spin(8) algebra.

Let us consider a specific generator, say J_{12} of Spin(7). We can write

$$J_{12} = \tfrac{1}{3}(J_{12} + J_{54} + J_{67}) + \tfrac{1}{3}(2J_{12} - J_{54} - J_{67}). \tag{5.7}$$

Note that the index 3 is missing on the right-hand side. Thus, one can define

$$K_3 = J_{12} + J_{54} + J_{67}, \quad \Gamma_{12} = 2J_{12} - J_{54} - J_{67}, \tag{5.8}$$

so that

$$J_{12} = \tfrac{1}{3}(K_3 + \Gamma_{12}). \tag{5.9}$$

The general definition reads

$$K_\gamma = \tfrac{1}{2}\psi_{\alpha\beta\gamma}J_{\alpha\beta}, \quad \Gamma_{\alpha\beta} = 2J_{\alpha\beta} - \tfrac{1}{2}\varphi_{\alpha\beta\rho\sigma}J_{\rho\sigma}, \tag{5.10}$$

where $\psi_{\alpha\beta\gamma}$ and its dual $\varphi_{\alpha\beta\rho\sigma}$ have been defined by Eqs. (2.4) and (3.2). This splitting corresponds to the decomposition

$$J_{\alpha\beta} = (\tfrac{1}{3}J_{\alpha\beta} + \tfrac{1}{6}\varphi_{\alpha\beta\rho\sigma}J_{\rho\sigma}) + (\tfrac{2}{3}J_{\alpha\beta} - \tfrac{1}{6}\varphi_{\alpha\beta\rho\sigma}J_{\rho\sigma}) = \tfrac{1}{3}(K_{\alpha\beta} + \Gamma_{\alpha\beta}), \tag{5.11}$$

where

$$K_{\alpha\beta} = J_{\alpha\beta} + \tfrac{1}{2}\varphi_{\alpha\beta\rho\sigma}J_{\rho\sigma} = \psi_{\alpha\beta\gamma}K_\gamma. \tag{5.12}$$

So, for instance, we have

$$K_{12} = K_{54} = K_{67} = K_3 \tag{5.13}$$

and

$$\Gamma_{12} + \Gamma_{54} + \Gamma_{67} = 0, \tag{5.14}$$

The latter being just a special case of

$$\psi_{\alpha\beta\gamma}\Gamma_{\alpha\beta} = 0. \tag{5.15}$$

It follows that there are seven-generators K_γ and 14 independent generators $\Gamma_{\alpha\beta}$ which add up to those of Spin(7). It has been shown previously[26] that $\Gamma_{\alpha\beta}$ are the elements of the Lie algebra of G_2 and that K_γ are the elements of the coset space $\text{Spin}(7)/G_2 = \Sigma^7$. We recover these results directly by going over to the octonionic representation of the operators J_{ab} acting on the units e_a ($a = 0,1...,7$) in the ensuing manner:

$$\begin{aligned} &J_{\alpha\beta}e_\alpha = e_\beta, \quad J_{\alpha\beta}e_\beta = -e_\alpha (\text{no sum}),\\ &J_{\alpha\beta}e_\gamma = 0 \quad (\gamma\neq\alpha, \gamma\neq\beta),\\ &J_{\gamma o}e_\gamma = e_o, \quad J_{\gamma o}e_o = -e_\gamma,\\ &J_{\gamma o}e_\rho = 0 \quad (\rho\neq o, \rho\neq\gamma). \end{aligned} \tag{5.16}$$

Introducing the full octonion

$$X = e_a x_a = x_o + e_\alpha x_\alpha = x_o + x \quad (x = \text{Vec}\, X), \tag{5.17}$$

we find the following representations of the various generators of Spin(8):

(a) $\text{Spin}(8)/\text{Spin}(7) = S^7$:

$$M_\alpha X = J_{\alpha o}X = -\tfrac{1}{2}\{e_\alpha, X\}. \tag{5.18}$$

(b) $\text{Spin}(7)/G_2 = \Sigma^7\ (\sim S^7)$:

$$K_\gamma X = \tfrac{1}{2}\psi_{\alpha\beta\gamma}J_{\alpha\beta}X = \tfrac{1}{2}[e_\gamma, X] = \tfrac{1}{2}[e_\gamma, x]. \tag{5.19}$$

(c) G_2:

$$\Gamma_{\alpha\beta}X = -\tfrac{3}{2}[e_\alpha, e_\beta, X] + \tfrac{1}{2}[[e_\alpha, e_\beta], X] \doteq E_{\alpha\beta}x. \tag{5.20}$$

Hence, using Eq. (5.11) we find

$$J_{\alpha\beta}x = \tfrac{1}{3}(\psi_{\alpha\beta\gamma}K_\gamma + \Gamma_{\alpha\beta})x = \tfrac{1}{2}\psi_{\alpha\beta\gamma}[e_\gamma, x] - \tfrac{1}{2}[e_\alpha, e_\beta, x] \tag{5.21}$$

for the Spin(7) generators. For completeness and future applications in connection with $S^7_{L,R}$, we also derive the form of the left and right octonion multiplication in terms of J_{ab}. Making use of Eqs. (5.18) and (5.19), we readily find

$$(M_\gamma + K_\gamma)X = -Xe_\gamma, \quad (M_\gamma - K_\gamma)X = -e_\gamma X. \tag{5.22}$$

Thus:

$$\begin{aligned} &Xe_3 = -(J_{30} + J_{12} + J_{54} + J_{67})X,\\ &e_2X = (-J_{30} + J_{12} + J_{54} + J_{67})X. \end{aligned} \tag{5.23}$$

Observe that M_γ and K_γ are associated with unit octonions, so are $M_\gamma \mp K_\gamma$. In the latter case, since $e_\gamma(Xe_\gamma) = (e_\gamma X)e_\gamma$, the operators $M_\gamma + K_\gamma$ and $M_\gamma - K_\gamma$ commute. They correspond to the independent seven-spheres S^7_R and S^7_L. In this manner the decomposition of J_{ab} into $M_\gamma \mp K_\gamma$ and $\Gamma_{\alpha\beta}$ mirrors that of the Spin(8) group space into $S^7_R \times S^7_L \times G_2$.

We are now equipped to given an O(8) covariant formulation of the above Spin(8) decomposition. Let us define two kinds of couplings for the second-rank skew symmetric tensors ω_{ab} in R^8 by use of the self-dual tensor f_{abcd} and the anti-self-dual tensor f'_{abcd} introduced in Sec. III.

$$\omega^\#_{ab} = \tfrac{1}{2}f_{abcd}\omega_{cd}, \quad \omega^\flat_{ab} = \tfrac{1}{2}f'_{abcd}\omega_{cd}. \tag{5.24}$$

We may call them, respectively, sharp-coupled and flat-coupled to ω_{ab}. Using the decomposition of f and f' into $\psi_{\alpha\beta\gamma}$ and $\psi_{\alpha\beta\gamma\delta}$, we obtain the following formulas for the sharp and flat couplings:

$$G_2{:}\Gamma^\#_{\alpha\beta} = -\Gamma_{\alpha\beta}, \quad \Gamma^\flat_{\alpha\beta} = -\Gamma_{\alpha\beta}, \tag{5.25}$$

$$S^7(\text{Spin}(8)/\text{Spin}(7)){:}\ M^\#_\gamma = K_\gamma, \quad M^\flat_\gamma = -K_\gamma, \tag{5.26}$$

$$\Sigma^7(\text{Spin}(7)/G_2){:}\ K^\#_\gamma = 3M_\gamma + 2K_\gamma, \quad K^\flat_\gamma = -3M_\gamma + 2K_\gamma. \tag{5.27}$$

At this point, we define the operators

$$S^R_\gamma = \tfrac{1}{2}(M_\gamma + K_\gamma), \quad S^L_\gamma = \tfrac{1}{2}(M_\gamma - K_\gamma), \tag{5.28}$$

which, for each γ, are commuting generators in $\text{Spin}(8)/G_2$ associated, respectively, with S^7_R and S^7_L. From Eqs. (5.26) and (5.27) we find

$$S^7_R{:}(S^R_\gamma)^\# = 3S^R_\gamma, \tag{5.29}$$

$$S^7_L{:}(S^L_\gamma)^\flat = 3S^L_\gamma. \tag{5.30}$$

The remaining couplings define the new operators

$$\Sigma^R_\gamma = -(S^R_\gamma)^\flat = \tfrac{1}{2}(3M_\gamma - K_\gamma) \in \Sigma^7_R, \tag{5.31a}$$

and

$$\Sigma^L_\gamma = (S^L_\gamma)^\# = -\tfrac{1}{2}(3M_\gamma + K_\gamma) \in \Sigma^7_L, \tag{5.31b}$$

which are clearly associated with the infinitesimal forms of $b'' = \bar{m}^2 b\bar{m}$ (Σ^7_R) and $a'' = mam^2$ (Σ^7_L) of Eq. (2.23) when the unit octonion m is near unity. These transformations satisfy the relations

$$(\Sigma^R_\gamma)^\# = -\Sigma^R_\gamma, \tag{5.32a}$$

$$(\Sigma^L_\gamma)^\flat = -\Sigma^L_\gamma. \tag{5.32b}$$

Consequently S^7_R, Σ^7_R, and G_2 are eigenstates of sharp coupling with respective eigenvalues 3, -1, and -1; while S^7_L, Σ^7_L, and G_2 are eigenstates of flat coupling with the same respective eigenvalues 3, -1, and -1. We can also define the mixed coupling transformation:

$$\omega'_{ab} = \tfrac{1}{2}(\omega^\#_{ab} + \omega^\flat_{ab}). \tag{5.33}$$

We obtain

$$M'_\gamma = 0, \quad K'_\gamma = 2K_\gamma, \quad \Gamma'_{\alpha\beta} = -\Gamma_{\alpha\beta}, \tag{5.34}$$

$$(M_\gamma \pm K_\gamma)' = \pm 2K_\gamma, \quad (3M_\gamma \pm K_\gamma)' = \pm 2K_\gamma. \tag{5.35}$$

It follows that the $S^7 = \text{Spin}(8)/\text{Spin}(7)$, $\Sigma^7 = \text{Spin}(7)/G_2$ and G_2 transformations are eigenstates of mixed coupling with eigenvalues, respectively, given by 0, 2, and -1. This discussion then completes our covariant treatment of the

decomposition Spin(8), Spin(8) $= S^7_R \times S^7_L G_2$, and into S^7, Σ^7, and G_2 in terms of the tensors f and f'.

We close this section by stating results connected with Triality. If we write the transformations T_i and R_i $(i = 1,2,3)$ of Eqs. (2.26) and (2.27) in infinitesimal form, to each generator $C \in$ Spin(8) are associated two other transformations $A \in$ Spin(8) and $B \in$ Spin(8) such that

$$A + B + C = 0. \tag{5.36}$$

Let $\Omega =$ Spin(8) and define

$$A = \tfrac{1}{2}(\Omega - \Omega^{\flat}), \quad B = \tfrac{1}{2}(\Omega - \Omega^{\#}),$$
$$C = -\Omega + \Omega' = -\Omega + \tfrac{1}{2}(\Omega^{\#} + \Omega^{\flat}). \tag{5.37}$$

Then A, B, and C are in Triality. We prove this feature by letting Ω range over all elements of Spin(8).

(G_2,G_2,G_2) Triality: Let $\Omega = \Gamma_{\alpha\beta} \in G_2$. Then

$$A = B = \Gamma_{\alpha\beta} \in G_2, \quad C = -2\Gamma_{\alpha\beta} \in G_2 \tag{5.38}$$

so that the G_2 space has self-Triality.

(S^7_R,S^7_L,S^7) Triality: Let $\Omega = M_\gamma \in S^7 =$ Spin(8)/Spin(7). Then

$$A = \tfrac{1}{2}(M_\gamma + K_\gamma) = S^R_\gamma \in S^7_R, \tag{5.39a}$$
$$B = \tfrac{1}{2}(M_\gamma - K_\gamma) = S^L_\gamma \in S^7_L, \tag{5.39b}$$
$$C = -M_\gamma \in S^7. \tag{5.39c}$$

Now let $\Omega = -S^R_\gamma$. Using Eqs. (5.31a) and (5.29) we find

$$A = -\frac{1}{2}\left(\frac{M_\gamma + K_\gamma}{2} + \frac{3M_\gamma - K_\gamma}{2}\right) = -M_\gamma \in S^7, \tag{5.40}$$

$$B = -\frac{1}{2}\left(\frac{M_\gamma + K_\gamma}{2} - \frac{3M_\gamma + 3K_\gamma}{2}\right) = S^R_\gamma \in S^7_R,$$

and by use of Eq. (5.35) we obtain

$$C = \tfrac{1}{2}(M_\gamma + K_\gamma) - K_\gamma = S^L_\gamma \in S^7_L. \tag{5.41}$$

Therefore, the new triple (A,B,C) is a cyclic permutation of Eqs. (5.39).

Finally, upon choosing $\Omega = -S^L_\gamma$ we get

$$A = S^L_\gamma, \quad B = -M_\gamma, \quad C = S^R_\gamma, \tag{5.42}$$

which gives yet another permutation of Eqs. (5.39).

Thus, we have shown that the generators associated with S^7_R, S^7_L, and $S^7 =$ SO(8)/SO(7) are in Triality. They correspond to the integrated transformations T_1, T_2, T_3 of Eq. (2.24). We now turn to the infinitesimal form of the Triality relations for R_1, R_2, and R_3 of Eq. (2.25):

$(\Sigma^7_R,\Sigma^7_L,\Sigma^7)$Triality: Let $\Omega = K_\gamma \in \Sigma^7 =$ Spin(7)/G_2. Then

$$A = \tfrac{1}{2}(K_\gamma - K^{\flat}_\gamma) = \tfrac{1}{2}(K_\gamma + 3M_\gamma - 2K_\gamma) = \Sigma^R_\gamma \in \Sigma^7_R, \tag{5.43a}$$
$$B = \tfrac{1}{2}(K_\gamma - K^{\#}_\gamma) = \tfrac{1}{2}(K_\gamma - 3M_\gamma - 2K_\gamma) = \Sigma^L_\gamma \in \Sigma^7_L, \tag{5.43b}$$
$$C = -K_\gamma + K'_\gamma = -K_\gamma + 2K_\gamma = K_\gamma \in \Sigma^7, \tag{5.43c}$$

where use has been made of Eqs. (5.27), (5.31), and (5.34).

Alternatively, we could have taken $\Omega = \Sigma^R_\gamma$ and obtain

$$A = K_\gamma, \quad B = \Sigma^R_\gamma, \quad C = \Sigma^L_\gamma. \tag{5.44}$$

Finally the choice $\Omega = \Sigma^L_\gamma$ yields

$$A = \Sigma^L_\gamma, \quad B = K_\gamma, \quad C = \Sigma^R_\gamma. \tag{5.45}$$

In all cases we obtain a permutation of the original (A,B,C) operators of Eq. (5.43), showing the Triality of the transformations $\frac{1}{2}(\pm 3M_\gamma - K_\gamma)$ and K_γ associated with the integral operators R_i of Eq. (2.25).

Since we let Ω range over all elements of the Lie algebra Spin(8), we have covered all cases of Triality and have shown that A,B,C as defined by Eq. (5.37) correspond to three transformations of Spin(8) in Triality. They are, respectively, the anti-self sharp coupled, flat coupled, and mixed coupled elements of the Spin(8) Lie algebra.

Finally we remark that Spin(8)/G_2 can have as basis either $M_\gamma \pm K_\gamma$ or $\pm 3M_\gamma - K_\gamma$. The latter corresponds to Spin(8) $= \Sigma^7_R \times \Sigma^7_L \times G_2$ and is a G_2 invariant decomposition.

VI. CONCLUDING REMARKS

In this work we have explicated the rich G_2 invariant algebraic structure underlying the existence of various exceptional vector cross products in eight dimensions. We have exhibited the intricate weaving of the notion of cross products, hence of generalized almost complex structure[18] and duality. In four dimensions the deep connection between complex, quaternionic structure, and duality[15,30] has enriched particle theory with fundamental structural insights and novel nonperturbative physical mechanisms in gauge and gravity theories. So a natural question is the following: Does there exist an eight-dimensional counterpart of this connection between the exceptional duality given here and an octonionic homomorphic structure? Along with the other works,[12,10,32] our own octonionic analysis[11] of Englert's solutions and recent advances on Clifford analysis[33,25] suggests an affirmative answer to the above query. So besides the applications mentioned in the Introduction, we envision that the rich octonionic symmetry laid bare here may well embody the analytic solutions to some interesting $d = 7$ or 8 differential equations governing the physics of Kaluza–Klein theories.

ACKNOWLEDGMENT

The research was supported in part by the U.S. Department of Energy under Contract No. DE-AC022-76ERO3075.

APPENDIX

For completeness we give here the Triality structure in the commutation relations of SO(8) as reduced with respect to G_2.

We have already decomposed O(8) with respect to the basis $M_\alpha \in$ SO(8)/SO(7) and $J_{\alpha\beta} \in$ SO(7) viz. Eqs. (5.6) and (5.3). We have further decomposed SO(7) into SO(7)/G_2 and G_2 represented, respectively, by the generators K_γ and $\Gamma_{\alpha\beta}$ with the constraint being $\psi_{\alpha\beta\gamma}\Gamma_{\alpha\beta} = 0$ [see Eqs. (5.10)–(5.12)]. Thus we find in the basis M_γ, K_γ, $\Gamma_{\alpha\beta}$ for O(8):

$$[M_\alpha,M_\beta] = \tfrac{1}{3}(\psi_{\alpha\beta\gamma}K_\gamma + \Gamma_{\alpha\beta}), \tag{A1}$$
$$[K_\alpha,K_\beta] = -\psi_{\alpha\beta\gamma}K_\gamma + \Gamma_{\alpha\beta}, \tag{A2}$$

$$[M_\alpha, K_\beta] = [K_\alpha, M_\beta] = \psi_{\alpha\beta\gamma} M_\gamma \tag{A3}$$

to which we add the commutation relation of $\Gamma_{\alpha\beta}$ given by Eq. (4.66),

$$[\Gamma_{\alpha\beta}, \Gamma_{\rho\sigma}] = C_{\alpha\beta\rho\sigma\kappa\lambda} \Gamma_{\kappa\lambda}. \tag{A4}$$

We have seen that $S^R_\mu = \frac{1}{2}(M_\mu + K_\mu)$, $S^L_\mu = \frac{1}{2}(M_\mu - K_\mu)$, and $-M_\mu$ form a triple of operators in Triality. We obtain a form for the O(8) commutation relations with Triality symmetry if we set

$$A_\mu = S^R_\mu,\ B_\mu = S^L_\mu,\ C_\mu = -M_\mu\ \ (A_\mu + B_\mu + C_\mu = 0). \tag{A5}$$

Further we introduce the alternative triple of operators in Triality given by

$$\begin{aligned} U_\mu &= \Sigma^R_\mu = \tfrac{1}{2}(3M_\mu - K_\mu), \\ V_\mu &= \Sigma^L_\mu = -\tfrac{1}{2}(3M_\mu + K_\mu), \\ W_\mu &= K_\mu \quad (U_\mu + V_\mu + W_\mu = 0). \end{aligned} \tag{A6}$$

We have the following relations:

$$U_\mu = B_\mu - C_\mu,\quad V_\mu = C_\mu - A_\mu,\quad W_\mu = A_\mu - B_\mu, \tag{A7}$$

$$\begin{aligned} A_\mu &= -\tfrac{1}{3}(V_\mu - W_\mu),\quad B_\mu = -\tfrac{1}{3}(W_\mu - U_\mu), \\ C_\mu &= -\tfrac{1}{3}(U_\mu - V_\mu). \end{aligned} \tag{A8}$$

In the (U,V,W,Γ) basis, the O(8) commutation relations take the highly symmetrical form [with the addition of the G_2 commutator Eq. (A4)]:

$$[V_\alpha, W_\beta] = -\psi_{\alpha\beta\gamma} U_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}, \tag{A9}$$

$$[W_\alpha, U_\beta] = -\psi_{\alpha\beta\gamma} V_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}, \tag{A10}$$

$$[U_\alpha, V_\beta] = -\psi_{\alpha\beta\gamma} W_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}, \tag{A11}$$

$$[U_\alpha, U_\beta] = -\psi_{\alpha\beta\gamma} U_\gamma + \Gamma_{\alpha\beta}, \tag{A12}$$

$$[V_\alpha, V_\beta] = -\psi_{\alpha\beta\gamma} V_\gamma + \Gamma_{\alpha\beta}, \tag{A13}$$

$$[W_\alpha, W_\beta] = -\psi_{\alpha\beta\gamma} W_\gamma + \Gamma_{\alpha\beta}. \tag{A14}$$

In the (A,B,C,Γ) basis, together with Eq. (A4) we obtain

$$[B_\alpha, C_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(B_\gamma - C_\gamma) - \tfrac{1}{6}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} U_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}), \tag{A15}$$

$$[C_\alpha, A_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(C_\gamma - A_\gamma) - \tfrac{1}{6}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} V_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}), \tag{A16}$$

$$[A_\alpha, B_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(A_\gamma - B_\gamma) - \tfrac{1}{6}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} W_\gamma - \tfrac{1}{2}\Gamma_{\alpha\beta}), \tag{A17}$$

$$[A_\alpha, A_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(B_\gamma - C_\gamma) + \tfrac{1}{3}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} U_\gamma + \Gamma_{\alpha\beta}), \tag{A18}$$

$$[B_\alpha, B_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(C_\gamma - A_\gamma) + \tfrac{1}{3}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} V_\gamma + \Gamma_{\alpha\beta}), \tag{A19}$$

$$[C_\alpha, C_\beta] = \tfrac{1}{3}\psi_{\alpha\beta\gamma}(A_\gamma - B_\gamma) + \tfrac{1}{3}\Gamma_{\alpha\beta} = \tfrac{1}{3}(\psi_{\alpha\beta\gamma} W_\gamma + \Gamma_{\alpha\beta}), \tag{A20}$$

$$A_\alpha + B_\alpha + C_\alpha = 0. \tag{A21}$$

In both the (U,V,W,Γ) and (A,B,C,Γ) bases, the Triality symmetry of the O(8) commutation relations is manifest.

[1] M. J. Duff, "Supergravity, the seven-sphere and spontaneous symmetry breaking," CERN Preprint TH. 3451, 1982.
[2] E. Cremmer and B. Julia, Nucl. Phys. B **159** 141 (1979).
[3] E. Witten, Nucl. Phys. B **187**, 412 (1981).
[4] See for instance, J. Milnor, Ann. Math. **64**, 399 (1956).
[5] T. Kugo and P. Townsend, Nucl. Phys. B **221**, 357 (1983) and references quoted therein.
[6] H.Freudenthal, "Oktaven, Ausnahmengruppen und Oktavengeometrie," Math. Inst., Rigi Univ.te Utrecht (1951); J. R. Faulkner and J. C. Farrar, Bull. London Math. Soc. **9**, 1 (1977).
[7] There have been numerous papers on this connection. For a comprehensive reference list, see L. Sorgsepp and J. Lohmus, "About Nonassociativity in Physics and Cayley–Graves' Octonions" Tartu Preprint F-7. 1978, published in Harvard Workshop 1979, p. 1388; Hadronic J. **2**, (1979) and references therein.
[8] F. Gürsey, lectures delivered at the 1977 International School of Subnuclear Physics, Ettore Majorana.
[9] F. Englert, Phys. Lett. B **119**, 339 (1982).
[10] R. D'Auria, P. Fré, and P. van Nieuwenhuizen, Phys. Lett. B **122**, 225 (1983); M. J. Duff and C. N. Pope, ICTP preprint 83–7; F. Englert, M. Rooman, and P. Spindel, Univ. Libre Bruxelles preprint, 1983; M. J. Duff, B. E. W. Nilsson, and C. N. Pope, Univ. Texas at Austin preprint, 1983.
[11] F. Gürsey and C.-H. Tze, Phys. Lett. B **127**, 191 (1983).
[12] Independently some of our results have also been obtained by T. Dereli, M. Panahimoghaddam, A. Sudbery, and R. W. Tucker, Lancaster Univ. preprint, 1983; J. Lukierski and P. Minnaert, Univ. Bordeaux Preprint PTB-128 1983; see, in particular, M. Rooman, Univ. Libre Bruxelles preprint, 1983.
[13] E. Cartan and J. A. Schouten, Proc. K. Ned. Akad. Wet. Amsterdam, **29**, 933 (1926).
[14] R. D. Schafer, *An Introduction to Non-Associative Algebras* (Academic, New York, 1966); V. D. Blij and T. A. Springer, Nieuw Archief voor Wiskunde (3), **VIII**, 158 (1960).
[15] M. F. Atiyah, N. Hitchin, and I. Singer, Proc. Roy. Soc. London A **362** 425 (1978); R. S. Ward, Commun. Math. Phys. **78**, 1 (1980).
[16] R. Hermann, Ed., *Spinors, Clifford and Cayley Algebras, Interdisciplinary Mathematics* (Math Sci Press, Brookline, MA, 1973), Vol. VII, Chap. 5.
[17] E. Calabi, Trans. Am. Math. Soc. **87**, 40 (1958). For a recent paper on S^6 in Kaluza–Klein theory, see S. Randbar-Daemi, A. Salam, and J. Strathdee, ICTP Preprint IC/83/7.
[18] A. Gray, Trans. Am. Math. Soc. **41**, 465 (1969); R. Brown and A. Gray, Comment. Math. Helv. **42**, 222 (1967).
[19] M. Berger, Bull. Soc. Math. France **83**, 279 (1955).
[20] E. Bonan, C. R. Acad. Sci. **262**, 127 (1966); D. V. Alekseevskii, Funktsional'nyi Analiz i Ego Prilozheniya **2**, 1 (1968).
[21] P. Zvengrowski, *Vector Fields and Vector Products*, Chap. 4, thesis, University of Chicago, 1965.
[22] E. Kleinfeld, in *Studies in Modern Algebra 2*, edited by A. A. Albert (Prentice–Hall, Englewood Cliffs, NJ, 1963), p. 1.
[23] V. Ogievetski and V. Tzeitlin, J. Phys. A (1979).
[24] L. Alvarez-Gaume and D. Z. Freedman, Commun. Math. Phys. **80**, 443 (1983); J. Bagger and E. Witten, "Matter Couplings in $N = 2$ Supergravity," Princeton Preprint, 1983, and references quoted therein.
[25] M. F. Atiyah, R. Bott, and A. Shapiro, Clifford Modules, Topology **3**, 1964 (1964); D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[26] M. Günaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973).
[27] I. Yokota, J. Fac. Sci. Shinshu Univ. **3**, (no. 1) 35 (1968); H. Toda, Y. Saito, and I. Yokota, Mem. Coll. Sci. Univ. Kyoto Ser. A **30**, 227 (1957).
[28] B. Eckmann, Comment Math. Helv. **15**, 358 (1942); G. W. Whitehead, Comment. Math. Helv. **37**, 1651 (1963).
[29] P. Hilton, *General Cohomology Theory and K-Theory* (Cambridge University, Cambridge, England, 1971).
[30] F. Gürsey and C.-H. Tze, Ann. Phys. **128**, 29 (1980).
[31] J. E. d'Atri and N. K. Nickerson, J. Diff. Geom. **2**, 393 (1968).
[32] L.Castellani and N. P. Warner, Caltech Preprint CALT-68-1033, May 1983.
[33] F. Brackx, R. Delanghe, and F. Sommen, *Clifford Analysis* (Pitman, New York, 1982).

On the relation between classical and quantum-thermodynamic entropy

Gian Paolo Beretta
Massachusetts Institute of Technology, Room 3-339, Cambridge, Massachusetts 02139

(Received 31 March 1983; accepted for publication 20 May 1983)

We discuss the unresolved problem of proving rigorously that in the classical limit $\hbar \to 0$, the quantum-thermodynamic entropy functional tends to the classical entropy functional. We state rather restrictive conditions that define the general problem of finding a complete classical phase-space representation of quantum kinematics. Whether the problem admits of solutions remains an unresolved question. We discuss a physically interesting attempt to relate the structure in the classical limit $\hbar \to 0$ of the well-known Blokhintzev, Wigner, and Wehrl phase-space functions to the spectral expansion of the quantum state operator.

PACS numbers: 05.30. − d, 02.50. + s

I. INTRODUCTION: PHYSICAL CONTEXT OF THE PROBLEM

The purpose of this paper is to discuss the problem of proving that in the classical limit $\hbar \to 0$ the quantum-thermodynamic entropy functional

$$s(\rho) = -k \operatorname{Tr} \rho \ln \rho \tag{1}$$

tends to the classical entropy functional

$$s^{\mathrm{cl}}(w) = -k \int \frac{dq\, dp}{2\pi\hbar} w(q,p) \ln w(q,p), \tag{2}$$

where ρ is the quantum state operator and w a properly defined classical state function. As recently stated by Wehrl,[1] "It is usually claimed that in the limit $\hbar \to 0$, the quantum-mechanical expression tends towards the classical one, however, a rigorous proof of this is nowhere found in the literature."

In this paper, we state conditions that define the problem of finding a classical phase-space representation of quantum kinematics. One such condition is that a properly defined classical state function w (the definition of which may involve limits as $\hbar \to 0$) should be such that $s^{\mathrm{cl}}(w) = s(\rho)$. Whether the problem thus defined admits of solutions remains an unresolved question which is worth further investigations. We gain some insight in the problem by studying the phase-space structure induced in the classical limit by the spectral expansion of the quantum state operator.

Our interest in this problem arises for physical reasons essentially distinct from the traditional. Indeed, even the physical meaning that we assign to the mathematical objects, especially ρ and w, is entirely different from the conventional. In our attempts to unify the laws of quantum mechanics and thermodynamics into a quantum thermodynamics,[2,3] an underlying hypothesis has been that no layer of statistical or information-theoretic reasoning should be required to bridge the gap between mechanics and thermodynamics. Indeed, in our theory there is no such gap. Quantum thermodynamics is a nonstatistical theory concerned exclusively with a causal description of the individual quantum states of a system, including a single particle.

A most important fundamental hypothesis[2] of quantum thermodynamics is that the general mathematical representation of the individual quantum states of a single isolated system cannot be in terms of the traditional state vectors $|\psi\rangle$ in the Hilbert space $\mathscr{H}$ of the system, but must be in terms of self-adjoint, nonnegative-definite, unit-trace operators ρ on $\mathscr{H}$ that are not necessarily idempotent. State operators ρ have the same mathematical properties ($\rho^\dagger = \rho$, $\rho \geqslant 0$, $\operatorname{Tr} \rho = 1$) as the statistical or density operators considered in traditional (von Neumann) quantum statistical mechanics.[4] But in quantum thermodynamics, their physical meaning is entirely different. The operators ρ represent individual states of a single system and not the index of statistics from a generally heterogeneous ensemble of identical systems. Thus, for example, the entropy functional $s(\rho)$, defined by Eq. (1), represents the entropy[5] of the single system in any of its states, equilibrium and nonequilibrium, dissipative and nondissipative,[3] and not a measure of statistical or information-theoretic uncertainty.

With this background, the problem of studying the classical limit $\hbar \to 0$ acquires for us a special physical meaning. The phase-space functions $w(q,p)$ represent individual classical states of a single system and not the index of statistics from a Gibbsian ensemble. The functional $s^{\mathrm{cl}}(w)$ represents the individual classical entropy of the single system in any of its classical states, and not the Gibbsian index of probability of phase.

However, we wish to emphasize that all the mathematical observations reported in this paper may obviously be interpreted also in the traditional way.

We restrict our discussion to the case of a single degree of freedom (e.g., a one-dimensional harmonic oscillator) so that the Hilbert space is $\mathscr{H} = L^2(\mathbb{R})$ and the classical phase space is $\Omega = \mathbb{R}^2$.

The paper is organized as follows. Coherent states and three well-known phase-space maps are briefly reviewed in Sec. II, conditions defining a complete phase-space representation of quantum kinematics are given in Sec. III, a discussion on the classical limit of the three phase-space maps is given in Sec. IV, and conclusions in Sec. V.

II. COHERENT STATES AND PHASE-SPACE FUNCTIONS

We denote the position and momentum operators by Q and P ($[Q,P] = i\hbar I$), and their eigenvalues by q and p so that $Q|q\rangle = q|q\rangle$ and $P|p\rangle = p|p\rangle$. The annihilation and creation

 0022-2488/84/051507-04$02.50

operators Z and $Z^\dagger$ ($[Z, Z^\dagger] = I$) are then defined by the relations[6]

$$Z \equiv \sqrt{1/2\hbar\omega}(\omega Q + iP), \quad Q = \sqrt{\hbar/2\omega}(Z^\dagger + Z), \tag{3}$$

$$Z^\dagger \equiv \sqrt{1/2\hbar\omega}(\omega Q - iP), \quad P = \sqrt{\omega\hbar/2}i(Z^\dagger - Z).$$

We call the normalized eigenvectors $|z\rangle$ of Z ($Z|z\rangle = z|z\rangle$) coherent vectors with "natural frequency" ω. For each eigenvalue z of Z, we define the real variables x and y (with dimensions of position and momentum, respectively) by the relations

$$x \equiv \sqrt{\hbar/2\omega}(z^* - z), \quad z = \sqrt{1/2\hbar\omega}(\omega x + iy), \tag{4}$$

$$y \equiv \sqrt{\omega\hbar/2}i(z^* + z), \quad z^* = \sqrt{1/2\hbar\omega}(\omega x - iy).$$

The representation of coherent vector $|z\rangle$ in configuration space is

$$\langle q|z\rangle = (\omega/\pi\hbar)^{1/4} \times \exp[-\omega(x-q)^2/2\hbar + iyq/\hbar - ixy/2\hbar], \tag{5}$$

whereas in momentum space it is

$$\langle p|z\rangle = (1/\pi\hbar\omega)^{1/4} \times \exp[-(y-p)^2/2\hbar\omega - ixp/\hbar + ixy/2\hbar]. \tag{6}$$

With $z = 0$ ($x = 0, y = 0$), Eqs. (5) and (6) give the representations of coherent vector $|0\rangle$. Moreover, we have the well-known relation $|z\rangle = W(z)|0\rangle$, where $W(z)$ is the unitary (Weyl) operator

$$W(z) \equiv \exp(zZ^\dagger - z^*Z) = W(x, y) = \exp[i(yQ - xP)/\hbar]. \tag{7}$$

We say that a system is in a coherent state if and only if its (individual) state operator is

$$\rho = P_z \equiv |z\rangle\langle z| = \rho^2. \tag{8}$$

These pure states are minimum uncertainty in phase space ($\Delta Q \Delta P = \hbar/2$).

Several linear mappings from the set of self-adjoint operators A on $\mathscr{H}$ to the set of complex-valued functions on the classical phase space Ω have been considered in the literature.[7] In our physical context, these mappings are attempts to find a phase-space representation of an individual quantum system. We will consider only three important examples, namely, the Blokhintzev phase-space map[8]

$$r(q, p; A) \equiv 2\pi\hbar\langle q|A|p\rangle\langle p|q\rangle \tag{9a}$$

$$= \iint \frac{d\theta\, d\tau}{2\pi\hbar} e^{-i(\theta q + \tau p)/\hbar} \mathrm{Tr}(e^{i\theta Q/\hbar} A e^{i\tau P/\hbar}), \tag{9b}$$

the Wigner phase-space map[9]

$$g(q, p; A) \equiv \iint \frac{d\theta\, d\tau}{2\pi\hbar} e^{-i(\theta q + \tau p)/\hbar} \mathrm{Tr}(A e^{i(\theta Q + \tau P)/\hbar}) \tag{10a}$$

$$= \iint \frac{d\xi\, d\eta}{\pi\hbar} r(q + \xi, p + \eta; A) e^{2i\xi\eta/\hbar}, \tag{10b}$$

and the Wehrl phase-space map[1]

$$f(q, p; A) \equiv \langle q, p|A|q, p\rangle \tag{11a}$$

$$= \iint \frac{d\xi\, d\eta}{\sqrt{2}\pi\hbar} r(q + \xi, p + \eta; A) e^{i\xi\eta/\hbar} e^{-(\omega^2\xi^2 + \eta^2)/2\hbar\omega}, \tag{11b}$$

where $|q, p\rangle$ is the coherent vector $|z\rangle$ with $x = q$ and $y = p$. The Blokhintzev and the Wigner maps are not real (i.e., $r^* \neq r$ and $g^* \neq g$), while the Wehrl map is real and nonnegative (i.e., $f^* = f \geqslant 0$). For every ρ and A, we have

$$\iint \frac{dq\, dp}{2\pi\hbar} r(q, p; \rho) = \iint \frac{dq\, dp}{2\pi\hbar} g(q, p; \rho) = \iint \frac{dq\, dp}{2\pi\hbar} f(q, p; \rho) = \mathrm{Tr}\,\rho = 1 \tag{12}$$

and

$$\iint \frac{dq\, dp}{2\pi\hbar} r^*(q, p; A) r(q, p; \rho) = \iint \frac{dq\, dp}{2\pi\hbar} g^*(q, p; A) g(q, p; \rho) = \mathrm{Tr}\, A\rho. \tag{13}$$

For the Wehrl map, instead, the relation

$$\iint \frac{dq\, dp}{2\pi\hbar} f(q, p; A) f(q, p; \rho) = \mathrm{Tr}\, A\rho \tag{14}$$

holds for every A only for the special class of (Wehrl) states ρ for which

$$\rho = \iint \frac{dq\, dp}{2\pi\hbar} f(q, p; \rho) |q, p\rangle\langle q, p|. \tag{15}$$

The usual interpretation of these relations is that the maps $r(q, p; \rho)$, $g(q, p; \rho)$, and $f(q, p; \rho)$ play a role analog to that of the classical phase-space state function, and the maps $r^*(q, p; A)$, $g^*(q, p; A)$, and $f(q, p; A)$ a role analog to that of the classical phase-space function associated with observable A. A discussion of the time evolution of $r(q, p; \rho)$ under Hamiltonian dynamics is given in Ref. 8.

III. CLASSICAL REPRESENTATION OF QUANTUM KINEMATICS

Ideally, a classical phase-space representation of a quantum system would be complete if it were possible to solve the following general problem. Given a system with quantum-mechanical Hilbert space $\mathscr{H}$ and classical-mechanical phase space Ω, find two mappings $w(q, p; \rho)$ and $a(q, p; A)$ that satisfy the following rather restrictive conditions. For every state operator ρ on $\mathscr{H}$, every well-defined self-adjoint operator A on $\mathscr{H}$, every point q, p in Ω, and every continuous real function $F(t)$ of the real variable t,

(i) $w(q, p; \rho)$ is real and nonnegative,

(ii) $a(q, p; A)$ is real,

(iii) $$\iint \frac{dq\, dp}{2\pi\hbar} F(w(q, p; \rho)) = \mathrm{Tr}\, F(\rho),$$

(iv) $$\iint \frac{dq\, dp}{2\pi\hbar} F(a(q, p; A)) w(q, p; \rho) = \mathrm{Tr}\, F(A)\rho.$$

Clearly, for $F(t) = -kt \ln t$ if $0 < t \leqslant 1$ and $F(t) = 0$ elsewhere, condition (iii) implies that $s^{cl}(w) = s(\rho)$.

No rigorous solution to this problem is known. To the best of our knowledge, even the physically interesting ques-

tion whether the problem admits of solutions, let alone to find them, remains unresolved. In what follows, we discuss the rudiments of an approach that may provide useful insight towards a resolution of the question.

IV. CLASSICAL LIMIT OF PHASE-SPACE FUNCTIONS

Let us consider the spectral expansion of the state operator

$$\rho = \sum_j w_j P_j, \tag{16}$$

$$I = \sum_j P_j, \tag{17}$$

$$P_j P_k = \delta_{jk} P_j, \tag{18}$$

where I is the identity and P_j the projector onto the eigenspace belonging to eigenvalue w_j, with degeneracy $d_j = \mathrm{Tr}\, P_j$. We readily verify that

$$\iint \frac{dq\,dp}{2\pi\hbar} r(q,p;P_j) = \mathrm{Tr}\, P_j = d_j, \tag{19}$$

$$\sum_j r(q,p;P_j) = 1, \tag{20}$$

$$r(q,p;\rho) = \sum_j w_j r(q,p;P_j), \tag{21}$$

and similar relations hold for the mappings g and f.

Using the relations

$$\iint \frac{d\xi\,d\eta}{2\pi\hbar} \xi^n \eta^m e^{i\xi\eta/\hbar} = \delta_{nm} i^n \hbar^n n!, \tag{22}$$

$$\iint \frac{d\xi\,d\eta}{\sqrt{2}\pi\hbar} \xi^n \eta^m e^{i\xi\eta/\hbar} e^{-(\omega^2\xi^2+\eta^2)/2\hbar\omega} = \begin{cases} 0 & \text{if } n+m \text{ odd}, \\ \dfrac{1\cdot 3\cdots(n+m-1)}{2^{n+m}} (\hbar\omega)^{m/2} (\hbar/\omega)^{n/2} & \text{if } n+m \text{ even}, \end{cases} \tag{23}$$

and the expansion

$$r(q+\xi, p+\eta; A) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{\partial^{n+m} r}{\partial q^n\, \partial p^m}\bigg|_{q,p;A} \frac{\xi^n \eta^m}{n!m!} \tag{24}$$

into Eqs. (10b) and (11b), we find

$$g(q,p;A) = r(q,p;A) + \sum_{n=1}^{\infty} \frac{\partial^{2n} r}{\partial q^n\, \partial p^n}\bigg|_{q,p;A} \frac{i^n \hbar^n}{n!}, \tag{25}$$

$$f(q,p;A) = r(q,p;A) + \sum_{n=1}^{\infty} \sum_{k=0}^{2n} \frac{\partial^{2n} r}{\partial q^k \partial p^{2n-k}}\bigg|_{q,p;A} \frac{1\cdot 3\cdots(2n-1)}{4^n} \omega^{n-k} \hbar^n, \tag{26}$$

and, after some manipulations involving Relations (18) and (22),

$$r(q,p;P_j) r(q,p;P_k) = \delta_{jk} r(q,p;P_j) - \sum_{n=1}^{\infty} \frac{(-i)^n \hbar^n}{n!} \frac{\partial^n r}{\partial q^n}\bigg|_{q,p;P_j} \frac{\partial^n r}{\partial p^n}\bigg|_{q,p;P_k}. \tag{27}$$

Relations (20) and (27) imply that the functions $r(q,p;P_j)$ have overlapping supports which cover the whole phase space Ω, i.e.,

$$\cup_j \Omega_j = \Omega, \tag{28}$$

$$\Omega_j \cap \Omega_k \neq \delta_{jk} \Omega_j, \tag{29}$$

where

$$\Omega_j \equiv \{q,p \mid r(q,p;P_j) \neq 0\}. \tag{30}$$

However, the relative importance of the overlap is small. For example, $|\int r_j r_k|^2 / |\int r_j^2| |\int r_k^2|$ is of order $\hbar^2/d_j d_k$ for $j \neq k$. Thus we conclude that, in the classical limit $\hbar \to 0$, the spectral expansion of the state operator ρ induces a partition of the phase-space Ω into disjoint cells Ω_j each belonging to a distinct eigenvalue w_j of ρ.

From Eqs. (25) and (26), it follows that if the Blokhintzev map $r(q,p;A)$, the Wigner map $g(q,p;A)$, and the Wehrl map $f(q,p;A)$ each admit of a classical limit as $\hbar \to 0$, then they tend to the same map. Assuming that such limits exist, we introduce the following notation

$$w(q,p) \equiv \lim_{\hbar\to 0} r(q,p;\rho) = \lim_{\hbar\to 0} g(q,p;\rho) = \lim_{\hbar\to 0} f(q,p;\rho), \tag{31}$$

$$a(q,p) \equiv \lim_{\hbar\to 0} r^*(q,p;A) = \lim_{\hbar\to 0} g^*(q,p;A) = \lim_{\hbar\to 0} f(q,p;A), \tag{32}$$

$$\alpha_j(q,p) \equiv \lim_{\hbar\to 0} r(q,p;P_j) = \lim_{\hbar\to 0} g(q,p;P_j) = \lim_{\hbar\to 0} f(q,p;P_j). \tag{33}$$

It then follows from Relations (20), (21), and (27) that

$$\sum_j \alpha_j(q,p) = 1, \tag{34}$$

$$w(q,p) = \sum_j w_j \alpha_j(q,p), \tag{35}$$

$$\alpha_j(q,p)\alpha_k(q,p) = \delta_{jk} \alpha_j(q,p), \tag{36}$$

and, therefore, the functions $\alpha_j(q,p)$ can only take the values 0 and 1, and have nonoverlapping supports covering the whole phase space. Thus the structure of the function w is such that

$$F(w(q,p)) = \sum_j F(w_j) \alpha_j(q,p) \tag{37}$$

in analogy with the general relation

$$F(\rho) = \sum_j F(w_j) P_j. \tag{38}$$

To proceed further, we conjecture that

$$\iint \frac{dq\,dp}{2\pi\hbar} \alpha_j(q,p) = d_j. \tag{39}$$

We have no proof for this. The conjecture is based only on some heuristic arguments. We first note that Relation (39) is consistent with the improper formal relations

$$\iint \frac{dq\,dp}{2\pi\hbar} 1 = \iint \frac{dq\,dp}{2\pi\hbar} r(q,p;I) = \mathrm{Tr}\, I = \sum_j d_j, \tag{40}$$

$$\iint \frac{dq\,dp}{2\pi\hbar}\,1 = \iint \frac{dq\,dp}{2\pi\hbar}\sum_j \alpha_j(q,p)$$

$$= \sum_j \iint \frac{dq\,dp}{2\pi\hbar}\,\alpha_j(q,p), \tag{41}$$

where we used that fact that $r(q,p;I)=1$ and Relation (34). Moreover, we note that it is consistent with the requirement that if $w(q,p)$ is to represent a classical state function, then it must be normalized and, therefore,

$$\iint \frac{dq\,dp}{2\pi\hbar}\,w(q,p) = \sum_j w_j \iint \frac{dq\,dp}{2\pi\hbar}\,\alpha_j(q,p)$$

$$= 1 = \sum_j w_j d_j. \tag{42}$$

Finally, we observe that the conjecture and the normalization condition for $w(q,p)$ would follow if there were a meaning to saying that the phase-space measure $dq\,dp/2\pi\hbar$ is independent of $\hbar$ so that the following relation would make sense at least for A in the trace class

$$\mathrm{Tr}\,A = \iint \frac{dq\,dp}{2\pi\hbar}\,r(q,p;A) = \lim_{\hbar\to 0}\iint \frac{dq\,dp}{2\pi\hbar}\,r(q,p;A)$$

$$= \iint \frac{dq\,dp}{2\pi\hbar}\lim_{\hbar\to 0} r(q,p;A). \tag{43}$$

If the conjecture could be proved, then from Relations (37) and (38) it would follow that

$$\iint \frac{dq\,dp}{2\pi\hbar}\,F(w(q,p)) = \iint \frac{dq\,dp}{2\pi\hbar}\sum_j F(w_j)\alpha_j(q,p)$$

$$= \sum_j F(w_j)d_j = \mathrm{Tr}\,F(\rho), \tag{44}$$

which would prove that the function $w(q,p)$ [Relation (31)] satisfies conditions (i) and (iii) of Sec. III and that, in particular, $s^{\mathrm{cl}}(w) = s(\rho)$. In a similar manner, and with similar conjectures, we would show that the function $a(q,p)$ [Relation (32)] satisfies conditions (ii) and (iv).

Because it is not clear whether Relations (39)–(43) admit of a rigorous justification, we conclude that the question of existence of solutions to the problem defined in Sec. III remains unresolved.

V. CONCLUSIONS

We have given restrictive conditions defining a complete classical phase-space representation of quantum kinematics for systems with both a classical and a quantum description. Whether such representations exist is an unresolved problem. We presented heuristic arguments in support of the usual unproved claim that $s(\rho)\to s^{\mathrm{cl}}(w)$ in the classical limit $\hbar\to 0$.

We have observed that, in the limit $\hbar\to 0$, the spectral expansion of the quantum state operator ρ induces a partition of the phase space Ω into disjoint cells Ω_j each belonging to a distinct eigenvalue w_j of ρ. Over cell Ω_j, the classical state function w has the constant value w_j. We conjectured that the phase-space volume of cell Ω_j equals the degeneracy d_j of eigenvalue w_j. Accordingly, the phase-space volume of the support of w, i.e., of the complement of cell Ω_0 belonging to the zero eigenvalue of ρ, cannot be smaller than the value 1 attained for every idempotent or pure state ($\rho^2=\rho$).

We conclude with a remark on dynamics, namely, on the distinction between conservation of volume in phase space and thermodynamic reversibility. It is true that the Liouville–von Neumann equation for the unitary evolution of ρ induces in the classical limit an evolution of w which preserves both the volume of the support of w and the value of the entropy. However, it is seldom recognized explicitly that conservation of volume in phase space is not equivalent to thermodynamic reversibility. For example, we could conceive of nonunitary evolution equations which preserve the volume of the support of w (i.e., equations that preserve the zero eigenvalues of ρ) but not the value of the entropy (i.e., such that the nonzero eigenvalues of ρ are not invariant). We believe[3] this to be an interesting and physically important feature of a recently proposed[10] nonlinear quantum equation of motion for a single elementary constituent of matter.

ACKNOWLEDGMENT

The author gratefully acknowledges financial support of the Thermo Electron Corporation.

[1] A. Wehrl, Rep. Math. Phys. **16**, 353 (1979). See also A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978) and E. H. Lieb, Commun. Math. Phys. **62**, 35 (1978).

[2] G. N. Hatsopoulos and E. P. Gyftopoulos, Found. Phys. **6**, 15, 127, 439, 561 (1976).

[3] G. P. Beretta, E. P.Gyftopoulos, J. L. Park, and G. N. Hatsopoulos (in press).

[4] A most instructive essay on the conceptual problems arising from the traditional use of the statistical operator is found in J. L. Park, Am. J. Phys. **36**, 211 (1968).

[5] A most persuasive proof that the entropy of any system (even a single particle) in any state ρ (equilibrium and nonequilibrium) is $s(\rho) = -k\,\mathrm{Tr}\,\rho\ln\rho$ is given in Ref. 2.

[6] See, e.g., J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), Chap. 7.

[7] L. Cohen, J. Math. Phys. **7**, 781 (1966); L. Cohen, in *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, edited by C. A. Hooker (Dordrecht, Holland, 1973), p. 66; J. L. Park, W. Band, and W. Yourgrau, Ann. Phys. **7.37**, 189 (1980) and references therein.

[8] D. I. Blokhintzev, J. Phys. **2**, 71 (1940); D. I. Blokhintzev and P. Nemirovsky, J. Phys. **3**, 291 (1940).

[9] E. Wigner, Phys. Rev. **40**, 749 (1932).

[10] G. P. Beretta, Thesis, (MIT, 1981) (unpublished). The new equation for a single elementary constituent of matter is presented in Ref. 3.

Linear transport in nonhomogeneous media. III

R. Daniel and G. C. Pomraning
School of Engineering and Applied Science, University of California, Los Angeles, California 90024

Edward W. Larsen
University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87544

(Received 26 July 1983; accepted for publication 16 December 1983)

A set of continuum and discrete eigenfunctions is constructed for linear half-space transport problems with single scatter albedos which are bilinear functions of position. For Hölder continuous distributions incident upon the half-space, these eigenfunctions are proved to be half-range complete and the expansion coefficient of the continuum eigenfunctions is proved to be continuous.

PACS numbers: 05.60. + w, 95.30.Jx, 42.68.Db

I. INTRODUCTION

In this article we examine some theoretical questions concerning the half-range completeness of a set of eigenfunctions arising from the half-space linear transport problem described by

$$\mu \frac{\partial \psi(x,\mu)}{\partial x} + \psi(x,\mu) = \frac{c(x)}{2} \int_{-1}^{1} \psi(x,\mu')d\mu', \quad x>0, \tag{1.1}$$

$$\psi(0,\mu) = g(\mu), \quad 0 \leqslant \mu \leqslant 1, \tag{1.2}$$

$$|\psi(x,\mu)| \leqslant M_0 \equiv \sup_{0<\mu<1} g(\mu), \quad 0 \leqslant x \leqslant \infty, \quad -1 \leqslant \mu \leqslant 1. \tag{1.3}$$

We assume that the function $g(\mu)$ is nonnegative and Hölder continuous for $0 \leqslant \mu \leqslant 1$, and that the single scatter albedo $c(x)$ is given by

$$c(x) = c_2 + \int_0^{\infty} \alpha(t) e^{-x/t}\, dt, \tag{1.4}$$

with

$$\alpha(t) = s(c_1 - c_2)e^{-s/t}/t^2, \quad 0 \leqslant c_1 < c_2 < 1, \quad s>0. \tag{1.5}$$

Explicitly we have

$$c(x) = (c_1 s + c_2 x)/(s + x). \tag{1.6}$$

We note the inequalities $c_1 \leqslant c(x) < c_2 < 1$, which ensure that the half-space is subcritical. This in turn implies that a unique bounded solution exists, and this solution satisfies the bound given by Eq. (1.3) (Ref. 1).

This problem has been considered in two recent articles.[2,3] In the first of these,[2] a set of continuum eigenfunctions is constructed which contains not only the usual delta and principal value functions, but also their derivatives. Additionally, a discrete eigenfunction is constructed, but only for c_1, c_2, and s satisfying a certain algebraic condition. These results are unsatisfactory in several respects. First, the extremely singular nature of the continuum solutions makes an accurate application of numerical methods a very difficult issue. Second, the fact that a discrete solution is constructed for only special values of c_1, c_2, and s places unphysical restrictions on the class of problems that can be analyzed. Third, no completeness proof for this set of eigenfunctions is given.

In a subsequent paper,[3] the same problem is considered, and progress is made on the first two of the above three issues. In particular, a new set of continuum eigenfunctions is constructed whose singular parts contain only the usual delta and principal value functions. Also, a discrete solution is constructed for all values of s if $c_1 < c_2$, and for an infinite denumerable set of values of s if $c_1 > c_2$. However, again, no completeness proof is given for this new set of eigenfunctions. The results in both of these papers,[2,3] as well as other articles containing results for transport problems with a variable $c(x)$, have recently been reviewed by Larsen.[4]

In the present paper, we return again to the problem defined by Eqs. (1.1)–(1.6), and for $c_1 < c_2$ we are able to settle all three issues mentioned above. First, we construct a third set of continuum eigenfunctions whose singular parts, like those obtained earlier,[3] contain only the usual delta and principal value functions. However, these eigenfunctions differ from the earlier ones in a crucial way, described later in Sec. II. Second, to this new set of continuum solutions we append the earlier discrete mode,[3] but modified in a simple way. Third, for this set of continuum plus one discrete solutions, we prove half-range completeness on the range $0 \leqslant \mu \leqslant 1$, and we also show that the expansion coefficient for the continuum modes $[b(\nu), 0 \leqslant \nu \leqslant 1]$ is a bounded and continuous function of ν.

Our analysis significantly borrows from and relates to two other previous articles[5,6] dealing with transport problems given by Eqs. (1.1)–(1.3). In particular, our derivation of the continuum modes closely follows an earlier derivation[5] in which $c(x)$ is given by Eq. (1.4) with $c_2 = 0$ and $\alpha(t)$ arbitrary. Also, our completeness proof is based upon the Laplace transform technique used earlier[6] to prove half-range completeness for a set of eigenfunctions arising from the exponential atmosphere $[c(x) = c_0 \exp(-x/s)]$. Interestingly, it was shown in that paper[6] that the expansion coefficient $b(\nu)$ for the continuum modes is a highly singular function of ν, whereas for the bilinear (a ratio of two linear functions) $c(x)$ given by Eq. (1.6) we prove here that the expansion coefficient is both bounded and continuous. This disparity is probably due to the existence of a discrete eigenfunction for the bilinear $c(x)$, whereas for the exponential $c(x)$ no discrete mode was included in the analysis.

The interest in eigenfunction analyses of linear transport equations with a variable $c(x)$ commenced in the late 1970's, with initial work focused on the exponential atmo-

 0022-2488/84/051511-08$02.50

sphere,[7] first introduced in the astrophysical literature.[8] The subsequent work reported for the bilinear $c(x)$ was motivated by the fact that such a $c(x)$ also yields a transport equation which is amenable to a singular eigenfunction treatment. Unfortunately, it appears at this time that these two functional forms for $c(x)$, or slight generalizations of these forms, are the only functions $c(x)$ for which eigenfunctions can be generated. In particular, no progress has been reported for a periodic $c(x)$, of great interest in nuclear reactor analysis. There is one important and unfinished aspect of the present and all previous analyses of the transport equation (1.1) with a variable $c(x)$. To date, no closed form solutions of boundary value problems have been constructed. This is due to the extreme complexity and nonstandard nature of the singular integral equations relating the expansion coefficients to the incident distribution $g(\mu)$. In the present article, as earlier,[6] we prove that these equations possess a unique solution, but we do not explicitly construct the solution.

An outline of the remainder of this article follows. In Sec. II we derive the new continuum eigenfunctions and in Sec. III we obtain the modified discrete mode. In Sec. IV we obtain an equation for the Laplace transform of the density [the integral over μ of $\psi(x,\mu)$]. In Sec. V we use the inverse Laplace transform to obtain a representation for the density, and in Sec. VI we prove half-range completeness. Additionally, an appendix is included which gives some technical details needed to obtain certain bounds on the Laplace transform of the density.

II. CONTINUUM EIGENFUNCTIONS

To construct continuum solutions to Eq. (1.1) with $c(x)$ given by Eq. (1.4), we introduce the ansatz

$$\psi_\nu(x,\mu) = f_\nu(\mu)e^{-x/\nu} + \int_0^\infty \beta_\nu(t,\mu)e^{-x/\omega}\,dt, \tag{2.1}$$

where

$$\frac{1}{\omega} = \frac{1}{\nu} + \frac{1}{t}, \tag{2.2}$$

and

$$\int_{-1}^{1} f_\nu(\mu)d\mu = 1, \tag{2.3}$$

$$\int_{-1}^{1} \beta_\nu(t,\mu)d\mu = 0. \tag{2.4}$$

We assume $0<\nu<1$, and since $0<t<\infty$, we have

$$0<\omega = \nu t/(\nu+t)<\nu<1. \tag{2.5}$$

Introduction of Eqs. (2.1)–(2.4) into Eq. (1.1) yields

$$\left(1-\frac{\mu}{\nu}\right)f_\nu(\mu)e^{-x/\nu} + \int_0^\infty \left(1-\frac{\mu}{\omega}\right)\beta_\nu(t,\mu)e^{-x/\omega}\,dt = \frac{c_2}{2}e^{-x/\nu} + \frac{1}{2}\int_0^\infty \alpha(t)e^{-x/\omega}\,dt. \tag{2.6}$$

Equation (2.6) is satisfied if

$$(1-\mu/\nu)f_\nu(\mu) = c_2/2, \tag{2.7}$$

$$(1-\mu/\omega)\beta_\nu(t,\mu) = \alpha(t)/2. \tag{2.8}$$

Using Eqs. (2.3)–(2.5), we find that the solutions of the above two equations are given by

$$f_\nu(\mu) = \lambda_2(\nu)\delta(\nu-\mu) + c_2\nu/2(\nu-\mu), \tag{2.9}$$

$$\beta_\nu(t,\mu) = [\omega\alpha(t)/2][1/(\omega-\mu) - \rho(\omega)\delta(\omega-\mu)], \tag{2.10}$$

where, for $0<\nu<1$,

$$\lambda_2(\nu) = 1-(c_2\nu/2)\ln[(1+\nu)/(1-\nu)], \tag{2.11}$$

$$\rho(\omega) = \ln[(1+\omega)/(1-\omega)]. \tag{2.12}$$

Combining all of these results, Eq. (2.1) gives for the continuum solutions

$$\psi_\nu(x,\mu) = \left(\lambda_2(\nu)\delta(\nu-\mu) + \frac{c_2\nu}{2(\nu-\mu)}\right)e^{-x/\nu} + \int_0^\infty \frac{\omega\alpha(t)}{2} \times\left[\frac{1}{\omega-\mu} - \delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right)\right]e^{-x/\omega}\,dt. \tag{2.13}$$

Equation (2.13) holds for a general class of functions $\alpha(t)$. Specializing this result to the form given by Eq. (1.5), we obtain

$$\psi_\nu(x,\mu) = \left(\lambda_2(\nu)\delta(\nu-\mu) + \frac{c_2\nu}{2(\nu-\mu)}\right)e^{-x/\nu} + \frac{s(c_2-c_1)}{2}\int_0^\infty \frac{\omega e^{-s/t}}{t^2}\delta(\omega-\mu) \times \ln\left(\frac{1+\omega}{1-\omega}\right)e^{-x/\omega}\,dt + \frac{s(c_1-c_2)}{2}\int_0^\infty \frac{\omega e^{-st}}{t^2}\left(\frac{1}{\omega-\mu}\right)e^{-x/\omega}\,dt. \tag{2.14}$$

Making a change in integration variables from t to ω according to

$$\frac{1}{\omega} = \frac{1}{\nu} + \frac{1}{t}, \tag{2.15}$$

and defining

$$H_\nu(\mu) = \begin{cases} 1, & 0<\mu<\nu<1, \\ 0, & 0<\nu<\mu<1, \quad \mu<0, \end{cases} \tag{2.16}$$

the first integral on the right hand side of Eq. (2.14) can be written

$$\int_0^\infty \frac{\omega e^{-s/t}}{t^2}\delta(\omega-\mu)\ln\left(\frac{1+\omega}{1-\omega}\right)e^{-x/\omega}\,dt = \frac{H_\nu(\mu)}{\mu}\ln\left(\frac{1+\mu}{1-\mu}\right)\exp\left(\frac{s}{\nu}-\frac{s}{\mu}-\frac{x}{\mu}\right). \tag{2.17}$$

Similarly, with the change of variables

$$u = (x+s)\left(\frac{1}{\nu}+\frac{1}{t}-\frac{1}{\mu}\right), \tag{2.18}$$

and defining

$$E(x) = P\int_x^\infty \frac{e^{-u}}{u}\,du, \tag{2.19}$$

the second integral on the right-hand side of Eq. (2.14) can be written

$$\int_0^\infty \frac{\omega e^{-st}}{t^2}\left(\frac{1}{\omega-\mu}\right)e^{-x/\omega}\,dt = -\frac{1}{\mu}\exp\left(\frac{s}{\nu}-\frac{s}{\mu}-\frac{x}{\mu}\right)E\left[(x+s)\left(\frac{1}{\nu}-\frac{1}{\mu}\right)\right]. \tag{2.20}$$

Combining Eqs. (2.14), (2.17), and (2.20) we obtain

$$\psi_\nu(x,\mu)=\left(\lambda_2(\nu)\delta(\nu-\mu)+\frac{c_2\nu}{2(\nu-\mu)}\right)e^{-x/\nu}+\frac{s(c_2-c_1)}{2\mu}\exp\left(\frac{s}{\nu}-\frac{s}{\mu}-\frac{x}{\mu}\right)\times\left\{H_\nu(\mu)\ln\left(\frac{1+\mu}{1-\mu}\right)+E\left[(x+s)\left(\frac{1}{\nu}-\frac{1}{\mu}\right)\right]\right\}. \tag{2.21}$$

We note from Eqs. (2.1)–(2.4) that $\psi_\nu(x,\mu)$ satisfies the simple (and, as we shall see, useful) condition

$$\int_{-1}^{1}\psi_\nu(x,\mu)d\mu=e^{-x/\nu},\quad 0<\nu<1. \tag{2.22}$$

The continuum eigenfunctions derived earlier[3] are given by

$$\psi_\nu(x,\mu)=\left(\lambda_2(\nu)\delta(\nu-\mu)+\frac{c_2\nu}{2(\nu-\mu)}\right)e^{-x/\nu}+\frac{(c_2-c_1)s}{c_2\mu^2}H_\nu(\mu)\exp\left[\frac{c_1 s}{c_2}\left(\frac{1}{\nu}-\frac{1}{\mu}\right)-\frac{x}{\mu}\right]. \tag{2.23}$$

These eigenfunctions have the same singular parts as those in Eq. (2.21), but different regular parts. It can be seen that these regular parts are simpler than those in Eq. (2.21). However, these eigenfunctions do not satisfy the simple condition Eq. (2.22), and we shall see that the analysis in both Secs. III and IV depends upon the simplicity of Eq. (2.22).

III. DISCRETE EIGENFUNCTION

A discrete eigenfunction of Eqs. (1.1), (1.3), and (1.6) has been given earlier.[3] To define this eigenfunction, we first introduce

$$\lambda_n(\nu)=1-\frac{c_n\nu}{2}\ln\left(\frac{\nu+1}{\nu-1}\right),\quad n=1,2,\quad 1<\nu<\infty, \tag{3.1}$$

$$\lambda_n(\nu_n)=0,\quad 1<\nu_1<\nu_2<\infty, \tag{3.2}$$

$$\theta(\nu)=\int_{\nu_1}^{\nu}\frac{\lambda_1(t)}{t^2\lambda_2(t)}dt,\quad 1<\nu<\nu_2, \tag{3.3}$$

$$B(\nu)=e^{s\theta(\nu)}/\nu^2\lambda_2(\nu),\quad 1<\nu<\nu_2, \tag{3.4}$$

$$A(\mu)=\begin{cases}-\dfrac{2}{c_2}\exp\left\{s\left[\theta(1)+\dfrac{c_1}{c_2}\left(1-\dfrac{1}{\mu}\right)\right]\right\}, & 0<\mu\leqslant 1,\\ 0, & -1\leqslant\mu<0.\end{cases} \tag{3.5}$$

Then the discrete eigenfunction $\tilde{\psi}_0(x,\mu)$ is given by[3]

$$\tilde{\psi}_0(x,\mu)=A(\mu)e^{-x/\mu}+\int_1^{\nu_2}\frac{\nu B(\nu)}{\nu-\mu}e^{-x/\nu}d\nu. \tag{3.6}$$

This implies

$$\int_{-1}^{1}\tilde{\psi}_0(x,\mu)d\mu=\int_0^1 A(\nu)e^{-x/\nu}d\nu+\int_1^{\nu_2}\nu B(\nu)\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\nu}d\nu. \tag{3.7}$$

We see that the density (the integral over μ) corresponding to this discrete mode contains a continuum part (that is, an integrand with exponentials decaying faster than e^{-x}).

We wish to construct a new discrete eigenfunction whose corresponding density does not contain a continuum part. To accomplish this, we define a new discrete mode $\psi_0(x,\mu)$ as

$$\psi_0(x,\mu)=\tilde{\psi}_0(x,\mu)-\int_0^1 A(\nu)\psi_\nu(x,\mu)d\nu, \tag{3.8}$$

where the $\psi_\nu(x,\mu)$ are the continuum eigenfunctions defined by Eq. (2.21). Equivalently, we may write

$$\psi_0(x,\mu)=\int_0^1 A(\nu)[\delta(\nu-\mu)e^{-x/\nu}-\psi_\nu(x,\mu)]d\nu+\int_1^{\nu_2}\frac{\nu B(\nu)}{\nu-\mu}e^{-x/\nu}d\nu, \tag{3.9}$$

and then by Eq. (2.22) we have the desired result

$$\int_{-1}^{1}\psi_0(x,\mu)d\mu=\int_1^{\nu_2}\nu B(\nu)\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\nu}d\nu. \tag{3.10}$$

An alternate way to obtain a discrete mode with no continuum part for the corresponding density is to introduce the ansatz

$$\bar{\psi}_0(x,\mu)=\int_1^{\nu_2}\frac{\nu\bar{B}(\nu)}{\nu-\mu}e^{-x/\nu}d\nu+\int_0^1 A(\nu,\mu)e^{-x/\nu}d\nu, \tag{3.11}$$

with

$$\int_{-1}^{1}A(\nu,\mu)d\mu=0. \tag{3.12}$$

Inserting Eqs. (3.11) and (3.12) into Eq. (1.1) one obtains, for a general $c(x)$ of the form given by Eq. (1.4),

$$\int_1^{\nu_2}\bar{B}(\nu)e^{-x/\nu}d\nu+\int_0^1\left(1-\frac{\mu}{\nu}\right)A(\nu,\mu)e^{-x/\nu}d\nu=\frac{c_2}{2}\int_1^{\nu_2}\nu\bar{B}(\nu)\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\nu}d\nu+\frac{1}{2}\int_0^\infty\alpha(t)\int_1^{\nu_2}\nu\bar{B}(\nu)\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\omega}d\nu\,dt. \tag{3.13}$$

We interchange the order of integration of ν and t in the last term in Eq. (3.13), and change integration variables from t to ω according to

$$\omega=\nu t/(\nu+t). \tag{3.14}$$

We then interchange the order of integration of ω and ν, relabel ω and ν as ν and u, and equate integrands in Eq. (3.13) for each value of ν. This leads to the two equations for $\bar{B}(\nu)$ and $A(\nu,\mu)$:

$$\lambda_2(\nu)\bar{B}(\nu)=\frac{1}{2}\int_\nu^{\nu_2}\frac{u^3}{(u-\nu)^2}\ln\left(\frac{u+1}{u-1}\right)\alpha\left(\frac{u\nu}{u-\nu}\right)\bar{B}(u)du, \tag{3.15}$$

$$\left(1-\frac{\mu}{\nu}\right)A(\nu,\mu)=\frac{1}{2}\int_1^{\nu_2}\frac{u^3}{(u-\nu)^2}\ln\left(\frac{u+1}{u-1}\right)\alpha\left(\frac{u\nu}{u-\nu}\right)\bar{B}(u)du. \tag{3.16}$$

Specializing to $\alpha(t)$ as given by Eq. (1.5), Eq. (3.15) can

be converted, by differentiation with respect to ν, to a first-order differential equation for $\overline{B}(\nu)$. The solution of this equation is

$$\overline{B}(\nu)=\frac{e^{-s/\nu}}{\nu^2\lambda_2(\nu)}\exp\left[\frac{s}{2}(c_2-c_1)\int_1^{\nu}\frac{1}{u\lambda_2(u)}\ln\left(\frac{u+1}{u-1}\right)du\right]. \tag{3.17}$$

It can be seen that, aside from an overall multiplicative constant, $\overline{B}(\nu)$ is identical to $B(\nu)$ as defined by Eq. (3.4). The corresponding function $A(\nu,\mu)$ follows from Eq. (3.16) as

$$A(\nu,\mu)=\kappa\left(\frac{e^{-s/\nu}}{\nu(\nu-\mu)}-\delta(\nu-\mu)\frac{e^{-s/\nu}}{\nu}\int_{-1}^{1}\frac{d\mu'}{\nu-\mu'}\right), \tag{3.18}$$

with

$$\kappa=1-\exp\left[\frac{s}{2}(c_2-c_1)\int_1^{\nu_2}\frac{1}{u\lambda_2(u)}\ln\left(\frac{u+1}{u-1}\right)du\right]. \tag{3.19}$$

In obtaining Eq. (3.18) from Eq. (3.16) we have appended the delta function $\delta(\nu-\mu)$ to assure that Eq. (3.12) is satisfied.

IV. THE LAPLACE TRANSFORM OF THE DENSITY

We define $\phi(x)$ as

$$\phi(x)=\frac{1}{2}\int_{-1}^{1}\psi(x,\mu)d\mu\,, \tag{4.1}$$

where $\psi(x,\mu)$ is the solution of Eqs. (1.1)–(1.6). Aside from the factor one-half, $\phi(x)$ is just the density associated with the distribution function $\psi(x,\mu)$. In this section, we perform a Laplace transform of the problem defined by Eqs. (1.1)–(1.6) to obtain a suitable equation for $\hat{\phi}(z)$, the Laplace transform of $\phi(x)$. To do this we first define the transforms of $\psi(x,\mu)$ and $\phi(x)$ as

$$\hat{\psi}(z,\mu)=\int_0^{\infty}e^{-zx}\psi(x,\mu)dx\,, \tag{4.2}$$

$$\hat{\phi}(z)=\int_0^{\infty}e^{-zx}\phi(x)dx\,. \tag{4.3}$$

From Eq. (1.3) we have

$$|\psi(x,\mu)|\leqslant M_0\,, \tag{4.4}$$

and hence by Eqs. (4.2) and (4.3) we have for $\mathrm{Re}\,z>0$,

$$|\hat{\psi}(z,\mu)|\leqslant M_0/\mathrm{Re}\,z\,, \tag{4.5}$$

$$|\hat{\phi}(z)|\leqslant M_0/\mathrm{Re}\,z\,. \tag{4.6}$$

To proceed, we operate on Eq. (1.1) by the Laplace transform operator

$$L=\int_0^{\infty}e^{-zx}(\cdot)dx\,, \tag{4.7}$$

and perform elementary operations to obtain

$$\hat{\psi}(z,\mu)=\frac{1}{1+\mu z}\left[\mu\psi(0,\mu)+c_2\hat{\phi}(z)+\int_0^{\infty}\alpha(t)\hat{\phi}\left(z+\frac{1}{t}\right)dt\right]. \tag{4.8}$$

Since $\hat{\psi}(z,\mu)$ is analytic for $\mathrm{Re}\,z>0$, the bracketed term in Eq. (4.8) must vanish for $-1\leqslant\mu<0$ and $z=-1/\mu$. This produces the result

$$\psi(0,-\mu)=\frac{1}{\mu}\left[c_2\hat{\phi}\left(\frac{1}{\mu}\right)+\int_0^{\infty}\alpha(t)\hat{\phi}\left(\frac{1}{\mu}+\frac{1}{t}\right)dt\right],\quad 0<\mu\leqslant 1. \tag{4.9}$$

Thus, from Eq. (4.6), the exiting angular distribution $\psi(0,-\mu)$, $0<\mu\leqslant 1$ is an analytic, bounded function of μ. We now integrate Eq. (4.8) over μ, using the indentities

$$\hat{\phi}(z)=\frac{1}{2}\int_{-1}^{1}\hat{\psi}(z,\mu)d\mu\,, \tag{4.10}$$

$$\lambda_2\left(\frac{-1}{z}\right)=1-\frac{c_2}{2}\int_{-1}^{1}\frac{d\mu}{1+\mu z}\,. \tag{4.11}$$

Specializing the result to $\alpha(t)$ as given by Eq. (1.5), we obtain after elementary operations

$$H_1(z)\hat{\phi}(z)=H_2(z)-\frac{s(c_2-c_1)}{c_2}\int_z^{\infty}e^{-su}\hat{\phi}(u)du\,, \tag{4.12}$$

where we have defined the functions

$$H_1(z)=\left[\frac{\lambda_2(-1/z)}{1-\lambda_2(-1/z)}\right]e^{-sz}\,, \tag{4.13}$$

$$G(z)=\frac{1}{2}\int_{-1}^{1}\frac{\mu\psi(0,\mu)}{1+\mu z}d\mu\,, \tag{4.14}$$

$$H_2(z)=G(z)e^{-sz}/[1-\lambda_2(-1/z)]\,. \tag{4.15}$$

By inspection, the functions $H_1(z)$ and $H_2(z)$ are analytic everywhere in the complex plane except for the branch cuts $(-\infty,-1]$ and $[1,\infty)$. To proceed, we differentiate Eq. (4.12) to obtain

$$\frac{d}{dz}[H_1(z)\hat{\phi}(z)]=\frac{dH_2(z)}{dz}+\frac{s(c_2-c_1)}{c_2}e^{-sz}\hat{\phi}(z)\,, \tag{4.16}$$

which is a first-order differential equation for $\hat{\phi}(z)$. To solve this equation, we introduce a new function

$$R(z)=\int_0^z\frac{e^{-st}}{c_2H_1(t)}dt=\int_0^z\frac{1-\lambda_2(-1/t)}{c_2\lambda_2(-1/t)}dt. \tag{4.17}$$

The integrand in Eq. (4.17) is an analytic function in the complex t plane, except for the branch cuts $(-\infty,-1)]$ and $[1,\infty)$, and simple poles at $t=\pm 1/\nu_2$. Therefore $R(z)$ is an analytic function in the complex plane except for the branch cuts $(-\infty,-1/\nu_2]$ and $[1/\nu_2,\infty)$. The path of integration in Eq. (4.17) is any simple curve from 0 to z which does not intersect these cuts (and, of course, z itself must not lie on either of these cuts).

Making use of Eq. (4.17), we rewrite Eq. (4.16) in the form

$$\frac{d}{dz}\{\hat{\phi}(z)H_1(z)\exp[s(c_1-c_2)R(z)]\}=H_2'(z)\exp[s(c_1-c_2)R(z)]\,. \tag{4.18}$$

To proceed, we use two results derived in the Appendix to this paper, namely [see Eqs. (A14) and (A19)],

$$H_1(z)=O(z-1/\nu_2),\quad |z-1/\nu_2|\ll 1, \tag{4.19}$$

$$\exp[s(c_1-c_2)R(z)]=O(z-1/\nu_2)^a,\quad |z-1/\nu_2|\ll 1, \tag{4.20}$$

where $a>0$ is given by Eq. (A15). Combining these two results, we have

$$H_1(z)\exp[s(c_1-c_2)R(z)]$$
$$=O(z-1/\nu_2)^{1+a},\quad |z-1/\nu_2|\ll 1. \tag{4.21}$$

Using this result, together with the fact that $\hat{\phi}(z)$ is analytic in the right half-plane, we integrate Eq. (4.18) from $1/\nu_2$ to z along any curve not intersecting the cuts $(-\infty,-1/\nu_2]$ and $[1/\nu_2,\infty)$, assuming that z itself does not lie on either of these cuts. We find

$$\hat{\phi}(z)H_1(z)\exp[s(c_1-c_2)R(z)]$$
$$=\int_{1/\nu_2}^{z} H_2'(t)\exp[s(c_1-c_2)R(t)]dt. \tag{4.22}$$

We consider a point z in the left half-plane and not on the cut $(-\infty,-1/\nu_2]$. At any such point, $H_1(z)$ is, by inspection, analytic and nonzero, and the same is true for $\exp[s(c_1-c_2)R(z)]$. Hence at this point $\hat{\phi}(z)$ is also analytic and given by

$$\hat{\phi}(z)=\frac{\int_{1/\nu_2}^{z}H_2'(t)\exp[s(c_1-c_2)R(t)]dt}{H_1(z)\exp[s(c_1-c_2)R(z)]}. \tag{4.23}$$

We conclude that $\hat{\phi}(z)$ is analytic everywhere in the complex plane except for the cut $(-\infty,-1/\nu_2]$. An integration by parts in Eq. (4.23), making use of Eq. (4.20), produces the following alternate and useful expression for $\hat{\phi}(z)$:

$$\hat{\phi}(z)=\frac{H_2(z)}{H_1(z)}+s(c_2-c_1)\frac{\int_{1/\nu_2}^{z}H_2(t)R'(t)\exp[s(c_1-c_2)R(t)]dt}{H_1(z)\exp[s(c_1-c_2)R(z)]}. \tag{4.24}$$

V. REPRESENTATION OF THE DENSITY

In the previous section we showed that $\hat{\phi}(z)$ is analytic everywhere in the complex plane except for the cut $(-\infty,-1/\nu_2]$. In the Appendix to this paper, we prove three additional facts concerning $\hat{\phi}(z)$ for z in the left half-plane and not on the cut. First, there exist constants $M_1>0$ and $\alpha<1$ such that [see Eq. (A29)]

$$|\hat{\phi}(z)|\leqslant M_1/|z+1/\nu_2|^{\alpha},\quad |z+1/\nu_2|\ll 1. \tag{5.1}$$

Second, there exists a constant M_2 such that [see Eq. (A30)]

$$|\hat{\phi}(z)|\leqslant M_2/|z|,\quad |z|\gg 1. \tag{5.2}$$

Third, the limiting values of $\hat{\phi}(z)$ as z approaches the cut from above and below, $\hat{\phi}^{\pm}(u)$ for $-\infty<u\leqslant -1/\nu_2$, are continuous for all values of u except possibly at $u=-1/\nu_2$ [where $\hat{\phi}^{\pm}(u)$ can be unbounded as in Eq. (5.1)], and there exists a constant $M_3>0$ such that [see Eq. (A31)]

$$|\hat{\phi}^{+}(u)-\hat{\phi}^{-}(u)|\leqslant M_3/u^2,\quad u<-1/\nu_2,\quad |u|\gg 1. \tag{5.3}$$

Using these results, it is a simple exercise to consider the inverse Laplace transform

$$\phi(x)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\hat{\phi}(z)e^{zx}\,dz,\quad \gamma>-1/\nu_2, \tag{5.4}$$

deform the contour into the left half-plane (wrapping it around the branch cut in the usual manner), and argue that only the integrals along this branch cut contribute. Hence we have

$$\phi(x)=\frac{1}{2\pi i}\int_{-\infty}^{-1/\nu_2}a(u)e^{ux}\,du, \tag{5.5}$$

where

$$a(u)=\hat{\phi}^{+}(u)-\hat{\phi}^{-}(u). \tag{5.6}$$

Making the change of variables

$$\nu=-1/u, \tag{5.7}$$

and defining

$$b(\nu)=(1/2\pi i\nu^2)a(-1/\nu), \tag{5.8}$$

we can rewrite Eq. (5.5) as

$$\phi(x)=\int_0^1 b(\nu)e^{-x/\nu}\,d\nu+\int_1^{\nu_2}b(\nu)e^{-x/\nu}\,d\nu. \tag{5.9}$$

We observe from Eqs. (5.6), (5.8), and the discussion above Eq. (5.3) that $b(\nu)$ is a continuous, bounded function of ν for $0<\nu\leqslant 1$.

Next we show that there exists a constant b_0 such that

$$b(\nu)=b_0\nu B(\nu)\ln(\nu+1)/(\nu-1),\quad 1<\nu<\nu_2, \tag{5.10}$$

where $B(\nu)$ is defined by Eq. (3.4). To do this, let t be a point on the cut satisfying $-1<t<-1/\nu_2$. We let z approach t from above and below the real axis in Eq. (4.12), and take the difference between the two results. This gives

$$H_1(t)a(t)=\frac{s(c_2-c_1)}{c_2}\int_{-1/\nu_2}^{t}e^{-sy}a(y)dy, \tag{5.11}$$

where $a(t)$ is given by Eq. (5.6) and we have used the fact that $H_1(z)$ and $H_2(z)$ are analytic for $-1<z<-1/\nu_2$. We differentiate Eq. (5.11) with respect to t and solve the resulting first-order equation for $a(t)$ to obtain

$$a(t)H_1(t)\exp\left(\frac{s(c_1-c_2)}{c_2}\int_{-1/\nu_1}^{t}\frac{e^{-sy}}{H_1(y)}dy\right)=2\pi i e^{s\eta}, \tag{5.12}$$

where η is a constant. Setting $t=-1/\nu$, $y=-1/u$, and using Eqs. (4.11), (4.13), and (5.8), we get

$$b(\nu)=\left(\frac{c_2\nu\ln[(\nu+1)/(\nu-1)]}{2\nu^2\lambda_2(\nu)}\right)\exp\left\{s\left(\eta-\frac{1}{\nu}+\frac{(c_2-c_1)}{c_2}\int_{\nu_1}^{\nu}\frac{1-\lambda_2(u)}{\lambda_2(u)}\frac{du}{u^2}\right)\right\},\quad 1<\nu<\nu_2. \tag{5.13}$$

On the other hand, for any constant b_0, Eq. (3.4) can be written

$$b_0\nu B(\nu)\ln[(\nu+1)/(\nu-1)]=\left(\frac{c_2\nu\ln[(\nu+1)/(\nu-1)]}{2\nu^2\lambda_2(\nu)}\right)\times\exp\left[\ln\left(\frac{2b_0}{c_2}\right)+s\theta(\nu)\right],\quad 1<\nu<\nu_2. \tag{5.14}$$

The left-hand sides of Eqs. (5.13) and (5.14) will be equal only if the corresponding right-hand sides are equal. This in turn implies that one must be able to choose the constant b_0 such that the function $F(\nu)$, defined by

$$F(\nu)=-\left(\eta-\frac{1}{\nu}+\frac{(c_2-c_1)}{c_2}\int_{\nu_1}^{\nu}\frac{1-\lambda_2(u)}{\lambda_2(u)}\frac{du}{u^2}\right)+\left[\frac{1}{s}\ln\left(\frac{2b_0}{c_2}\right)+\theta(\nu)\right], \tag{5.15}$$

is identically zero. This is possible if and only if [see Eq. (3.3)]

$$F'(\nu)=\frac{1}{\nu^2}\left(\frac{\lambda_1(\nu)}{\lambda_2(\nu)}-1-\frac{(c_2-c_1)}{c_2\lambda_2(\nu)}[1-\lambda_2(\nu)]\right) \tag{5.16}$$

is identically zero. Introducing the definition of $\lambda_n(\nu)$, $n=1,2$ [see Eq. (3.1)] into Eq. (5.16), we find

$$F'(\nu)=0. \tag{5.17}$$

Thus there exists a unique constant b_0 such that Eq. (5.10) holds.

To summarize, we have shown that there exists a continuous bounded function $b(\nu)$ defined on $0<\nu<1$, and a unique constant b_0 such that

$$\phi(x)=\int_0^1 b(\nu)e^{-x/\nu}\,d\nu+b_0\int_1^{\nu_2}\nu B(\nu)\times\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\nu}\,d\nu. \tag{5.18}$$

VI. HALF-RANGE COMPLETENESS

Let the function $b(\nu)$ and the constant b_0 be as described in Eq. (5.18), and define the function

$$\chi(x,\mu)=2\int_0^1 b(\nu)\psi_\nu(x,\mu)d\nu+2b_0\psi_0(x,\mu). \tag{6.1}$$

where $\psi_\nu(x,\mu)$ is the continuum eigenfunction derived in Sec. II, and $\psi_0(x,\mu)$ is the discrete eigenfunction derived in Sec. III. By construction, $\chi(x,\mu)$ satisfies the transport equation

$$\mu\frac{\partial\chi(x,\mu)}{\partial x}+\chi(x,\mu)=\frac{c(x)}{2}\int_{-1}^1\chi(x,\mu')d\mu', \tag{6.2}$$

with $c(x)$ given by Eq. (1.6). Further, by Eqs. (2.22) and (3.10), we have

$$\frac{1}{2}\int_{-1}^1\chi(x,\mu)d\mu=\int_0^1 b(\nu)e^{-x/\nu}\,d\nu+b_0\int_1^{\nu_2}\nu B(\nu)\ln\left(\frac{\nu+1}{\nu-1}\right)e^{-x/\nu}\,d\nu. \tag{6.3}$$

Comparing Eqs. (4.1), (5.18), and (6.3) we conclude

$$\int_{-1}^1[\psi(x,\mu)-\chi(x,\mu)]d\mu=0. \tag{6.4}$$

We subtract Eq. (6.2) from Eq. (1.1) and make use of Eq. (6.4) to get

$$\mu\frac{\partial(\psi-\chi)}{\partial x}+(\psi-\chi)=0, \tag{6.5}$$

which has the solution

$$\psi(x,\mu)-\chi(x,\mu)=f(\mu)e^{-x/\mu}, \tag{6.6}$$

where $f(\mu)$ is to be determined. Since both $\psi(x,\mu)$ and $\chi(x,\mu)$ vanish as x increases without bound, we must take

$$f(\mu)=0,\quad -1\leqslant\mu\leqslant 0. \tag{6.7}$$

Introducing Eqs. (6.6) and (6.7) into Eq. (6.4) gives

$$\int_0^1 f(\mu)e^{-x/\mu}\,d\mu=0. \tag{6.8}$$

If we make a change of integration variables according to

$$\mu=1/(\eta+1), \tag{6.9}$$

Eq. (6.8) becomes

$$\int_0^\infty\frac{1}{(\eta+1)^2}f\left(\frac{1}{\eta+1}\right)e^{-x\eta}\,d\eta=0. \tag{6.10}$$

Thus the Laplace transform of the function $f[(\eta+1)^{-1}]/(\eta+1)^2$ vanishes, which implies

$$f(\mu)=0,\quad 0<\mu\leqslant 1. \tag{6.11}$$

Equations (6.7) and (6.11) in turn then imply

$$\psi(x,\mu)=\chi(x,\mu). \tag{6.12}$$

Thus, the unique solution of Eqs. (1.1)–(1.6) is given by Eq. (6.1). In particular, setting $x=0$ gives the result

$$g(\mu)=2\int_0^1 b(\nu)\psi_\nu(0,\mu)d\nu+2b_0\psi_0(0,\mu),\quad 0<\mu\leqslant 1. \tag{6.13}$$

This proves that the set of eigenfunctions derived in Secs. II and III is half-range complete on the interval $0<\mu\leqslant 1$, and by the results in Sec. V we have that the continuum expansion coefficient $2b(\nu)$ is continuous and bounded for $0<\nu\leqslant 1$.

Finally we note that this simple half-range completeness proof depends explicitly on Eq. (6.3), which in turn depends upon the simple condition Eq. (2.22) satisfied by the continuum eigenfunctions developed in Sec. II. The continuum eigenfunctions derived earlier,[3] given by Eq. (2.23), do not satisfy a simple condition of the form of Eq. (2.22), and therefore the present half-range completeness proof is not applicable to them.

ACKNOWLEDGMENTS

The work of two of the authors (R.D. and G.C.P.) was partially supported by the National Science Foundation. The work of the third author (E.W.L.) was performed under the auspices of the U.S. Department of Energy.

APPENDIX

In this Appendix we obtain results, concerning the bounds and smoothness of certain functions, which are used in Secs. IV and V. For ease of notation, we define the function

$$\xi(z)=\begin{cases} i\pi, & \operatorname{Im} z>0,\\ -i\pi, & \operatorname{Im} z<0.\end{cases} \tag{A1}$$

First, from Eq. (4.11) we have

$$\lambda_2\left(\frac{-1}{z}\right)=1-\frac{c_2}{2z}\ln\left(\frac{1+z}{1-z}\right). \tag{A2}$$

This function is analytic except for the two branch cuts $(-\infty,-1]$ and $[1,\infty)$, and $\lambda_2^{-1}(-1/z)$ is analytic except for the same two cuts and the two simple poles at $z=\pm 1/\nu_2$. It is easily shown from Eq. (A2) that

$$\lambda_2\left(\frac{-1}{z}\right)=1-\frac{i\pi c_2\xi}{2z}-\frac{c_2}{z^2}+O(z^{-3}),\quad |z|\gg 1. \tag{A3}$$

From Eq. (4.17) we have

$$R'(z)=[1-\lambda_2(-1/z)]/c_2\lambda_2(-1/z), \tag{A4}$$

and based upon the properties of $\lambda_2(-1/z)$ we conclude that

$R'(z)$ is analytic everywhere in the complex plane except for the two cuts $(-\infty,-1]$ and $[1,\infty)$, and the two simple poles at $z=\pm 1/\nu_2$. For large $|z|$ we have

$$R'(z)=\xi/2z+O(z^{-2}),\quad |z|\gg 1, \tag{A5}$$

and moreover the limiting values of $R'(z)$ as Im z goes to zero from above and below the real axis, $[R'(u)]^{\pm}$, are continuous functions of u except for the two simple poles at $u=\pm 1/\nu_2$. In the vicinity of $z=-1/\nu_2$, we have

$$\begin{aligned}\lambda_2(-1/z)&=\lambda_2(\nu_2)+\nu_2^2\lambda_2'(\nu_2)(z+1/\nu_2)+O(z+1/\nu_2)^2\\&=\nu_2^2\lambda_2'(\nu_2)(z+1/\nu_2)+O(z+1/\nu_2)^2,\end{aligned} \tag{A6}$$

and hence

$$R'(z)=\left(\frac{-1}{c_2\nu_2^2\lambda_2'(\nu_2)}\right)\left(\frac{1}{z+1/\nu_2}\right)+O(1),\quad \left|z+\frac{1}{\nu_2}\right|\ll 1. \tag{A7}$$

Similarly,

$$R'(z)=\left(\frac{-1}{c_2\nu_2^2\lambda_2'(\nu_2)}\right)\left(\frac{1}{z-1/\nu_2}\right)+O(1),\quad \left|z-\frac{1}{\nu_2}\right|\ll 1. \tag{A8}$$

The results of the above paragraph imply that $R(z)$, given by

$$R(z)=\int_0^z R'(t)dt, \tag{A9}$$

is analytic everywhere in the complex plane except for the two cuts $(-\infty,-1/\nu_2]$ and $[1/\nu_2,\infty)$, and

$$R(z)=(\xi/2)\ln z+O(1),\quad |z|\gg 1. \tag{A10}$$

In addition, the limiting values of $R(z)$ as Im z goes to zero from above and below the real axis, $R^{\pm}(u)$, are continuous functions of u except for logarithmic singularities at the two branch points $u=\pm 1/\nu_2$. Near these points, we have

$$R(z)=\left(\frac{1}{c_2\nu_2^2\lambda_2'(\nu_2)}\right)\ln\left(z+\frac{1}{\nu_2}\right)+O(1),\quad \left|z+\frac{1}{\nu_2}\right|\ll 1, \tag{A11}$$

$$R(z)=\left(\frac{-1}{c_2\nu_2^2\lambda_2'(\nu_2)}\right)\ln\left(z-\frac{1}{\nu_2}\right)+O(1),\quad \left|z-\frac{1}{\nu_2}\right|\ll 1, \tag{A12}$$

and hence

$$\exp[s(c_1-c_2)R(z)]=O(z+1/\nu_2)^{-a},\quad |z+1/\nu_2|\ll 1, \tag{A13}$$

$$\exp[s(c_1-c_2)R(z)]=O(z-1/\nu_2)^{a},\quad |z-1/\nu_2|\ll 1, \tag{A14}$$

where

$$a=s(c_2-c_1)/c_2\nu_2^2\lambda_2'(\nu_2)>0. \tag{A15}$$

Moreover, Eq. (A10) implies

$$\exp[s(c_1-c_2)R(z)]=O(1),\quad |z|\gg 1. \tag{A16}$$

Next, from Eqs. (4.13) and (A4) we have

$$\frac{1}{H_1(z)}=\left(\frac{1-\lambda_2(-1/z)}{\lambda_2(-1/z)}\right)e^{sz}=c_2R'(z)e^{sz}. \tag{A17}$$

This function $H_1^{-1}(z)$ is analytic everywhere in the complex plane except for the two branch cuts $(-\infty,-1]$ and $[1,\infty)$, and two simple poles at $z=\pm 1/\nu_2$. Near these points, we have

$$\frac{1}{H_1(z)}=\left(\frac{e^{-s/\nu_2}}{\nu_2^2\lambda_2'(\nu_2)}\right)\left(\frac{1}{z+1/\nu_2}\right)+O(1),\quad \left|z+\frac{1}{\nu_2}\right|\ll 1, \tag{A18}$$

$$\frac{1}{H_1(z)}=\left(\frac{-e^{s/\nu_2}}{\nu_2^2\lambda_2'(\nu_2)}\right)\left(\frac{1}{z-1/\nu_2}\right)+O(1),\quad \left|z-\frac{1}{\nu_2}\right|\ll 1, \tag{A19}$$

and for large z we have

$$1/H_1(z)=[\xi c_2/2z+O(z^{-2})]e^{sz},\quad |z|\gg 1. \tag{A20}$$

The limiting values of $H_1^{-1}(z)$ as Im z goes to zero from above and below the real axis, $[H_1^{-1}(u)]^{\pm}$, are continuous functions of u except for two simple poles at $u=\pm 1/\nu_2$.

Next, the function $G(z)$, defined as [see Eq. (4.14)]

$$G(z)=\frac{1}{2}\int_{-1}^{1}\frac{\mu\psi(0,\mu)}{1+\mu z}d\mu, \tag{A21}$$

is analytic everywhere in the complex plane except for the cuts $(-\infty,-1]$ and $[1,\infty)$. For $|z|\gg 1$ we have

$$\begin{aligned}G(z)&-\frac{1}{2z}\int_{-1}^{1}\psi(0,\mu)d\mu\\&=-\frac{1}{2z^2}\int_{-1}^{1}\frac{\psi(0,\mu)}{\mu+z^{-1}}d\mu\\&=-\frac{1}{2z^2}\left(\int_{-1}^{0}\frac{\psi(0,0^-)}{\mu+z^{-1}}d\mu\right.\\&\quad\left.+\int_{0}^{1}\frac{\psi(0,0^+)}{\mu+z^{-1}}d\mu+O(1)\right)\\&=(\ln z/2z^2)[\psi(0,0^-)-\psi(0,0^+)]+O(z^{-2}),\end{aligned} \tag{A22}$$

and hence

$$G(z)=\frac{1}{2z}\int_{-1}^{1}\psi(0,\mu)d\mu+\frac{\ln z}{2z^2}\langle\psi\rangle+O(z^{-2}),\quad |z|\gg 1, \tag{A23}$$

where

$$\langle\psi\rangle=\psi(0,0^-)-\psi(0,0^+). \tag{A24}$$

Moreover, the limiting values of $G(z)$ as Im z goes to zero from above and below the real axis, $G^{\pm}(u)$, are continuous functions of u except for logarithmic singularities at $u=\pm 1$.

Finally, we consider the function $H_2(z)$, defined as [see Eq. (4.15)]

$$H_2(z)=\{G(z)/[1+\lambda_2(-1/z)]\}e^{-sz}. \tag{A25}$$

This function is analytic everywhere in the complex plane except for the cuts $(-\infty,-1]$ and $[1,\infty)$. From Eqs. (A3) and (A22) we have

$$\begin{aligned}H_2(z)=\frac{1}{\xi c_2}\left(\int_{-1}^{1}\psi(0,\mu)d\mu+\frac{2\ln z}{z}\langle\psi\rangle\right.\\\left.+O(z^{-1})\right)e^{-sz},\quad |z|\gg 1,\end{aligned} \tag{A26}$$

and the limiting values of $H_2(z)$ as Im z goes to zero from above and below the real axis, $H_2^{\pm}(u)$, are continuous functions of u.

Combining Eqs. (A5), (A16), and (A26) we deduce

$$H_2(z)R'(z)\exp[s(c_1-c_2)R(z)] = O(e^{-sz}/|z|), \quad \mathrm{Re}\, z<0, \quad |z|\gg 1, \tag{A27}$$

and hence

$$\left|\int_0^z H_2(t)R'(t)\exp[s(c_1-c_2)R(t)]dt\right| = O(e^{-s\,\mathrm{Re}\,z}/|z|), \quad \mathrm{Re}\, z<0, \quad |z|\gg 1. \tag{A28}$$

From all of the above results and Eq. (4.24), $\hat{\phi}(z)$ is analytic everywhere in the complex plane except on the cut $(-\infty, -1/\nu_2]$. In the vicinity of the point $z=-1/\nu_2$, we have [see Eqs. (A13), (A18), and (4.23), whose numerator is $O(1)$ at $z=-1/\nu_2$]

$$\hat{\phi}(z) = O[(z+1/\nu_2)^{a-1}], \quad |z+1/\nu_2|\ll 1, \tag{A29}$$

where $a>0$ is defined by Eq. (A15). Also, by introducing Eqs. (A16), (A20), (A26), and (A28) into Eq. (4.24), we conclude

$$\hat{\phi}(z) = \frac{1}{2z}\int_{-1}^{1}\psi(0,\mu)d\mu + \frac{\ln z}{z^2}\langle\psi\rangle + O(z^{-2}), \quad \mathrm{Re}\, z<0, \quad |z|\gg 1. \tag{A30}$$

Moreover, the limiting values of $\hat{\phi}(z)$ as Im z goes to zero from above and below the real axis, $\hat{\phi}^{\pm}(u)$, are continuous functions of u for $-\infty<u<-1/\nu_2$. At the point $u=-1/\nu_2$, $\hat{\phi}^{\pm}(u)$ can have a singularity of the type described by Eq. (A29). Finally, from Eq. (A30) we have

$$\hat{\phi}^{+}(u)-\hat{\phi}^{-}(u) = O(u^{-2}), \quad u<0, \quad |u|\gg 1. \tag{A31}$$

In obtaining Eq. (A31) we have used the fact that ln z evaluated just above the cut differs from its value just below the cut by a term of $O(1)$, namely $2\pi i$, for any value u.

[1] E. W. Larsen, SIAM J. Math. Anal. **12**, 282 (1981).
[2] G. C. Pomraning and E. W. Larsen, J. Math. Phys. **21**, 1603 (1980).
[3] E. W. Larsen, G. C. Pomraning, and V. C. Badham, J. Math. Phys. **23**, 815 (1982).
[4] E. W. Larsen, Prog. Nucl. Energy **8**, 215 (1981).
[5] E. W. Larsen, J. Math. Phys. **22**, 1463 (1981).
[6] E. W. Larsen and T. W. Mullikin, J. Math. Phys. **22**, 856 (1981).
[7] T. W. Mullikin and C. E. Siewert, Ann. Nucl. Energy **7**, 205 (1980).
[8] J. B. Chamberlain and M. B. McElroy, Astrophys. J. **144**, 1148 (1966).

Proof of the decoupling theorem of field theory in Minkowski space[a)]

Edward B. Manoukian
Department of National Defence, Royal Military College of Canada, Kingston, Ontario, K7L 2W3, Canada

(Received 4 February 1983; accepted for publication 9 December 1983)

The decoupling theorem of quantum field theory is proved in Minkowski space. It states the vanishing property, in the distributional sense, of the renormalized Feynman amplitudes when any subset of the underlying masses are scaled to infinity. To prove the theorem we were able, in the process, to bound the corresponding integrals, in the $\epsilon \to +0$ limit, by similar Euclidean integrals. All subtractions of renormalization are assumed to be carried out at the origin of momentum space with the degree of divergence of a subtraction coinciding with the dimensionality of the corresponding subdiagram.

PACS numbers: 11.10.Jj

I. INTRODUCTION

A rigorous proof[1] (see also Ref. 2) of the decoupling[3] theorem of field theory in Euclidean space, directly in momentum space, is now available in the literature established in a model independent way. The usefulness of this result lies in the fact that renormalized Feynman amplitudes depending, as a subset of their masses, on some "heavy" masses may be "neglected." The proof[1] of the theorem in Euclidean space is quite involved as it makes a detailed use of the intricacies of subtractions of renormalization. (It is important to realize that the proof in Ref. 1 holds true with a Minkowski metric as well as for $\epsilon > 0$ with an $i\epsilon$ prescription in the denominators.) An earlier attempt was made to prove this theorem in Minkowski space as well. In the latter the theorem was proved[4] only in the case when all the masses appearing in the amplitude become large and is, therefore, of limited applicability. To prove the theorem, we use an elementary proof[5] for the existence of the $\epsilon \to +0$ limit of renormalized Feynman amplitudes, and we bound the resulting integrals in this limit by "similar Euclidean integrals." We then make use of our earlier results[1] already established in Euclidean space. We prove the vanishing property of the amplitudes in Minkowski space, in the distributional sense, when any subset of the underlying masses are scaled to infinity. We consider only nonzero masses. All the subtractions of renormalization are carried out at the origin of momentum space. The proof applies to theories with higher spin fields and with derivative couplings as well.

II. PROOF OF THE THEOREM

The renormalized Feynman amplitude (absolutely convergent for $\epsilon > 0$) may be written in the form

$$F_\epsilon(P,\mu^1,\mu^2) = \int_{\mathbf{R}^{4n}} dK\, A(P,K,\mu^1,\mu^2) \prod_{l=1}^{L} D_l^{-1}, \quad \epsilon > 0, \tag{1}$$

where

$$D_l = [Q_l^2 + \mu_l^2 - i\epsilon(\mathbf{Q}_l^2 + \mu_l^2)], \tag{2}$$

$Q_l = \Sigma_{i=1}^m a_{li} p_i + \Sigma_{i=1}^n b_{li} k_i$, $P = (p_1^0,\ldots,p_m^3)$; A is a polynomial in its arguments and, in general, the μ_j^{-1} as well.

Without loss of generality we take $\mu^1 = \{\mu_1,\ldots,\mu_s\}$, $\mu^2 = \{\mu_{s+1},\ldots,\mu_L\}$, where the latter denotes the set of those masses we wish to scale to infinity. In terms of Feynman parameters (1) may be rewritten as[6,7,5]

$$F_\epsilon(P,\mu^1,\mu^2) = \int_D d\alpha\, N(\alpha,P,\mu^1,\mu^2,\epsilon)[G_\epsilon(\alpha,P,\mu^1,\mu^2)]^{-t}, \tag{3}$$

where

$$G_\epsilon(\alpha,P,\mu^1,\mu^2) = pUp + M^2 - i\epsilon(\mathbf{p}\cdot U\mathbf{p} + M^2), \tag{4}$$

$$M^2 = M_1^2 + M_2^2,$$

$$M_1^2 = \sum_{i=1}^{s} \alpha_i \mu_i^2, \tag{5}$$

$$M_2^2 = \sum_{i=s+1}^{L} \alpha_i \mu_i^2,$$

U is a matrix rational in α, and continuous almost everywhere in D, where $D = \{\alpha = (\alpha_1,\ldots,\alpha_L), \alpha_i \geqslant 0, \Sigma_{i=1}^L \alpha_i = 1\}$. U may be extended[6] to a continuous function everywhere in D. $N(\alpha,P,\mu^1,\mu^2,\epsilon)$ is rational in α, and is a polynomial in the elements in P,μ^1,μ^2 in ϵ and, in general, in the μ_j^{-1} as well. t is some positive integer.

We prove the following theorem:

Theorem

$$\lim_{\eta\to\infty}\left(\lim_{\epsilon\to+0} T_\epsilon(f;\eta)\right) = 0,$$

where

$$T_\epsilon(f;\eta) = \int_{\mathbf{R}^{4m}} dP f(P) F_\epsilon(P,\mu^1,\eta\mu^2)$$

and $f(P) \in \mathcal{S}(\mathbf{R}^{4m})$, $\eta\mu^2 = \{\eta\mu_{s+1},\ldots,\eta\mu_L\}$.

To prove the theorem we use the identity[5]

$$\begin{aligned}[G_\epsilon(\alpha,P,\mu^1,\mu^2)]^{-t} = &-(\mathbf{p}\cdot U\mathbf{p} + M^2)^2 t(t+1)\int_\epsilon^1 d\lambda_1 \\ &\times \int_{\lambda_1}^1 d\lambda\, [G_\lambda(\alpha,P,\mu^1,\mu^2)]^{-t-2} \\ &+ (\mathbf{p}\cdot U\mathbf{p} + M^2) it(\epsilon - 1) \\ &\times [G_1(\alpha,P,\mu^1,\mu^2)]^{-t-1} \\ &+ [G_1(\alpha,P,\mu^1,\mu^2)]^{-t},\end{aligned} \tag{6}$$

[a)] Work supported by the Department of National Defence Award under CRAD No. 3610-637.

 0022-2488/84/051519-05$02.50

and substitute, in turn, the three expressions on the right-hand side of (6) for $[G_\epsilon]^{-t}$ in (3).

In general, A in (1) may be a polynomial in ϵ: $A=\Sigma_a(\epsilon)^a A_a$. From the existence[5-7] of the limit $\epsilon\to+0$ of $T_\epsilon(f;\eta)$ with A in it replaced in turn by A_a, we may without loss of generality suppose in the subsequent analysis that A is independent of ϵ.

We note

$$\frac{\mathbf{p}\cdot U\mathbf{p}+M^2}{|G_1(\alpha,P,\mu^1,\mu^2)|^{t+1}}\leqslant\frac{C}{[p_E Up_E+M^2]^t}, \tag{7}$$

where C is some positive constant, and $p_E Up_E\equiv \mathbf{p}\cdot U\mathbf{p}+p^0Up^0$. Accordingly, when the second term on the right-hand side of (6) is substituted for $[G_\epsilon]^{-t}$ in (3), we may bound the resulting integral in absolute value by

$$C'\int_D d\alpha|N(\alpha,P,\mu^1,\mu^2,\epsilon)|[p_E Up_E+M^2]^{-t}. \tag{8}$$

In turn we may readily bound the integral in (8) by the expression

$$C_0''\int_D d\alpha\int_{\mathbf{R}^{4n}}dK\,|A(P,K,\mu^1,\mu^2)|\left(\sum_{l=1}^{L}\alpha_l D_{lE}\right)^{-L}$$
$$=C''\int_{\mathbf{R}^{4n}}dK\,|A(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L}D_{lE}^{-1}\quad(<\infty), \tag{9}$$

where $D_{lE}=\mathbf{Q}_l^2+(Q_l^0)^2+\mu_l^2$. The validity of the transition from (8) to (9) follows from the following analysis. By using the Lagrange interpolating formula,[8] we may introduce a finite number of distinct nonvanishing values ϵ_i^* and write

$$N(\alpha,P,\mu^1,\mu^2,\epsilon)=\sum_\alpha(\epsilon)^\alpha N_a(\alpha,P,\mu^1,\mu^2)$$
$$=\sum_a(\epsilon)^a\sum_i c_a(\epsilon_i^*)N(\alpha,P,\mu^1,\mu^2,\epsilon_i^*).$$

Since

$$N(\alpha,P,\mu^1,\mu^2,\epsilon^*)[G_{\epsilon^*}(\alpha,P,\mu^1,\mu^2)]^{-t}$$
$$=C_{\epsilon^*}\int_{\mathbf{R}^{4n}}dK\,A(P,K,\mu^1,\mu^2)\left(\sum_l\alpha_l D_{l\epsilon^*}\right)^{-L}$$

and

$$|G_{\epsilon^*}(\alpha,P,\mu^1,\mu^2)|/[p_E Up_E+M^2]\leqslant C'_{\epsilon^*},$$

we may bound ($\epsilon\geqslant0$)

$$|N(\alpha,P,\mu^1,\mu^2,\epsilon)|[p_E Up_E+M^2]^{-t}$$
$$\leqslant\left(\sum_i\sum_a(\epsilon)^a|c_a(\epsilon_i^*)|\,|C_{\epsilon_i^*}|\,|C'_{\epsilon_i^*}|^t|C''_{\epsilon_i^*}|\right)$$
$$\times\int_{\mathbf{R}^{4n}}dK\,|A(P,K,\mu^1,\mu^2)|\left(\sum_{l=1}^{L}\alpha_l D_{lE}\right)^{-L}.$$

Upon integration of the latter on α, we obtain the bound in (9) for (8). Hence if we denote the resulting expression for $F_\epsilon(P,\mu^1,\mu^2)$ when the second term on the right-hand side of (6) is substituted for $[G_\epsilon(\alpha,P,\mu^1,\mu^2)]^{-t}$ by $F_\epsilon^{(2)}(P,\mu^1,\mu^2)$, we obtain (for all $\epsilon\geqslant0$):

$$|F_\epsilon^{(2)}(P,\mu^1,\mu^2)|\leqslant C''\int_{\mathbf{R}^{4n}}dK\,|A(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L}D_{lE}^{-1}, \tag{10}$$

where

$$\int_{\mathbf{R}^{4n}}dK\,A(P,K,\mu^1,\mu^2)\prod_{l=1}^{L}D_{lE}^{-1}\equiv F_E(P,\mu^1,\mu^2), \tag{11}$$

denotes a Euclidean version of $F_\epsilon(P,\mu^1,\mu^2)$ in (1). The $\epsilon\to+0$ limit of $F_\epsilon^{(2)}(P,\mu^1,\mu^2)$ (trivially) exists. Similarly when the third expression on the right-hand side of (6) is substituted for $[G_\epsilon]^{-t}$ in (3), and the resulting integral is denoted by $F_\epsilon^{(3)}(P,\mu^1,\mu^2)$, we obtain for the latter the bound (for all $\epsilon\geqslant0$)

$$|F_\epsilon^{(3)}(P,\mu^1,\mu^2)|\leqslant C'''\int_{\mathbf{R}^{4n}}dK\,|A(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L}D_{lE}^{-1}, \tag{12}$$

and the limit $\epsilon\to+0$ of $F_\epsilon^{(3)}(P,\mu^1,\mu^2)$ (trivially) exists.

To handle the first term on the right-hand side of (6) we introduce[7] a $\mathscr{C}^\infty$-function $\chi(x)$, $0\leqslant\chi(x)\leqslant1$, defined by $\chi(x)=0$ if $x<\frac{1}{3}$ and $\chi(x)=1$ if $x>\frac{2}{3}$, and we set $x\equiv p^0Up^0/(\mathbf{p}\cdot U\mathbf{p}+M^2)$. Accordingly, when the first term on the right-hand side of (6) is substituted for $[G_\epsilon]^{-t}$ in (3), the resulting expression may be written as $F_\epsilon^{(1)}=F_{\epsilon1}^{(1)}+F_{\epsilon2}^{(1)}$,

$$F_{\epsilon1}^{(1)}(P,\mu^1,\mu^2)=-t(t+1)\int_D d\alpha\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1 d\lambda\,N$$
$$\times(\alpha,P,\mu^1,\mu^2,\epsilon)[1-\chi(x)](\mathbf{p}\cdot U\mathbf{p}+M^2)^2$$
$$\times[G_\lambda(\alpha,P,\mu^1,\mu^2)]^{-t-2}, \tag{13}$$

$$F_{\epsilon2}^{(1)}(P,\mu^1,\mu^2)=-t(t+1)\int_D d\alpha\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1 d\lambda\,N$$
$$\times(\alpha,P,\mu^1,\mu^2,\epsilon)\chi(x)(\mathbf{p}\cdot U\mathbf{p}+M^2)^2$$
$$\times[G_\lambda(\alpha,P,\mu^1,\mu^2)]^{-t-2}. \tag{14}$$

Due to the property of the function $[1-\chi(x)]$, we may use the following bound ($\underline{\mu}^2\equiv\min_j\mu_j^2$):

$$|pUp+M^2-i\lambda(\mathbf{p}\cdot U\mathbf{p}+M^2)|\geqslant|pUp+M^2|$$
$$\geqslant\tfrac{1}{3}(\mathbf{p}\cdot U\mathbf{p}+M^2)\geqslant\underline{\mu}^2/3,$$

in (13). Also by using the continuity property of U in D, we may use

$$\left|\frac{\mathbf{p}\cdot U\mathbf{p}+M^2}{pUp+M^2-i\lambda(\mathbf{p}\cdot U\mathbf{p}+M^2)}\right|$$
$$\leqslant\left|\frac{\mathbf{p}\cdot U\mathbf{p}+M^2}{pUp+M^2}\right|\leqslant1+\left|\frac{p_0Up_0}{pUp+M^2}\right|$$
$$\leqslant\left[1+C\sum_{i=1}^{m}\frac{(p_i^0)^2}{\underline{\mu}^2}\right], \tag{15}$$

in (13), again due to the property of the function $[1-\chi(x)]$, where C is some positive constant and may be taken to be greater than one. Similarly we may also use the following bound in (13):

$$\left|\frac{p_E Up_E+M^2}{pUp+M^2-i\lambda(\mathbf{p}\cdot U\mathbf{p}+M^2)}\right|\leqslant\left[1+2C\sum_{i=1}^{m}\frac{(p_i^0)^2}{\underline{\mu}^2}\right]. \tag{16}$$

Accordingly, we may then derive the following bound for $F_{\epsilon1}^{(1)}(P,\mu^1,\mu^2)$ valid for all $0\leqslant\epsilon\leqslant1$:

$$|F^{(1)}_{\epsilon 1}(P,\mu^1,\mu^2)|$$
$$\leqslant\left[1+2C\sum_{i=1}^{m}\frac{(p_i^0)^2}{\mu^2}\right]^{t+2}$$
$$\times\left|\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1 d\lambda\right|\left|\int_D d\alpha|N(\alpha,P,\mu^1,\mu^2,\epsilon)|\right.$$
$$\times[\,p_E Up_E+M^2\,]^{-t}$$
$$\leqslant C^{(1)}\left[1+2C\sum_{i=1}^{m}\frac{(p_i^0)^2}{\mu^2}\right]^{t+2}$$
$$\times\int_{\mathbf{R}^{4n}}dK\,|A\,(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L}D_{lE}^{-1}. \tag{17}$$

The $\epsilon\to+0$ limit of $F^{(1)}_{\epsilon 1}(P,\mu^1,\mu^2)$ then follows by an elementary application of the Lebesgue dominated convergence theorem.

We now consider the integral $F^{(1)}_{\epsilon 2}(P,\mu^1,\mu^2)$ in (14). To this end we define

$$T^{(1)}_{\epsilon 2}(f)=-t(t+1)\int_{\mathbf{R}^{4m}}dPf(P)\int_D d\alpha\int_\epsilon^1$$
$$\times d\lambda_1\int_{\lambda_1}^1 d\lambda\,N(\alpha,\mathbf{p},\mu^1,\mu^2,\epsilon)(\mathbf{p}\cdot U\mathbf{p}+M^2)^2$$
$$\times\chi(x)[\,G_\lambda(\alpha,P,\mu^1,\mu^2)\,]^{-t-2}, \tag{18}$$

where $f(P)\in\mathscr{S}(\mathbb{R}^{4m})$. We also write[5-7]

$$N(\alpha,P,\mu^1,\mu^2,\epsilon)=\sum_{a,b}N_{ab}(\alpha,\mathbf{P},\mu^1,\mu^2)(p^0)^a\epsilon^b, \tag{19}$$

where[6]

$$\int_D d\alpha|N_{ab}(\alpha,\mathbf{P},\mu^1,\mu^2)|<\infty. \tag{20}$$

We may then set

$$T^{(1)}_{\epsilon 2ab}(f)=-t(t+1)\int_{\mathbf{R}^{4m}}(p^0)^a f(P)\int_D d\alpha$$
$$\times\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1 d\lambda\,N_{ab}(\alpha,\mathbf{P},\mu^1,\mu^2)(\mathbf{p}\cdot U\mathbf{p}+M^2)^2$$
$$\times\chi(x)[\,G_\lambda(\alpha,P,\mu^1,\mu^2)\,]^{-t-2}. \tag{21}$$

Using the identity

$$[\,G_\lambda(\alpha,P,\mu^1,\mu^2)\,]^{-t-2}=\frac{(-\frac{1}{2})^{t+1}}{(t+1)!}$$
$$\times\left[(p_0Up_0)^{-1}\sum_{i=1}^{m}p_i^0\frac{\partial}{\partial p_i^0}\right]^{t+1}$$
$$\times[\,G_\lambda(\alpha,P,\mu^1,\mu^2)\,]^{-1} \tag{22}$$

and the vanishing property of $f(p)$ at infinity together with all of its derivatives, we may integrate (21) by parts over P to rewrite the latter as (for $\epsilon>0$)

$$-\frac{t(t+1)}{(t+1)!}\left(\frac{1}{2}\right)^{t+1}\int_{\mathbf{R}^{4m}}dP\int_D d\alpha\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1 d\lambda$$
$$\times N_{ab}(\alpha,\mathbf{P},\mu^1,\mu^2)[\,G_\lambda(\alpha,P,\mu^1,\mu^2)\,]^{-1}$$
$$\times(\,\mathbf{p}\cdot U\mathbf{p}+M^2)^2\left\{\left[\sum_{i=1}^{m}\frac{\partial}{\partial p_i^0}p_i^0\frac{1}{p^0Up^0}\right]^{t+1}\right.$$
$$\left.\times\,\chi(x)(p^0)^a f(P)\right\}. \tag{23}$$

The expression in the curly brackets may be rewritten as a finite sum[5,7]

$$\frac{1}{(\,P^0Up^0)^{t+1}}\sum_i\chi^i_a(\,p^0,x,\alpha)h^i_a(P), \tag{24}$$

where the $\chi^i_a(\,p^0,x,\alpha)$ may be bounded by a polynomial in p^0 independent of α, and $h^i_a(P)\in\mathscr{S}(\mathbb{R}^{4m})$. To take the limit $\epsilon\to+0$ of (23), we use in it the bounds

$$|p^0Up^0|^{-t}\leqslant(4)^t(\,p_EUp_E+M^2)^{-t}, \tag{25}$$
$$|p^0Up^0|^{-1}\leqslant 3/[\,\mathbf{p}\cdot U\mathbf{p}+M^2] \tag{26}$$

[recall the definition of $\chi(x)$], and hence

$$|(\,p^0Up^0)^{-1}(\,p^0Up^0)^{-t}|$$
$$\leqslant A\,(\,\mathbf{p}\cdot U\mathbf{p}+M^2)^{-1}(\,p_EUp_E+M^2)^{-t}, \tag{27}$$

where $p_EUp_E\equiv\mathbf{p}\cdot U\mathbf{p}+p^0Up^0$. Also we may bound

$$|G_\lambda(\alpha,P,\mu^1,\mu^2)|^{-1}\leqslant 1/\lambda\,(\,\mathbf{p}\cdot U\mathbf{p}+M^2). \tag{28}$$

Accordingly we may bound (23) in absolute value by

$$C\int_{\mathbf{R}^{4m}}dP\sum_i|\tilde{h}^i_a(P)|\left|\int_\epsilon^1 d\lambda_1\int_{\lambda_1}^1\frac{d\lambda}{\lambda}\right|$$
$$\times\int_D d\alpha\,\frac{|N_{ab}(\alpha,\mathbf{P},\mu^1,\mu^2)|}{[\,p_EUp_E+M^2\,]^t}\quad(<\infty), \tag{29}$$

and the latter exists for all $0\leqslant\epsilon\leqslant 1$, where C is some positive constant. The existence of the $\epsilon\to+0$ limit of (23) then follows by an application of the Lebesgue dominated convergence theorem. Note that all of our analysis holds true whether μ^2 is scaled by a parameter $\eta\geqslant 1$ or not.

All told we then have

$$\lim_{\epsilon\to+0}T_\epsilon(f;\eta)=\hat{T}_0(f;\eta)$$
$$=T^{(1)}_{01}(f;\eta)+T^{(1)}_{02}(f;\eta)+T^{(2)}_0(f;\eta)$$
$$+T^{(3)}_0(f;\eta), \tag{30}$$

where

$$T^{(1)}_{01}(f;\eta)=-t(t+1)\int_{\mathbf{R}^{4m}}dPf(P)\int_D d\alpha\int_0^1 d\lambda_1\int_{\lambda_1}^1 d\lambda$$
$$\times N(\alpha,P,\mu^1,\eta\mu^2,0)[\,1-\chi(x)]$$
$$\times(\mathbf{p}\cdot U\mathbf{p}+M_1^2+\eta^2M_2^2)^2$$
$$\times[\,G_\lambda(\alpha,P,\mu^1,\eta\mu^2)\,]^{-t-2}, \tag{31}$$

$$T^{(1)}_{02}(f;\eta)=-\frac{(\frac{1}{2})^{t+1}}{(t-1)!}\sum_a\int_{\mathbf{R}^{4m}}dP\int_D d\alpha\int_0^1 d\lambda_1\int_{\lambda_1}^1 d\lambda$$
$$\times N_{a0}(\alpha,\mathbf{P},\mu^1,\eta\mu^2)$$
$$\times[\,G_\lambda(\alpha,P,\mu^1,\eta\mu^2)\,]^{-1}$$
$$\times(\,\mathbf{p}\cdot U\mathbf{p}+M_1^2+\eta^2M_2^2)^2$$
$$\times\left\{\left[\sum_{i=1}^{m}\frac{\partial}{\partial p_i^0}p_i^0\frac{1}{p^0Up^0}\right]^{t+1}\chi(x)(p^0)^a f(P)\right\}, \tag{32}$$

$$T^{(2)}_0(f;\eta)=-it\int_{\mathbf{R}^{4m}}dPf(P)\int_D d\alpha$$
$$\times N(\alpha,P,\mu^1,\eta\mu^2,0)(\mathbf{p}\cdot U\mathbf{p}+M_1^2+\eta^2M_2^2)$$
$$\times[\,G_1(\alpha,P,\mu^1,\eta\mu^2)\,]^{-t-1}, \tag{33}$$

$$T_0^{(3)}(f;\eta)=\int_{\mathbb{R}^{4m}} dP f(P)\int_D d\alpha$$

$$\times N(\alpha,P,\mu^1,\eta\mu^2,0)[G_1(\alpha,P,\mu^1,\eta\mu^2)]^{-t}. \quad (34)$$

We also have the bound

$$|T_{01}^{(1)}(f;\eta)+T_0^{(2)}(f;\eta)+T_0^{(3)}(f;\eta)|$$

$$\leqslant C_1\int_{\mathbb{R}^{4m}} dP\,|\tilde{f}(P)|\int_{\mathbb{R}^{4n}} dK$$

$$\times|A(P,K,\mu^1,\eta\mu^2)|\prod_{l=1}^{L} D_{lE\eta}^{-1}, \quad (35)$$

$$\tilde{f}(P)\in\mathscr{S}(\mathbb{R}^{4m})$$

where

$$D_{lE\eta}^{-1}=\mathbf{Q}_l^2+(Q_l^0)^2+c_l\mu_l^2, \quad (36)$$

$$c_l=\begin{cases}1, & l=1,\ldots,s,\\ \eta^2, & l=s+1,\ldots,L,\end{cases} \quad (37)$$

and C_1 is some positive constant. On the other hand we have from (29),

$$|T_{02}^{(1)}(f;\eta)|\leqslant C_2\sum_a\int_{\mathbb{R}^{4m}} dP\,|h_a(P)|$$

$$\times\int_D d\alpha|N_{a0}(\alpha,\mathbf{P},\mu^1,\eta\mu^2)|$$

$$\times[\,p_E Up_E+M_1^2+\eta^2M_2]^{-t} \quad (38)$$

for some positive constant C_2, and $h_a(P)=\Sigma_i\,|\tilde{h}_a^{\,i}(P)|\in\mathscr{S}(\mathbb{R}^{4m})$.

On the other hand we may use the generalized Lagrange interpolating formula[6,8] and introduce distinct finite values $p_{*t'}^0$ for p^0 and write

$$N_{a0}(\alpha,\mathbf{P},\mu^1,\eta\mu^2)$$

$$=\sum_{t'}N(\alpha,\mathbf{P},P_{*t'}^0,\mu^1,\eta\mu^2)A_a(P_{*t'}^0), \quad (39)$$

where the coefficients $A_a(P_{*t'}^0)$ are finite. We also use the continuity property of U in D to derive the bound

$$\left|\frac{p_{*t'}^0\,Up_{*t'}^0+\mathbf{p}\cdot U\mathbf{p}+M_1^2+\eta^2M_2^2}{p^0Up^0+\mathbf{p}\cdot U\mathbf{p}+M_1^2+\eta^2M_2^2}\right|$$

$$\leqslant 1+c'\sum_{i=1}^{m}\frac{(p_{*t'i}^0)^2}{\mu^2}, \quad (40)$$

and the property of $h_a(P)$ to write

$$|h_a(P)|\leqslant C^{(a)}\left[1+c\sum_{i=1}^{m}(p_i^0)^2+c\sum_{i=1}^{m}\mathbf{p}_i^2\right]^{-N_1}$$

$$\times\left[1+c\sum_{i=1}^{m}(p_i^0)^2+c\sum_{i=1}^{m}\mathbf{p}_i^2\right]^{-N_2}$$

$$\leqslant C^{(a)}\left[1+c\sum_{i=1}^{m}(p_i^0)^2\right]^{-N_1}\left[1+c\sum_{i=1}^{m}\mathbf{p}_i^2\right]^{-N_2} \quad (41)$$

for $c>0$; and arbitrary positive integers N_1 and N_2 which may be chosen as large as we please. We may then bound $T_{02}^{(1)}(f;\eta)$ in absolute value as

$$|T_{02}^{(1)}(f;\eta)|\leqslant C_2\sum_a C^{(a)}\sum_{t'}|A_a(P_{*t'}^0)\left[1+c'\sum_{i=1}^{m}\frac{(P_{*t'i}^0)^2}{\mu^2}\right]^t$$

$$\times\int_{\mathbb{R}^m} dP^0\left[1+c\sum(p_i^0)^2\right]^{-N_1} \quad (42)$$

$$\times\int_{\mathbb{R}^{3m}} d\mathbf{P}\left[1+c\sum\mathbf{p}_i^2\right]^{-N_2}\int_D d\alpha$$

$$\times|N(\alpha,\mathbf{P},P_{*t'}^0,\mu^1,\eta\mu^2)|$$

$$\times[\,\mathbf{p}\cdot U\mathbf{p}+p_{*t'}^0\,Up_{*t'}^0+M_1^2+\eta^2M_2^2]^{-t}$$

$$\leqslant C''\sum_a C^{(a)}\sum_{t'}|A_a(P_{*t'}^0)|\left[1+c'\sum_{i=1}^{m}\frac{(P_{*t'i}^0)^2}{\mu^2}\right]^t$$

$$\times\int_{\mathbb{R}^{3m}} d\mathbf{P}\left[1+c\sum_{i=1}^{m}\mathbf{p}_i^2\right]^{-N_2}$$

$$\times\int_{\mathbb{R}^{4n}} dK\,|A(\mathbf{P},P_{*t'}^0,K,\mu^1,\eta\mu^2)|\prod_{l=1}^{L} D_{lE\eta}^{*\,-1} \quad (43)$$

for N_1 sufficiently large, and where $D_{lE\eta}^{*}$ is identical to $D_{lE\eta}$ with p^0 in the latter replaced by $p_{*t'}^0$; C'' is some positive constant.

From (30), (35), and (43) we note that we have reduced our analysis to one in Euclidean space and the vanishing property of $\hat{T}_0(f;\eta)$ for $\eta\to\infty$ is then readily established from our earlier analysis.[1] To this end we consider first the integral

$$\int_{\mathbb{R}^{4m}} dP\,|\tilde{f}(P)|\left[\int_{\mathbb{R}^{4n}} dK\,|A(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L} D_{lE\eta}^{-1}\right], \quad (44)$$

appearing on the right-hand side of (35), and use the well-known property[9,10] of the following integral as a function of P and μ^1,μ^2:

$$\int_{\mathbb{R}^{4n}} dK\,|A(P,K,\mu^1,\mu^2)|\prod_{l=1}^{L} D_{lE\eta}^{-1}. \quad (45)$$

For any P the integral in (45) is rigorously bounded by η^{-1} times a polynomial in $\ln\eta$ for $\eta\to\infty$.[1] Quite generally we may repeat word for word the induction proof in Ref. 9 by applying in the process the Heine–Borel covering theorem for the P integral in (44) and investigate the property of the integral in (44) for $\eta\to\infty$. To this end we know[1,9,10] that the integral in (45) may be bounded by a constant times η^{-1} and some powers of the elements in P times arbitrary powers of the logarithm of these variables as well for $\eta\to\infty$ and any of the elements in P large enough. Due to the very welcome property of $\tilde{f}(P)\in\mathscr{S}(\mathbb{R}^{4m})$ appearing in the P integrand in (44), the power counting criterion[9] is trivially satisfied for the P integral in (44), and we may then find a constant $b>1$ and a constant $H>0$ such that for $\eta\geqslant b$ the integral in (44) may be bounded above as

$$\int_{\mathbb{R}^{4m}} dP\,|\tilde{f}(P)|\int_{\mathbb{R}^{4n}} dK\,|A(P,K,\mu^1,\eta\mu^2)|\prod_{l=1}^{L} D_{lE\eta}^{-1}$$

$$\leqslant\frac{H}{\eta}\times[\text{polynomial in }\ln\eta], \quad (46)$$

$$\eta\geqslant b,\ H>0.$$

We may repeat the above reasoning for each of the t' terms in (43) for the $\mathbf{P}$ integral with $p_{*t'}^0$ fixed, and choose the

positive integer N_2 in (43) sufficiently large to conclude that we may find constants $b_{t'} > 1$ such that[1] for $\eta \geqslant b_{t'}$,

$$\int_{\mathbf{R}^{3m}} d\mathbf{P}\left[1 + c\sum_{i=1}^{m} \mathbf{p}_i^2\right]^{-N_2} \int_{\mathbf{R}^{4n}} dK \times |A(\mathbf{P}, P^0_{*t'}, K, \mu^1, \eta\mu^2)| \prod_{l=1}^{L} D^{*-1}_{lE\eta} \leqslant \frac{H_{t'}}{\eta} \times [\text{polynomial in } \ln \eta], \quad \eta \geqslant b_{t'},\ H_{t'} > 0. \quad (47)$$

The proof of the theorem is then completed by taking the limit $\eta \to \infty$ in (46) and (47) and carrying out the finite sums over t' and a in (43). It is interesting to note that the upper bound rate of decrease in (46) and (47) is as for the situation in Euclidean space.[1]

[1] E. B. Manoukian, J. Math. Phys. **22**, 2258 (1981).
[2] J. Ambjørn, Commun. Math. Phys. **67**, 109 (1979).
[3] T. Appelquist and J. Carazzonne, Phys. Rev. D **11**, 2856 (1975); E. C. Poggio, H. R. Quinn, and J. B. Zuber, Phys. Rev. D **15**, 1630 (1977); E. G. J. Collins, F. Wilczek, and A. Zee, Phys. Rev. D **18**, 242 (1978); D. Toussaint, Phys. Rev. D **18**, 1626 (1978); Y. Kazama and Y. -P. Yao, Phys. Rev. D **21**, 1116, 1138 (1980); *ibid.* **22**, 514 (1980); T. Hagiwara and N. Nakazawa, Phys. Rev. D **23**, 959 (1981); B. A. Ovrut and H. J. Schnitzer, Nucl. Phys. B **179**, 381 (1981).
[4] E. B. Manoukian, Int. J. Theor. Phys. **22**, 315 (1983).
[5] E. B. Manoukian, J. Phys. A **16**, 4131 (1983).
[6] Y. Hahn and W. Zimmermann, Commun. Math. Phys. **10**, 330 (1968); W. Zimmermann, Commun. Math. Phys. **11**, 1 (1968).
[7] J. H. Lowenstein and E. Speer, Commun. Math. Phys. **47**, 43 (1976).
[8] E. Isaacson and H. B. Keller, *Analysis of Numerical Methods* (Wiley, New York, 1966).
[9] S. Weinberg, Phys. Rev. **118**, 838 (1960).
[10] E. B. Manoukian, J. Phys. G **8**, 599 (1982).

On the relation between complex manifolds and soliton theoretic constructions for self-dual fields

Jurgen Burzlaff[a),b)] and J. E. M. Hornos[c),d)]
J. W. Gibbs Laboratory, Physics Department, Yale University, New Haven, Connecticut 06511

(Received 29 April 1983; accepted for publication 29 July 1983)

We study the relation between the Atiyah–Drinfeld–Hitchin–Manin construction for self-dual fields, Ward's twistor method, and the Belavin–Zakharov and Forgacs–Horvath–Palla soliton theoretic methods. Starting from the vector spaces, in terms of which the Atiyah–Drinfeld–Hitchin–Manin construction is formulated, we calculate the transition matrices, in terms of which the twistor method is formulated, for the Belavin–Polyakov–Schwartz–Tyupkin instanton and the Prasad–Sommerfield monopole explicitly. Given the transition matrices, we show how to solve the inverse scattering problem à la Belavin–Zakharov and Forgacs–Horvath–Palla, respectively.

PACS numbers: 11.15. – q

1. INTRODUCTION

In recent years different methods have been successfully applied to construct self-dual gauge fields. Atiyah, Hitchin, Drinfeld, and Manin[1] (ADHM) solved the instanton problem for Euclidean Yang–Mills theory, in the sense that they reduced the self-duality equations to purely algebraic conditions. Using Ward's twistor method[2] and an inverse scattering method, respectively, Corrigan and Goddard[3] and Forgacs, Horvath, and Palla[4] have found static SU(2) Yang–Mills–Higgs n monopole solutions in the limit of vanishing Higgs potential with the maximal number of degrees of freedom. The question of regularity, which remains unanswered in the latter two approaches, could be more easily decided with the method of Nahm[5] who adapts the ADHM construction for instantons to monopoles. Nahm's method, on the other hand, does not yield explicit formulas for the gauge potentials.

It is, therefore, of interest to understand the relation between these methods. A better understanding of the connections might even help to solve the main open problem which is the proof of regularity of the monopole solutions. A first step in establishing the connection between the ADHM and the twistor scheme was taken by Osborn[6] who showed how to construct, in principle, certain vector spaces corresponding to self-dual fields and from these vector spaces, covariantly constant fields on anti-self-dual planes. For the 't Hooft and the Witten–Peng solutions, Osborn constructed the vector spaces explicitly.

In this paper we construct the covariantly constant fields on anti-self-dual planes explicitly for the Belavin–Polyakov–Schwartz–Tyupkin (BPST) instanton[7] and the Prasad–Sommerfield (PS) monopole.[8] From these fields we calculate the corresponding transition matrices and show how they lead to a linear system associated with the nonlinear self-duality equations. In this way we arrive at the soliton theoretic results of Belavin and Zakharov[9] for the BPST instanton and of Forgacs, Horvath, and Palla[4] for the PS monopole.

Thus starting from the ADHM method, we reproduce the solution to the inverse scattering problem via the twistor theoretic construction.

2. GENERAL FORMALISM

Following Ref. 6, we consider self-dual fields

$$F_{\mu\nu} = F^{*}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}, \quad \mu,\nu,\ldots = 1,2,3,4, \tag{2.1}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \tag{2.2}$$

holds with A_μ an element of the Lie algebra of the gauge group G. For self-dual fields, the linear equations

$$\begin{aligned} &D_{\alpha\alpha'}\psi^{\alpha'} = 0, \quad \alpha,\alpha' = 1,2, \\ &D_{\alpha\alpha'} = (D_\mu \bar{e}_\mu)_{\alpha\alpha'}, \quad D_\mu = \partial_\mu + A_\mu, \\ &e = (i\vec{\sigma},1_2), \quad \bar{e} = (-i\vec{\sigma},1_2), \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} &D^{\alpha'\alpha}\Omega = \psi^{\alpha'}\Delta^{\prime\alpha}, \\ &\Delta^{\prime\alpha} = a^{\prime\alpha} + b^{\prime}_{\alpha'}x^{\alpha'\alpha} \end{aligned} \tag{2.4}$$

are consistent and their solutions span linear spaces. For $G = \mathrm{SU}(N)$ and topological quantum number n there are n linearly independent solutions $\psi^{\alpha'}$ and $2n + N$ linearly independent solutions Ω.

Given these solutions it is possible to construct covariantly constant fields on anti-self-dual planes S_θ which are defined by homogeneous coordinates $\theta = (\eta_\alpha, \chi^{\alpha'})$ and the conditions

$$x^{\alpha'\alpha}\eta_\alpha = \chi^{\alpha'}, \quad \eta \neq 0. \tag{2.5}$$

In fact, $\omega = \Omega v$ with v satisfying

$$\psi^{\alpha'}(a^{\prime\alpha}\eta_\alpha + b^{\prime}_{\beta'}\chi^{\beta'})v = 0, \tag{2.6}$$

is covariantly constant on the anti-self-dual plane S_θ. That means

$$n_{l\mu}D_\mu\omega|_{x\in S_\theta} = 0 \tag{2.7}$$

holds for the two tangent vectors n_l to the anti-self-dual plane.

---

a) Max Kade Foundation Fellow.
b) Address after October 1982: Dublin Institute for Advanced Studies.
c) Supported in part by CNPq Brazil.
d) Permanent address: Instituto de Fisica e Quimica de S. Carlos, CP 369, S. Carlos (SP), Brazil CEP 06514.

 0022-2488/84/051524-04$02.50

Having found the above covariantly constant fields according to Ref. 6, we can calculate the transition matrix using the formulas of Corrigan *et al.*[10] To this end we cover the space of anti-self-dual planes by two coordinate patches

$$x^1_\theta = \begin{pmatrix} \chi^1/\eta_1 & 0 \\ \chi^2/\eta_1 & 0 \end{pmatrix}, \quad \eta_1 \neq 0,$$
$$x^2_\theta = \begin{pmatrix} 0 & \chi^1/\eta_2 \\ 0 & \chi^2/\eta_2 \end{pmatrix}, \quad \eta_2 \neq 0. \tag{2.8}$$

Then the equation

$$\omega_l(x^1_\theta) = g\omega_l(x^2_\theta), \tag{2.9}$$

defines the transition matrix g, which depends on $\zeta = \eta_1/\eta_2$, $\hat{\nu} = i\chi^1/\eta_2$, and $\mu = i\chi^2/\eta_2$ if $\omega_l(x)$ is homogeneous in θ.

From g we can extract the gauge potentials explicitly if it is possible to split g or any equivalent $\tilde{g}$ in the following way:

$$g[\hat{\mu} = i(x^{21}\zeta + x^{22}),\ \hat{\nu} = i(x^{11} + x^{12}\zeta^{-1}), \zeta\,] = g(x,\zeta) = h^{-1}(x,\zeta)k(x,\zeta), \tag{2.10}$$

where h is a Taylor series in ζ^{-1} and k is a Taylor series in ζ and $\det h = \det k = 1$ holds. An equivalent transition matrix can be defined by

$$\tilde{g} = \Lambda g \Xi, \tag{2.11}$$

where $\Lambda(\hat{\mu},\hat{\nu},\zeta) = \Lambda(x,\zeta)$ is a Taylor series in ζ^{-1}, and $\Xi(\hat{\mu},\hat{\nu},\zeta) = \Xi(x,\zeta)$ is a Taylor series in ζ with $\det \Lambda = \det \Xi = 1$.

Now $\tilde{g}$ can be split in the sense of Eq. (2.10) if $\tilde{g}$ is of the special triangular form

$$\tilde{g} = \begin{pmatrix} \zeta^n & \hat{\rho} \\ 0 & \zeta^{-n} \end{pmatrix} \tag{2.12}$$

with a Laurent series $\hat{\rho}(x,\zeta) = \hat{\rho}(\hat{\mu},\hat{\nu},\zeta)$. The potentials are in this case explicitly given by the equations

$$A_{\alpha'1} - \zeta A_{\alpha'2} = -(\partial_{\alpha'1}k - \zeta\partial_{\alpha'2}k)k^{-1}. \tag{2.13}$$

This completes the description of the twistor method.

The pair of Eq. (2.13) is nothing but a linear system associated with the nonlinear self-duality equations since the consistency condition between them is the self-duality condition. Looked at in this way, Ward's twistor method which codes the self-dual fields in terms of the transition matrix g is a way to solve the inverse scattering problem. It yields k's for which the right-hand side of Eq. (2.13) is linear in ζ and, therefore, allows the construction of the ζ-independent $A_{\alpha'\alpha}$ in terms of these k's. The consistency condition then guarantees that this construction leads to self-dual fields.

3. THE BELAVIN–POLYAKOV–SCHWARTZ–TYUPKIN INSTANTON

For the 't Hooft 1 instanton potential,

$$A_\mu = \tfrac{1}{2}(\bar{e}_\mu e_\nu - \delta_{\mu\nu}1_2)\partial_\nu \ln\phi \tag{3.1}$$

with

$$\phi = \sum_{i=0}^{1} \frac{\lambda_i^2}{|x - y_i|^2}, \tag{3.2}$$

the vector spaces defined by (2.3) and (2.4) are explicitly given by[6]

$$\epsilon_{\alpha'\beta'}\psi^{\beta'}_{i\alpha} = \tilde{\psi}_{i\alpha\alpha'} = \phi^{1/2}\partial_{\alpha\alpha'}(1/\phi\,|x - y_i|^2), \tag{3.3}$$

$$\Delta'^{\beta}_{ij\alpha'}\epsilon_{\beta\alpha} = \tilde{\Delta}'_{ij\alpha\alpha'} = \delta_{ij}(\bar{x} - \bar{y}_i)_{\alpha\alpha'} - \delta_{0j}(\lambda_i^2/\lambda_0^2)(\bar{x} - \bar{y}_0)_{\alpha\alpha'}, \tag{3.4}$$

$$\Omega_{i\alpha\alpha'} = \phi^{-1/2}(\bar{x} - \bar{y}_i)_{\alpha\alpha'}/|x - y_i|^2, \quad i,j = 0,1. \tag{3.5}$$

For $y_0 = (\vec{0}, \lambda_0)$ in the limit $\lambda_0 \to \infty$, the 't Hooft 1 instanton solution reduces to the BPST instanton with $\lambda = \lambda_1$ and $y = y_1$.

It is straightforward to check that Eqs. (2.3) and (2.4) which, with all indices displayed, read

$$(D_\mu)^\beta_\alpha \tilde{\psi}_{i\beta\alpha'}(e_\mu)^{\alpha'\gamma} = 0, \tag{3.6}$$

$$(D_\mu)^\beta_\alpha \Omega_{i\beta\alpha'} = -\tfrac{1}{2}\tilde{\psi}_{j\alpha\beta'}(e_\mu)^{\beta'\beta}\tilde{\Delta}'_{ji\beta\alpha'}, \tag{3.7}$$

are satisfied by the functions (3.2)–(3.4). Equation (2.6) implies the equation

$$\epsilon_{\alpha'\beta'}\epsilon_{ij}\lambda_i^2(\chi^{\alpha'} - y_j^{\alpha'\alpha}\eta_\alpha)v^{\beta'j} = 0, \tag{3.8}$$

which holds for

$$v_1^{\alpha'0} = (1/\eta_1)(\chi^{\alpha'} - y_0^{\alpha'\alpha}\eta_\alpha), \quad v_1^{\alpha'1} = 0, \tag{3.9}$$

and

$$v_2^{10} = -(\lambda_0/\eta_1)(\chi^2 - y_1^{2\alpha}\eta_\alpha),$$
$$v_2^{11} = -(\lambda_1^2/\lambda_0\eta_1)(\chi^2 - y_0^{2\alpha}\eta_\alpha), \quad v_2^{2i} = 0. \tag{3.10}$$

We have constructed two fields $\omega_{al} = \Omega_{i\alpha\alpha'}v_l^{\alpha'i}, l = 1,2$, which are covariantly constant on the anti-self-dual plane S_θ. For $y_0 = (\vec{0}, \lambda_0)$, $\lambda_1 = \lambda$, $y_i = y$ these fields read

$$\omega_1 = \phi^{-1/2}\begin{pmatrix} 1 \\ \zeta^{-1} \end{pmatrix}, \tag{3.11}$$

$$\omega_2 = \phi^{-1/2}\zeta^{-1}\begin{pmatrix} -i\hat{\mu} - y^{21}\zeta - y^{22} + \lambda^2\dfrac{(\bar{x} - \bar{y})_{11}}{|x - y|^2} \\ \lambda^2\dfrac{(\bar{x} - \bar{y})_{21}}{|x - y|^2} \end{pmatrix} \tag{3.12}$$

in the limit $\lambda_0 \to \infty$.

Given ω_1 and ω_2, we obtain the transition matrix g defined by Eq. (2.9) in terms of Λ, Ξ, and g according to Eq. (2.11) as

$$\Lambda = \begin{pmatrix} i\sqrt{\phi_1} & 0 \\ (i/\sqrt{\phi_1})\zeta^{-1} & -i/\sqrt{\phi_1} \end{pmatrix}, \tag{3.13}$$

$$\Xi = \begin{pmatrix} -(i/\sqrt{\phi_2})\zeta & -i\sqrt{\phi_2} \\ -i/\sqrt{\phi_2} & 0 \end{pmatrix}, \tag{3.14}$$

$$\tilde{g} = \begin{pmatrix} \zeta & \hat{\rho} \\ 0 & \zeta^{-1} \end{pmatrix} \tag{3.15}$$

with

$$\hat{\rho} = 1 + \lambda^2 y^2/|x_\theta^1 - y|^2\,|x_\theta^2 - y|^2, \quad \phi_m = \phi(x_\theta^m). \tag{3.16}$$

The reality condition is satisfied by g,

$$g(\hat{\mu},\hat{\nu},\zeta) = g^\dagger(-\hat{\nu}, -\hat{\mu}, -\bar{\zeta}^{-1}) \tag{3.17}$$

for real x_μ and y_μ and the equivalent $\tilde{g}$ is of the special trian-

gular form which makes it possible to split it.

For

$$\left|\frac{a_0}{a_1}\right| := \left|\frac{y^{12}x^{22}-y^{22}x^{12}}{y^2+y^{12}x^{21}-y^{22}x^{11}}\right| < |\zeta| < \left|\frac{y^2+y^{21}x^{12}-y^{11}x^{22}}{y^{21}x^{11}-y^{11}x^{21}}\right| =: \left|\frac{b_0}{b_1}\right|, \tag{3.18}$$

$$\hat{\rho} = 1 + \frac{\lambda^2}{|x-y|^2}\left(\frac{b_0}{b_0+b_1\zeta} - \frac{a_0\zeta^{-1}}{a_1+a_0\zeta^{-1}}\right), \tag{3.19}$$

$$k = \frac{1}{\sqrt{\rho_0}}\begin{pmatrix} 1 & -\dfrac{\lambda^2}{|x-y|^2}\dfrac{b_1}{b_0+b_1\zeta} \\ \zeta & 1+\dfrac{\lambda^2}{|x-y|^2}\dfrac{b_0}{b_0+b_1\zeta} \end{pmatrix} \tag{3.20}$$

with

$$\rho_0 = 1 + \lambda^2/|x-y|^2 \tag{3.21}$$

hold. A gauge transformation

$$k \to k' = \Gamma k, \quad h \to h' = \Gamma h \tag{3.22}$$

with

$$\Gamma = \frac{1}{|x-y|}\begin{pmatrix} (x-y)^{22} & (x-y)^{21} \\ (x-y)^{12} & (x-y)^{11} \end{pmatrix} \tag{3.23}$$

puts the gauge potentials into the familiar form of Ref. 7,

$$A'_{\alpha'\alpha} = -((x-y)^2/((x-y)^2+\lambda^2))(\partial_{\alpha'\alpha}\Gamma)\Gamma^{-1}. \tag{3.24}$$

We now look at the BPST instanton in the soliton theoretic context. We have found that Γk, with the Γ of Eq. (3.23) and the k of Eq. (3.20), solves the inverse scattering problem of Eq. (2.13) for the potentials (3.24). So does, of course, then any $F = \Gamma k K(\hat{\mu},\hat{\nu},\zeta)$. For $\lambda = 2$ and

$$K = \begin{pmatrix} \dfrac{1}{4}(i\hat{\nu}+y^{11}+y^{12}\zeta^{-1}) + \dfrac{2y^{11}}{y^2 - iy^{21}\hat{\nu}\zeta + iy^{11}\hat{\mu}} & -\dfrac{1}{4}(i\hat{\mu}\zeta^{-1}+y^{21}+y^{22}\zeta^{-1}) - \dfrac{2y^{11}}{y^2 - iy^{21}\hat{\nu}\zeta + iy^{11}\hat{\mu}} \\ -\dfrac{1}{2}(i\hat{\nu}\zeta + y^{11}\zeta + y^{12}) & \dfrac{1}{2}(i\hat{\mu} + y^{21}\zeta + y^{22}) \end{pmatrix}, \tag{3.25}$$

F reads

$$F = uI - \zeta fA - \zeta^{-1}fA^{\dagger} \tag{3.26}$$

with

$$u = -\frac{|x-y|^2+8}{4(|x-y|^2+4)^{1/2}}, \quad f = \frac{|x-y|^2}{4(|x-y|^2+4)^{1/2}},$$

$$A = \frac{1}{|x-y|^2}\begin{pmatrix} -(x^{11}-y^{11})(x^{21}-y^{21}) & (x^{21}-y^{21})^2 \\ -(x^{11}-y^{11})^2 & (x^{11}-y^{11})(x^{21}-y^{21}) \end{pmatrix}. \tag{3.27}$$

This is the solution to the inverse scattering problem for the BPST instanton by Belavin and Zakharov[9] which we have recovered from our formulation in terms of the transition matrix (3.15).

4. THE PRASAD–SOMMERFIELD MONOPOLE

The gauge potentials for the PS monopole can be written in the form (3.1).

$$\phi = (1/r)\sinh(r)e^{it} \tag{4.1}$$

holds in this case. The vector spaces defined by (2.3) and (2.4) are spanned by[6]

$$\tilde{\psi}_{\alpha\alpha'} = \phi^{1/2}\partial_{\alpha\alpha'}(\rho/\phi), \tag{4.2}$$

$$\tilde{\Delta}'_{\alpha\alpha'} = -i\overleftarrow{\partial}_{\gamma}\delta_{\alpha\alpha'} - \bar{x}_{\alpha\alpha'}, \tag{4.3}$$

$$\Omega_{\alpha\alpha'} = \phi^{-1/2}(e^{i\gamma x})_{\alpha\alpha'} \tag{4.4}$$

with

$$\rho = -(i/r)\sinh(\gamma r)e^{i\gamma t}. \tag{4.5}$$

With Ω given, the gauge potentials can be reconstructed using the formula

$$A_{\mu} = \int_0^1 d\gamma\, \Omega(x,1-\gamma)\partial_{\mu}\Omega^{c}(x,\gamma), \tag{4.6}$$

where

$$\Omega^{c} = \phi^{-1/2}e^{i\gamma\bar{x}} \tag{4.7}$$

holds in our case. It is true, in general, that given the linear spaces explicitly, the gauge potentials can be calculated explicitly. Thus the purpose of going from the Eqs. (4.2), (4.3), and (4.4) in the ADHM scheme to the transition matrix for the twistor construction cannot be to find the gauge potentials. The purpose is to yield a better understanding of the relation between these schemes which might be helpful because in general the linear spaces are not given explicitly in the ADHM or in the ADHM-Nahm construction.

It is, therefore, instructive to continue and calculate the fields

$$\omega_{\alpha} = \int_0^1 d\gamma\, \Omega_{\alpha\alpha'}v^{\alpha'}, \tag{4.8}$$

which are covariantly constant on anti-self-dual planes if v satisfies Eq. (2.6). Equation (2.6) leads to the condition

$$i\epsilon^{\alpha}_{\alpha'}\eta_{\alpha}\partial_{\gamma}v^{\alpha'} - \epsilon_{\alpha'\beta'}\chi^{\alpha'}v^{\beta'} = 0 \tag{4.9}$$

which admits the two solutions

$$v_1 = \begin{pmatrix} 0 \\ \exp(-i\gamma\chi^1/\eta_1) \end{pmatrix}, \tag{4.10}$$

$$v_2 = \begin{pmatrix} \exp(-i\gamma\chi^2/\eta^2) \\ 0 \end{pmatrix}. \tag{4.11}$$

From the covariantly constant fields (4.8), we can again calculate the transition matrix g using (4.10), (4.11), and definition (2.9). The result of this calculation is $g = \Lambda^{-1}\tilde{g}\Xi^{-1}$, with

$$\Lambda = \begin{pmatrix} (2(1-e^{\hat{\nu}})/\hat{\nu})^{1/2} & 0 \\ -(\hat{\nu}/2(1-e^{\hat{\nu}}))^{1/2}\zeta^{-1} & (\hat{\nu}/2(1-e^{\hat{\nu}}))^{1/2} \end{pmatrix}, \tag{4.12}$$

$$\Xi = \begin{pmatrix} -(\hat{\mu}/2(1-e^{\hat{\mu}}))^{1/2}\zeta & (2(1-e^{\hat{\mu}})/\hat{\mu})^{1/2} \\ -(\hat{\mu}/2(1-e^{\hat{\mu}}))^{1/2} & 0 \end{pmatrix}, \tag{4.13}$$

and

$$\tilde{g} = \begin{pmatrix} \zeta & 2(e^{\hat{\mu}} - e^{\hat{\nu}}/(\hat{\mu} - \hat{\nu}) \\ 0 & \zeta^{-1} \end{pmatrix}. \tag{4.14}$$

We have recovered the transition matrix for the PS monopole.[11]

The discussion of the soliton theoretic description of the PS monopole proceeds along similar lines as the corresponding discussion for the instanton. Given the transition matrix g, k and h from Eq. (2.10) yield solutions to the inverse scattering problem of Eq. (2.13). This is true particularly in the gauge defined by $A_{\alpha'2} = 0$, i.e., for $h'(\zeta^{-1} = 0) = \Gamma h = 1$, $k' = \Gamma k$. In this gauge, Forgacs, Horvath, and Palla[4] have given a solution to the same inverse scattering problem. Their solution, therefore, differs from k' by a matrix which only depends on $\hat{\mu}$, $\hat{\nu}$, and ζ. This holds true not only for the PS monopole but for the whole $4n - 1$ parameter family of n monopole solutions found in Ref. 3 and Ref. 4.

ACKNOWLEDGMENT

We are indebted to D. H. Tchrakian for clarifying discussions.

[1] M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, Phys. Lett. A **65**, 185 (1978).
[2] R. S. Ward, Phys. Lett. A **61**, 81 (1977); M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[3] E. Corrigan and P. Goddard, Commun. Math. Phys. **80**, 575 (1981).
[4] P. Forgacs, Z. Horvath, and L. Palla, *Proceedings of the Monopole Meeting*, Trieste 1981, edited by N. S. Craigie, P. Goddard, and W. Nahm (World Scientific, Singapore, 1982), p. 21.
[5] W. Nahm, in Ref. 4, p. 87.
[6] H. Osborn, Commun. Math. Phys. **86**, 195 (1982); also CERN preprint Ref. TH 3120, 1981.
[7] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[8] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[9] A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).
[10] E. F. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).
[11] R. S. Ward, Commun. Math. Phys. **79**, 317 (1981).

Composite gluons and scalar fields

W. R. Gutierrez[a] and L. F. Saez
Department of Physics, City College of the City University of New York, New York, New York 10031

(Received 12 April 1983; accepted for publication 29 July 1983)

We study a model of scalar fields in four dimensions which has been introduced previously as a representation of gauge vectors in terms of scalar fields. We argue that conventional weak coupling constant perturbation theory is not renormalizable; however, strong coupling expansion based on the lattice formulation is specially well suited to the model.

PACS numbers: 11.15.Ha

The scalar field model we are going to consider has been introduced as a representation of gauge vectors in terms of composite scalar fields.[1-3] The model can be viewed too as a generalization of two-dimensional sigma models to four-dimensional space-time.

First we introduce the model and discuss its various classical aspects. It is defined at the classical level as a parametrization of gauge vectors in terms of scalar fields as $A_\mu = V^+(x)\partial_\mu V(x)$; $V^+V = I_N$, where $V(x)$ is a rectangular matrix of $K \times N$ $(K > N)$ size of complex scalar fields. The constraint $V^+V = I$ is needed to obtain the usual local gauge transformation, $V' = VU^+$ and $A'_\mu = UA_\mu U^+ + UU^+{}_{,\mu}$.

The explicit form of $F^2_{\mu\nu}(V, V^+)$ is rather cumbersome; however, there is a simple and important parametrization in terms of proyector matrices $P(x)$,[4] $P(x) = V(x)V^+(x)$ from which we get

$$F_{\mu\nu} = V^+[\partial_\mu P, \partial_\nu P]V,$$

and (1)

$$\mathrm{Tr}(F^2_{\mu\nu}) = \mathrm{Tr}\{P[\partial_\mu P, \partial_\nu P]\}^2.$$

The matrices $P(x)$ satisfy $P^+ = P$, $\mathrm{Tr}\,P = N$ and they are invariant under the local group transformation, $P'(x) = P(x)$. It can be conversely shown that taking P as an independent variable subject to the constraints $P^+ = P$, $P^2 = P$, and $\mathrm{Tr}\,P = N$, then we have $P = VV^+$ and $V^+V = I(N \times N)$. The invariance of $P(x)$ under gauge transformations shows that beside $F^2_{\mu\nu}$, there are more gauge invariant terms that could be added to a Lagrangian.

The case of the abelian group U(1) is given by $A_\mu = C^*_a(x)C_{a,\mu}$, $C^*_a C_a = 1$. The Lagrangian becomes

$$F^2_{\mu\nu} = (C_{a,\mu}C^*_{a,\nu} - C^*_{a,\mu}C_{a,\nu})^2, \tag{2}$$

where we have identified $V_{a1}(x)|_{N=1} = C_a(x)$, $a = 1,\ldots,K$.

We want to show the equivalence of the previous representation with the proyector formalism and that there is a well-defined Hamiltonian. Using the basis of the 2×2 Hermitian matrices, we have $P = Ip_0/2 + s_a p_a$, $a = 1,2,3$ $(s_a = \sigma_a/2)$, where the conditions on the proyectors imply that $p_0 = 1$, $p_a^2 = 1$. Thus we find

$$F^2_{\mu\nu} = -\tfrac{1}{4}(p^2_{,\mu}p^2_{,\nu} - (p_{,\mu}p_{,\nu})^2). \tag{3}$$

Now we can check that the following form of A_μ gives the same Lagrangian:

$$A_\mu = \tfrac{1}{2}i(p_{1,\mu}p_2 - p_{2,\mu}p_1)/(1 + p_3). \tag{4}$$

This last form of A_μ can be derived also in terms of $C_a(x)$ $(a = 1,2)$ if we fix a gauge such that $C_1 = p_3$ and $C_2 = (p_1 + ip_2)(1 + p_3)^{-1/2}$.

In order to get a well-defined Hamiltonian, we solve explicitly the remaining constraints $p_a^2 = 1$, using spherical coordinates, $p = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. In terms of the angular variables we find

$$F^2_{\mu\nu} = -\tfrac{1}{4}(\theta_{,\mu}\phi_{,\nu} - \theta_{,\nu}\phi_{,\nu})^2\sin^2\theta$$

and (5)

$$A_\mu = -\tfrac{1}{2}i(1 - \cos\theta)\phi_{,\mu}.$$

This last form of $F^2_{\mu\nu}$ allows an explicit calculation of the Hamiltonian. Note that the Lagrangian contains, at most, quadratic powers of the temporal derivative. In fact, the canonical momenta are well defined, from where the Hamiltonian is given by

$$H = \begin{pmatrix}\pi_\theta\\ \pi_\phi\end{pmatrix}^T \frac{2g^2}{\sin^2\theta}\begin{pmatrix}\phi^2_{,i} & -\theta_{,i}\phi_{,i}\\ -\theta_{,i}\phi_{,i} & \theta^2_{,i}\end{pmatrix}^{-1}\begin{pmatrix}\pi_\theta\\ \pi_\phi\end{pmatrix} - \frac{1}{16g^2}(\theta_{,i}\phi_{,j} - \theta_{,j}\phi_{,i})^2\sin^2\theta. \tag{6}$$

In general any proyector of size $K \times N$ could be used to represent a theory locally invariant under U(N) or SU(N). However, the explicit algebraic solution of the independent scalar fields of such a general proyector is quite a difficult task and for us is unknown if this general solution has been found. Instead we have chosen a case where the proyector space has the attractive feature of containing the same number of independent variables associated with the canonical U(N) gauge vector model, namely $2N^2$ independent components. This situation arises from considering a proyector of $2N \times 2N$ size to represent a U(N) locally invariant model.

We already have shown the validity of this statement for the case of U(1) and next it will be shown for the U(2) group. It can be shown from general arguments based on the unitary gauge fixing and separately on the algebra of proyectors, that proyectors of $2N \times 2N$ size that represent U(N) local invariance have only $2N^2$ degrees of freedom.

To study the nonabelian U(2) case using proyectors of 4×4 dimension is useful to work in the base of Hermitian matrices, where the proyector acquires the following form:

$$P = \begin{pmatrix}(1 + t)(I/2) + x^a s^a & z^{0*}(I/2) + z^{a*}s^a\\ z^0(I/2) + z^a s^a & (1 - t)(I/2) + y^a s^a\end{pmatrix}. \tag{7}$$

[a] Supported in part by NSF grant PHY-78-24888 and CUNY Faculty Research Award 13956.

 0022-2488/84/051528-03$02.50

The conditions derived from $P^2 = P$ are satisfied by the following solutions, with $t \neq 0$,

$$x^a = -(1/2t)(z^{0*} z^a + z^0 z^{a*} + i z^{c*} z^b \epsilon_{cba}),$$
$$y^a = +(1/2t)(z^{0*} z^a + z^0 z^{a*} - i z^{c*} z^b \epsilon_{cba}), \tag{8}$$

and the condition $t^2 + x^2 + |z^K|^2 = 1$, $K = 0,1,2,3$.

At this point we have solved all the constraints associated to the $P^2 = P$ condition, and we still have the eight independent variables contained in z^K, which is the number of variables that corresponds exactly to the number of independent variables of U(2) Yang–Mills gauge theory after fixing gauge. There are solutions with fewer independent variables. For example, we can establish the relation between this general solution and the particular solution given by Kafiev[3] for the SU(2) group. That solution has four independent variables, which correspond to $t = -\phi^5$, $z^0 = \phi^4$, $z^a = -i\phi^a$, and $\phi^A \phi^A = 1$, $A = 1,...,5$. Further details related to the model resulting from this particular solution are given in the same reference. For completeness we give here the resulting Lagrangian, known as the Skyrme's model:

$$-4g^2 L = F_{\mu\nu}^2 = -c[(\partial_\mu \phi^A)^2 (\partial_\nu \phi^B)^2 - (\partial_\mu \phi^A \partial_\nu \phi^B)^2],$$
$$A = 1,...,n. \tag{9}$$

As we have seen previously, Skyrme's model is an important special case for the U(1) group if $n = 3$, $c = \frac{1}{4}$ and for the SU(2) group if $n = 5$, $c = \frac{1}{2}$.

Next, we present some general remarks to show the problems related with renormalizability that we have found in the attempt to perform the weak coupling constant expansion of the Skyrme's model, which is a particular case of the class of models under consideration, though, these types of difficulties should be encountered in the general case as well. We devote the final part to the lattice formulation of the Skyrme model. Our main goal there is to show that the strong coupling constant expansion of the generating functional is well defined in the sense of not containing divergent terms. The first significant term of the expansion is calculated and is compared with the corresponding term of Wilson's model.

To study the quantum properties of the model in the small coupling constant regime taking ϕ^A as field variables, we would need a term like the conventional kinetic energy for the free scalar field, to take it as the unperturbed Hamiltonian. Changing variables $\phi^A = \phi^A(\rho^B)$, we can indeed generate a term of the type $\partial_\mu \rho^A \partial_\nu \rho_B M^{AB}_{\mu\nu}$, where $M^{AB}_{\mu\nu}$ are constants; for example, shifting the variables as $\phi^A = \phi_0^A(x) + g\rho^A(x)$, where ϕ_0^A is a classical solution of the model. However, as we can see immediately, the model contains interactions of the type $(\partial\rho)^4$, which are the very typical structures of the Lagrangian. Therefore, after changing variables we have the Lagrangian terms of type $(\partial\rho)^2$, and $g^2(\partial\rho)^4$. Considering the canonical dimensions of ρ, the interaction $(\partial\rho)^4$ acquires a dimension (length)$^{-8}$ which will be in conflict with the conventional criterion of renormalizability. This criterion is based on the fact that all known renormalizable models in four dimensions only contain interactions with canonical dimensions of (length)$^{-a}$, $a \leqslant 4$. Moreover, if we try to eliminate the constraint $\phi^2 = 1$ to obtain the independent field variables, we would necessarily get interactions which contain all power of the fields. These types of interactions are known to be nonrenormalizable since only up to the fourth power of a scalar field is acceptable in four dimensions to get a renormalizable model.

However, the strong coupling constant expansion in the lattice formulation is very well suited to the model since it is finite term by term. The lattice formulation[5] is done in Euclidean space that is considered as the analytical continuation of Minkowski space. In analogy with statistical mechanics, the generating functional is defined as

$$Z = \int D\phi \exp(-\beta\varepsilon(\phi)), \tag{10}$$

where $\varepsilon = \int L$ (Euclidean).

Next we consider the usual hypercubical lattice with N sites. The derivatives are replaced by finite differences of variables located at neighbor sites.

$$\partial_\mu \phi(x) = \lim_{a\to 0}(1/a)[\phi(x_k - a_\mu) - \phi(x_k)] \to (1/a) \times (\phi_{k+\mu} - \phi_k), \tag{11}$$

where a is the lattice spacing. Therefore, the lattice action is given by

$$\int d^4x L = \lim_{N\to\infty, a\to 0} \sum_k \frac{a^4 L_k}{a^4} = \lim_{N\to\infty} \sum_{k=1}^N L_k,$$

where

$$L(\partial_\mu \phi^A_{(x)}) \to L_k = L(\phi^A_{k+\mu} - \phi^A_k). \tag{12}$$

The cancellation of the lattice spacing parameter in four dimensions and the condition $(\phi^A_k)^2 = 1$ will guarantee the finiteness of the strong coupling constant expansion. The generating functional acquires the form

$$Z = \int \prod_{k=1}^N d\phi_k \, \delta((\phi^A_k)^2 - 1) \exp\left(\beta \sum_{k=1}^N L_k\right). \tag{13}$$

A very important quantity that can be evaluated in the limit of $N \to \infty$ ($a \to 0$) is the *free energy* that is defined[5] in analogy with statistical mechanics theory by $F = \lim_{N\to\infty}(1/N)\ln Z$, from where all the statistical quantities can be derived as for example the average energy $E = \partial F/\partial\beta$.

The power expansion in $\beta = c/4g^2$ can be given systematically with the use of the usual partial derivative technique;

$$Z = \exp\left[\beta \sum_k L_k\left(\frac{\partial}{\partial j^A}\right)\right] \prod_{k=1}^N Z_0[j_k]\Big|_{j^A_k = 0}, \tag{14}$$

where $Z_0(j) = ((n/2) - 1)! I_{n/2-1}(j) j^{(1-(n/2))}$. $I_m(x)$ is the modified Bessel function and $j = (j^A j^A)^{1/2}$. Some additional formulas as well as the diagram technique can be found in Ref. 5. Note that the first term of the expansion which is proportional to β can always be eliminated by adding an appropriate constant to the lattice action. Therefore, the first significant term of the expansion is proportional to β^2 and the free energy up to this approximation is given by

$$F = \beta^2 6(52/n + 268/n^2 + 39/n(n+2)). \tag{15}$$

Now we want to compare with Wilson's[5] formulation,

which up to the same order is $F_{U(1)} = 1.5/g^4$ and $F_{SU(2)} = 3/g^4$. In the Skyrme's model[6] we would have for U(1), $\beta = 1/8g^2$, $n = 3$, and for SU(2), $\beta = 1/2g^2$, $n = 5$, from which we get $F_{U(1)} = 3.1(1.5/g^4)$ and $F_{SU(2)} = 11.1(3/g^4)$.

As we can see, there is not agreement in the numerical factors; however, this comparison cannot be taken as final. To make a formal comparison between the type of model we have considered and the conventional gauge invariant model we have to take into account other terms that have been neglected up to now, as for example, the determinant coming from the variables change from $A_\mu(x)$ to $P(x)$, which may modify the functional measure. This last problem is under investigation.

[1]M. Dubois-Violette and Y. Georgelin, Phys. Lett. B **82**, 251 (1979).
[2]I. Bars and M. Gunaydin, Phys. Lett. B **95**, 373 (1980).
[3]Yu. N. Kafiev, Phys. Lett. B **96**, 337.
[4]M. F. Atiytah, N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, Phys. Lett. A **65**, 185 (1978).
[5]K. Wilson, Phys. Rev. D **10**, 2445 (1974); R. Balian, J. M. Drouffe, and C. Itzykson, Phys. Rev. D **11**, 2104 (1975); M. Creutz, Rev. Mod. Phys. **50**, 561 (1978).
[6]T. H. Skyrme, Proc. R. Soc. (London) **262**, 237 (1961); J. Math. Phys. **12**, 1735 (1971).

Local gauge invariant Lagrangians in classical field theories

D. R. Grigore
Central Institute of Physics, Institute for Physics and Nuclear Engineering, Bucharest, P. O. Box MG-6, Romania

(Received 19 April 1983; accepted for publication 19 August 1983)

We investigate the most general local gauge invariant Lagrangian in the framework of classical field theory. We rederive essentially Utiyama's result. Our proof makes clear the importance of the so-called current condition, i.e., the requirement that the Noether currents are not identically zero. This condition is of importance both in the general motivation for the introduction of the Yang–Mills fields and for the actual proof. Some comments are made about the basic mathematical structure of the problem—the gauge group.

PACS numbers: 11.15.Kc

1. INTRODUCTION

In recent years there has been increasing evidence that gauge theories[1,2] are a serious candidate for a unified theory of elementary particles. These theories rely on the so-called gauge principle which asserts that the invariance of the theory must be a local rather than a global one. This hypothesis strongly restricts the type of possible interactions between the elementary fields. The only paper which investigates the most general Lagrangian (in the framework of classical field theory) which satisfies the gauge principle is, to our knowledge, that of Utiyama.[3]

In this paper we want to give an alternative proof to that of Utiyama using the theory of first-order partial differential equations. The necessity of giving a new proof stems from two factors. Firstly, Utiyama's proof although essentially correct is based on some redundant suppositions. Secondly, one of the basic, although not explicitly stated assumptions of Ref. 3 is that the introduction of gauge fields is a necessary ingredient for realizing a local gauge invariance. This assumption is not true because one can construct Lagrangians with local gauge invariance but without gauge fields.[4] The peculiarity of these Lagrangians is that the Noether currents are identically zero, so a basic motivation for the introduction of gauge fields is, besides the local gauge invariance, the requirement that these currents are not zero.[5] Our proof exhibits clearly that this condition, beside its role in the general motivation, is very important in the elaboration of the proof.

Let us state our main result. Given a simple Lie group G with n parameters, we define the so-called gauge group

$$\tilde{G} = \{\tilde{g}:\mathbb{R}^d \to G \,|\, \tilde{g} \in \mathscr{C}^\infty(\mathbb{R}^d)\}.$$

We are looking for the most general Lagrangian

$$\mathscr{L}(\chi^A, \chi^A_{,\mu}, A^J, A^J_{,\mu}) \quad A = 1,...,N;$$
$$J = 1,...,M; \quad \mu = 0,...,d-1$$

which is invariant under the transformation

$$\chi^A \to \chi'^A = \chi^A + \delta\chi^A, \quad \delta\chi^A = \xi^i (T_i)^A{}_B \chi^B,$$
$$A^J \to A'^J = A^J + \delta A^J, \quad \delta A^J = \xi^i (U_i)^J{}_K A^K + C^{J,\mu}{}_i \xi^i{}_{,\mu}$$

and satisfies the current condition.

Following the terminology of Pauli[6] we call the first transformation a gauge transformation of the first kind, and the second transformation a gauge transformation of the second kind. We want these transformations to be the infinitesimal form of some representation of the gauge group $\tilde{G}$ acting on χ's and A's, respectively. (ξ's are the infinitesimal x-dependent parameters.) Imposing some natural requirements that $\mathscr{L}$ depends nontrivially on A's, and the gauge transformation of the second kind is also nontrivial (these conditions will be formulated more clearly in Sec. 3) we find that the most general Lagrangian is of the following form: $\mathscr{L} = \mathscr{L}'(\chi^A, \chi^A_{;\mu}, F^k_{\mu\nu})$ and satisfies the relation (3.54). Here $\chi^A_{;\mu} \equiv \chi^A_{,\mu} - A^i_\mu (T_i)^A{}_B \chi^B$ and $F^k_{\mu\nu} \equiv A^k_{\mu,\nu} - A^k_{\nu,\mu} + \mathscr{F}^k{}_{ij} A^i_\mu A^j_\nu$. ($A^i_\mu$ are some linear combinations of the original A's with the transformation law $\delta A^k_\mu = \xi^i \mathscr{F}^k{}_{ij} A^j_\mu + \xi^k{}_{,\mu}$, $\mathscr{F}^k{}_{ij}$ are the structure constants of the group G, and summation over the dummy indices is used.) The result can be generalized to a direct product between a semisimple and an abelian group.

In Ref. 3, the same result is derived in two particular situations.

A. $\mathscr{L} = \mathscr{L}(\chi^A, \chi^A_{,\mu}, A^J)$.

Our proof shows that the supposition that $M = dn$ and the matrix C of elements $C^{J,\mu}{}_i$ is invertible, are not necessary.

B. $\mathscr{L} = \mathscr{L}(A^J, A^J_{,\mu})$.

The assumption that the A's are a linear combination of the Yang–Mills potentials A^i_μ with the transformation law stated above, is not necessary.

Our proof does not exclude *a priori*, terms of Pauli type $\chi F \chi$, and does not suppose that we have some additional invariance, like Poincaré invariance. If this condition is imposed, we prove that A^i_μ is a *cuadrivector* (for $d = 4$).

The paper is organized as follows. In Sec. 2, we give a brief account on the theory of first-order partial differential equations, and we illustrate the method with the example of a global invariant Lagrangian. In Sec. 3, we derive our main result, i.e., the most general form of a Lagrangian with a

 0022-2488/84/051531-08$02.50

local invariance. In Sec. 4, we try to exploit more completely the group structure of the theory. In Sec. 5, we draw some conclusions. Two lemmas are proved in the Appendix.

2. THE CASE OF A GLOBAL INVARIANCE

(a). First we give some details on the theory of first-order partial differential equations.[7] We consider the homogeneous system of equations

$$X_\alpha(\mathscr{L}) = D_\alpha \mathscr{L} = 0, \quad \alpha = 1,...,r, \tag{2.1}$$

where $\mathscr{L}$ is an unknown function of the independent variables μ^a $(a = 1,...,s)$ and $D_\alpha \equiv X_\alpha{}^a \partial/\partial\mu^a$. The given coefficients $X_\alpha{}^a$ are permitted to be functions of μ's. We suppose that Eqs. (2.1) are linear independent (LI), i.e., if we have $\lambda_\alpha(\mu)$ so that

$$\sum_\alpha \lambda_\alpha X_\alpha(\mathscr{L}) = 0,$$

then $\lambda_\alpha = 0$ for all α.

If $r \geqslant s$, from the first s equations of (2.1) we get $\partial\mathscr{L}/\partial\mu^a = 0$ (use is made of the hypothesis of linear independence), so (2.1) has only the trivial solution $\mathscr{L} = \text{const}$.

Let us consider the case $r < s$. It is easy to see that (2.1) has the following consequences:

$$[D_\alpha, D_\beta]\mathscr{L} = 0, \quad \alpha,\beta = 1,...,r, \tag{2.2}$$

and $[D_\alpha, D_\beta]$ is also a linear differential operator.

We have $r(r-1)/2$ new linear and homogeneous equations which are a consequence of the system (2.1). If among these equations we have some which are LI with respect to (2.1), we adjoin them to the initial equations and iterate the procedure. There are two possibilities.

It may happen that after a number of iterations we get a number greater than s of LI equations, in which case only the trivial solution exists.

It may happen that after a number of iterations we have a number smaller than s of LI equations (2.1'), and the procedure (2.2) does not furnish new equations. The system (2.1') is of the form (2.1) and verifies

$$[D_\alpha, D_\beta] = \mathscr{F}_{\alpha\beta}{}^{\gamma} D_\gamma, \quad \alpha = 1,...,r', \quad r' \geqslant r. \tag{2.3}$$

We call such a system complete. A complete system can have a nontrivial solution.

(b) Let $\mathscr{L}(\chi^A, \chi^A_{,\mu})$ be a Lagrangian invariant under the global transformation

$$\chi^A \to \chi'^A = T(g)^A{}_B \chi^B. \tag{2.4}$$

Here χ^A are some classical fields in $\mathbb{R}^d$, and $T(g)$ is a finite-dimensional representation of a n-parameter Lie group G. The infinitesimal form of (2.4) is

$$\chi'^A = \chi^A + \delta\chi^A, \quad \delta\chi^A = \xi^i (T_i)^A{}_B \chi^B, \tag{2.5}$$

where ξ^i are the infinitesimal parameters and T_i a representation of the Lie algebra $\mathscr{G}$ of G which corresponds to $T(g)$. The matrices T_i satisfy the relation

$$[T_i, T_j] = \mathscr{F}_{ij}{}^{k} T_k, \quad i,j = 1,...,n, \tag{2.6}$$

where $\mathscr{F}_{ij}{}^{k}$ are the structure constants of the group G.

It is easy to see that the invariance condition of under (2.5) reads

$$\frac{\partial\mathscr{L}}{\delta\chi^A}(T_i)^A{}_B \chi^B + \frac{\partial\mathscr{L}}{\partial\chi^A_{,\mu}}(T_i)^A{}_B \chi^B_{,\mu} = 0, \quad i = 1,...,n. \tag{2.7}$$

We suppose that $T(g)$ is a faithful representation, so that Eqs. (2.7) are LI. [If $T(g)$ is not faithful, then the set $H = \{g \in G \mid T(g) = 1\}$ is a normal subgroup of G, so we can build a faithful representation of the group G/H by putting $\hat{T}(\hat{g}) \equiv T(g)$; $\hat{g} \in G/H, \forall\, g \in \hat{g}$. One can see that the group and the representation relevant to our problem are G/H and $\hat{T}(\hat{g})$, so we can consider that $T(g)$ in the relation (2.4) is faithful.]

Let us introduce the vector

$$\mu^a = \begin{pmatrix} \chi^A \\ \chi^A_{,\mu} \end{pmatrix}, \quad a = 1,...,(d+1)N.$$

The system (2.7) can be written in the form (2.1), with $\alpha \to (i)$, $i = 1,...,n$, and

$$X_{(i)}{}^a = Y_{(i)}{}^a{}_b \mu^b, \quad i = 1,...,n. \tag{2.8}$$

The constants $Y_{(i)}{}^a{}_b$ which are different from zero are

$$Y_{(i)}{}^{\chi^A}{}_{\chi^B} = (T_i)^A{}_B, \quad Y_{(i)}{}^{\chi^A_{,\mu}}{}_{\chi^B_{,\nu}} = \delta^\nu_\mu (T_i)^A{}_B \quad i = 1,...,n \tag{2.9}$$

(we are using the more transparent notation $Y_{(i)}{}^{\mu^a}{}_{\mu^b} \equiv Y_i{}^a{}_b$).

Introducing the $(d+1)N \times (d+1)N$ matrices $Y_{(i)}$ with components $Y_{(i)}{}^a{}_b$ it is easy to prove that

$$[D_{(i)}, D_{(j)}] = [Y_{(i)}, Y_{(j)}]^a{}_b \mu^b \frac{\partial}{\partial\mu^a} \quad i,j = 1,...,n. \tag{2.10}$$

But from (2.9) we get

$$[Y_{(i)}, Y_{(j)}] = \mathscr{F}_{ij}{}^{k} Y_{(k)}, \quad i,j = 1,...,n, \tag{2.11}$$

so we have

$$[D_{(i)}, D_{(j)}] = \mathscr{F}_{ij}{}^{k} D_{(k)}, \quad i,j = 1,...,n. \tag{2.12}$$

The system is complete and, for $n \leqslant (d+1)N$, can have nontrivial solutions.

3. THE CASE OF LOCAL INVARIANCE

The Noether theorem[8] asserts that the existence of an invariance of the type (2.4), implies the existence of some conserved currents. In our case one can prove easily that the expressions

$$\mathscr{T}^\mu_i \equiv \frac{\partial\mathscr{L}}{\partial\chi^A_{,\mu}}(T_i)^A{}_B \chi^B, \quad i = 1,...,n, \ \mu = 0,...,d-1 \tag{3.1}$$

verify

$$\mathscr{T}^\mu_{i,\mu} = 0, \quad i = 1,...,n, \tag{3.2}$$

if χ^A is a solution of the Euler–Lagrange equation of motion.

If one tries to promote the global invariance (2.5) of $\mathscr{L}$ to a local one, i.e., ξ^i's become functions of x, then one gets identically

$$\mathscr{T}^\mu_i = 0, \quad i = 1,...,n; \quad \mu = 0,...,d-1 \tag{3.3}$$

(see Ref. 5).

A way out of this unusual situation is to introduce in $\mathscr{L}$, beside the fields χ, some "compensating" fields. These are the Yang–Mills fields. In fact let us suppose that the Lagrangian $\mathscr{L}(\chi^A, \chi^A_{,\mu}, A^J, A^J_{,\mu})$, $A = 1,...,N$; $J = 1,...,M$; $\mu = 0,..., d-1$, is invariant under the transformation

$$\chi^A \to \chi'^A = \chi^A + \delta\chi^A, \quad \delta\chi^A = \xi^i (T_i)^A{}_B \chi^B, \tag{3.4}$$

$$A^J \to A'^J = A^J + \delta A^J, \quad \delta A^J = \xi^i (U_i)^J{}_K A^K + C_i^{J,\mu} \xi^i{}_{,\mu}. \tag{3.5}$$

We are going to prove that the A-type fields are LI combination of the usual Yang–Mills fields.

We impose the following natural requirements.

(a) The dependence of $\mathscr{L}$ on A^J's and $A^J_{,\mu}$'s is nontrivial, i.e., we cannot find constants λ^J, not all of them zero so that $\mathscr{L}$ does not depend on $A = \Sigma_{J=1}^M \lambda^J A^J$ and $A_{,\mu}$.

(b) If $\Sigma_{J=1}^M \lambda^J C_i^{J,\mu} = 0$, for all μ and i, then $\lambda^J = 0$ for all J. The meaning of this condition is the following: If we are in the opposite situation, we introduce new independent fields $\tilde{A}^J = \Sigma_{J'=1}^M \mathscr{D}^{JJ'} A^{J'}$ ($\det \mathscr{D} \neq 0$) so that one of the new A's is $\Sigma_{J=1}^M \lambda^J A^J$. Then (3.5) tells us that this particular field has a first type transformation law, so we did not split the fields properly in the two classes from the beginning.

This condition tells us that the M dimensional vectors C_i^μ of components $C^{J,\mu}{}_i$ span the whole $\mathbb{R}^M$, so we have $dn \geqslant M$. We will see that in fact $dn = M$.

(c) We suppose that (3.4) and (3.5) are the infinitesimal form of some representation of the gauge group $\tilde{G}$. It is evident that if the T_i's satisfy (2.6), (3.4) is the infinitesimal form of the following representation of $\tilde{G}$:

$$\chi^A(x) \to \chi'^A(x) = T(g(x))^A{}_B \chi^B(x),$$

where $T(g)$ is the representation of G from Sec. 2(b). If we take in (3.5) $\xi^i = \text{const}$, we obtain that U^i's must satisfy a relation analogous to (2.6).

(d) The group G is simple. This condition is of a technical nature and can be relaxed.

(e) The current condition prevents (3.3) being true. Actually we will need a stronger assumption which we prefer to formulate and comment in the course of the proof.

Now we proceed with the proof of the assertion stated in the Introduction.

The invariance conditions are easy to find writing the variation of $\mathscr{L}$ under (3.4) and (3.5), and equating to zero the coefficients of $\xi^i{}_{,\mu\nu}, \xi^i{}_{,\mu}$, and ξ^i, respectively. One gets the following system:

$$C^{J,\mu}{}_i \frac{\partial \mathscr{L}}{\partial A^J_{,\nu}} + (\mu \leftrightarrow \nu) = 0 \quad \text{for all } i,\mu,\nu, \tag{3.6}$$

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}} (T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial A^J_{,\mu}} (U_i)^J{}_K A^K + C^{J,\mu}{}_i \frac{\partial \mathscr{L}}{\partial A^J} = 0$$

for all i,μ, (3.7)

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}} (T_i)^A{}_B \chi^B_{,\mu} + \frac{\partial \mathscr{L}}{\partial \chi^A} (T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial A^J_{,\mu}} (U_i)^J{}_K A^K_{,\mu} + \frac{\partial \mathscr{L}}{\partial A^J} (U_i)^J{}_K A^K = 0$$

for all i. (3.8)

Like in Sec. (2b), we can arrange so that (3.6)–(3.8) are LI.

One can put the system (3.6)–(3.8) under the standard form (2.1), introducing the vector

$$\mu^a = \begin{pmatrix} \chi^A \\ \chi^A_{,\mu} \\ A^J \\ A^J_{,\mu} \end{pmatrix}, \quad a = 1,\ldots,(d+1)(M+N).$$

The coefficients $X_\alpha{}^a$ are of the form

$$X_\alpha{}^a = Y_\alpha{}^a{}_b \mu^b + Z_\alpha{}^a. \tag{3.9}$$

Inspecting (3.6)–(3.8), we see that the constants $Y_\alpha{}^a{}_b$, $Z_\alpha{}^a$ different from zero are the following.

(i) From Eq. (3.6), $\alpha = (i;\mu,\nu)$ ($\frac{1}{2} d(d+1) n$ values)

$$Z_{(i;\mu,\nu)}{}^{A^J_{,\rho}} = \delta^\nu_\rho C^{J,\mu}{}_i + (\mu \leftrightarrow \nu).$$

(ii) From Eq. (3.7), $\alpha = (i;\mu)$ (dn values)

$$Y_{(i;\mu)}{}^{\chi^A_{,\nu}}{}_{\chi^B} = \delta^\mu_\nu (T_i)^A{}_B,$$

$$Y_{(i;\mu)}{}^{A^J_{,\nu}}{}_{A^K} = \delta^\mu_\nu (U_i)^J{}_K,$$

$$Z_{(i;\mu)}{}^{A^J} = C^{J,\mu}{}_i.$$

(iii) From Eq. (3.8), $\alpha = (i)$ (n values)

$$Y_{(i)}{}^{\chi^A_{,\mu}}{}_{\chi^B_{,\nu}} = \delta^\nu_\mu (T_i)^A{}_B, \quad Y_{(i)}{}^{\chi^A}{}_{\chi^B} = (T_i)^A{}_B,$$

$$Y_{(i)}{}^{A^J_{,\mu}}{}_{A^K_{,\nu}} = \delta^\nu_\mu (U_i)^J{}_K, \quad Y_{(i)}{}^{A^J}{}_{A^K} = (U_i)^J{}_K.$$

(We have used again the more transparent notations

$$Y_\alpha{}^a{}_b \equiv Y_\alpha{}^{\mu^a}{}_{\mu^b}, \quad Z_\alpha{}^a \equiv Z_\alpha \mu^a.)$$

The index α takes $(d+1)(d+2)n/2$ values. The following comment is necessary: In deriving Eqs. (3.6)–(3.8) we have implicitly assumed that $\mathscr{L}$ has a nontrivial dependence on all $(d+1)(M+N)$ variables μ^a. The situation in which $\mathscr{L}$ does not depend on μ^{a_0}, can be treated by adding to (3.6)–(3.8) the equation $\partial \mathscr{L}/\partial \mu^{a_0} = 0$. So we assume from now on that $\mathscr{L}$ does depend nontrivially on all the variables μ^a without losing the generality.

We proceed now with the iterative program outlined in Sec. (2a). A simple calculation yields

$$[D_\alpha, D_\beta] = -[Y_\alpha, Y_\beta]^a{}_b \mu^b \frac{\partial}{\partial \mu^a} - (Y_\alpha Z_\beta - Y_\beta Z_\alpha)^a \frac{\partial}{\partial \mu^a}, \tag{3.10}$$

where Y_α is the $(d+1)(M+N) \times (d+1)(M+N)$ matrix of elements $Y_\alpha{}^a{}_b$, and Z_α the $(d+1)(M+N)$ vector of components $Z_\alpha{}^a$.

Now, a simple but rather long calculation yields

$$[Y_\alpha, Y_\beta] = \mathscr{F}^\gamma_{\alpha\beta} Y_\gamma, \tag{3.11}$$

where the nonzero constants $\mathscr{F}^\gamma_{\alpha\beta}$ are

$$\mathscr{F}^{(k;\nu)}_{(i;\mu)(j)} = -\mathscr{F}^{(k;\nu)}_{(j)(i;\mu)} = \delta^\mu_\nu \mathscr{F}^k_{ij},$$

and

$$\mathscr{F}^{(k)}_{(i)(j)} = \mathscr{F}^k_{ij}.$$

Using (3.11) in (3.10) one obtains

$$[D_\alpha, D_\beta] = -\mathscr{F}^\gamma_{\alpha\beta} D_\gamma - (Y_\alpha Z_\beta - Y_\beta Z_\alpha - \mathscr{F}^\gamma_{\alpha\beta} Z_\gamma)^a \frac{\partial}{\partial \mu^a}, \tag{3.12}$$

so the system (2.1) has the following consequences:

$$(Y_\alpha Z_\beta - Y_\beta Z_\alpha - \mathscr{F}^\gamma_{\alpha\beta} Z_\gamma)^a \frac{\partial \mathscr{L}}{\partial \mu^a} = 0. \tag{3.13}$$

Using the explicit form of Y_α and Z_α, one obtains that (3.13) gives something nontrivial in the following situations:

I. For $\alpha = (i;\mu)$ and $\beta = (j;\nu)$:

$$(U_i)^J{}_K C^{K,\nu}{}_j \frac{\partial \mathscr{L}}{\partial A^J{}_{,\mu}} - (U_j)^J{}_K C^{K,\mu}{}_i \frac{\partial \mathscr{L}}{\partial A^J{}_{,\nu}} = 0 \tag{3.14}$$

for all i,j,μ,ν.

II. For $\alpha = (i)$ and $\beta = (j;\mu,\nu)$:

$$(U_i)^J{}_K \left(C^{K,\mu}{}_j \frac{\partial \mathscr{L}}{\partial A^J{}_{,\nu}} + (\mu \leftrightarrow \nu) \right) = 0 \tag{3.15}$$

for all i,j,μ,ν.

III. For $\alpha = (i)$ and $\beta = (j;\mu)$:

$$[(U_i)^J{}_K C^{K,\mu}{}_j - \mathscr{F}^{\ k}_{ij} C^{J,\mu}{}_k] \frac{\partial \mathscr{L}}{\partial A^J} = 0 \tag{3.16}$$

for all i,j,μ.

An immediate consequence of (3.16) is that we have

$$U_i \neq 0. \tag{3.17}$$

Proof: If we suppose the contrary, i.e., $U_i = 0$ for all i, then (3.16) gives

$$\mathscr{F}^{\ k}_{ij} C^{J,\mu}{}_k \frac{\partial \mathscr{L}}{\partial A^J} = 0 \quad \text{for all } i,j,\mu.$$

Now we apply the following lemma.

Lemma 1: Given a simple n-dimensional Lie algebra with structure constants $\mathscr{F}^{\ k}_{ij}$ and v_k some n-dimensional vector so that $\mathscr{F}^{\ k}_{ij} v_k = 0$ holds for all i,j, then $v_k = 0$ for all k (see the Appendix for the proof).

We get

$$C^{J,\mu}{}_k \frac{\partial \mathscr{L}}{\partial A^J} = 0 \quad \text{for all } k \text{ and } \mu.$$

which inserted in (3.7) gives (remember that we supposed $U_i = 0$)

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}} (T_i)^A{}_B \chi^B \equiv 0 \quad \text{for all } i \text{ and } \mu,$$

or $\mathscr{T}^\mu_i \equiv 0$ for all i and μ, which is precisely what we want to avoid. ■

Another consequence of (3.16) can be obtained by using the following lemma.

Lemma 2: If $Z^J(\partial \mathscr{L}/\partial A^J) = 0$, then $U_i Z = 0$ for all i.

Proof: Denote $\mathscr{J} = \{J \mid Z^J \neq 0\}$. If $\mathscr{J} = \varnothing$ then $Z = 0$ and the proof is finished. If $\mathscr{J} \neq \varnothing$, then we introduce instead of A^J some LI combinations $\tilde{A}^J = \Sigma_{J'=1}^M \mathscr{D}^{JJ'} A^{J'}$ (det $\mathscr{D} \neq 0$) so that $A = \Sigma_{J \in \mathscr{J}} (1/Z^J) A^J$ is one of the $\tilde{A}$'s. It is easy to see that $Z^J(\partial \mathscr{L}/\partial A^J) = 0$ is equivalent to $\partial \mathscr{L}/\partial A = 0$, i.e., $\mathscr{L}$ does not depend on A. Let us differentiate (3.7) with respect to A. Keeping in mind that $\mathscr{L}$ does not depend on A, one gets

$$Z^J_i \frac{\partial \mathscr{L}}{\partial A^J{}_{,\mu}} = 0 \quad \text{for all } i \text{ and } \mu, \tag{*}$$

where $Z_i \equiv U_i Z$.

Analogously, differentiating (3.8) to A one gets

$$Z^J_i \frac{\partial \mathscr{L}}{\partial A^J} = 0 \quad \text{for all } i. \tag{**}$$

If we suppose that there is an i so that $Z_i \neq 0$, one can introduce $\mathscr{J}_i = \{J \mid Z^J_i \neq 0\}$ and $A_i = \Sigma_{J \in \mathscr{J}_i} (1/Z^J_i) A^J$. Then (*) and (**) tell us that $\partial \mathscr{L}/\partial A_{i,\mu} = 0$ for all μ and $\partial \mathscr{L}/\partial A_i = 0$, respectively. This contradicts the hypothesis a. ■

Let us decompose the space $\mathbb{R}^M$ in which U_i acts, in irreducible subspaces with respect to the representation U_i. (The existence of this decomposition is assured by the Weyl theorem, see Ref. 9, p. 428.)

$$\mathbb{R}^M = \bigoplus_q V_q, \quad U_i = \bigoplus_q U_i^{(q)} \quad \text{for all } i. \tag{3.18}$$

From now on we consider that (3.5) is given in this form. We denote $v^{(q)} \equiv v|_{V_q}$, where $v \in \mathbb{R}^M$. Now the relations $U_i Z = 0$ are equivalent to $U_i^{(q)} Z^{(q)} = 0$ for all q and i. We have two possibilities:

(1) $Z^{(q)} = 0$,

(2) $Z^{(q)} \neq 0$.

In this situation $U_i^{(q)} Z^{(q)} = 0$ tells us that $Z^{(q)}$ generates a one-dimensional linear subspace invariant under the irreducible representation $U^{(q)}$, so V_q is one dimensional, generated by $Z^{(q)}$ and $U^{(q)} = 0$.

Taking $Z = U_i C^\mu_j - \mathscr{F}^{\ k}_{ij} C^\mu_k$ we can apply Lemma 2 and the considerations that follow to (3.16). One obtains that we can have two types of indices q:

(1) $q \in Q$ for which

$$U_i^{(q)} C^{(q),\mu}{}_j = \mathscr{F}^{\ k}_{ij} C^{(q)\mu}{}_k, \quad i,j = 1,\ldots,n, \ \mu = 0,\ldots,d-1, \tag{3.19}$$

(2) $q \in Q'$ for which $\dim V_q = 1$ and

$$U_i^{(q)} = 0, \quad i = 1,\ldots,n. \tag{3.20}$$

We cannot have $Q = \varnothing$ because (3.17) would be contradicted. Now we can draw some conclusions: if $U_i^{(q)}$'s and $C^{(q)M}{}_i$'s which are the building blocks of U_i and C^μ_i from (3.5) do not fall into one out of the two categories from above then we get a contradiction and our problem does not have a solution. If we have for every q (3.19) or (3.20), the problem can have solutions and in the following we suppose that we are in this situation.

Let us see what are the relations (3.14)–(3.16) in this case. We use the decomposition (3.18) and (3.19) and get immediately that (3.14) and (3.15) are equivalent to

$$\mathscr{F}^{\ k}_{ij} \sum_{q \in Q} \left(C^{(q)J,\nu}{}_k \frac{\partial \mathscr{L}}{\partial A^{(q)J}{}_{,\mu}} + \mu \leftrightarrow \nu \right) = 0,$$
$$ij = 1,\ldots,n, \ \mu,\nu = 0,\ldots,d-1,$$

or, applying Lemma 1,

$$\sum_{q \in Q} \left(C^{(q)J,\nu}{}_k \frac{\partial \mathscr{L}}{\partial A^{(q)J}{}_{,\mu}} + \mu \leftrightarrow \nu \right) = 0,$$

$$k = 1,...,n,\ \mu,\nu = 0,...,d-1. \tag{3.21}$$

Combining (3.21) with (3.6) we get also

$$\sum_{q\in Q'}\left(C^{(q)J,\nu}{}_{k}\frac{\partial\mathscr{L}}{\partial A^{(q)J}{}_{,\mu}} + \mu\leftrightarrow\nu\right) = 0,$$

$$k = 1,...,n,\mu,\nu = 0,...,d-1. \tag{3.22}$$

The relation (3.16) gives (remember that dim $V_q = 1$)

$$\sum_{q\in Q'}\mathscr{F}^{\ k}_{ij}C^{(q)\mu}{}_{k}\frac{\partial\mathscr{L}}{\partial A^{(q)}} = 0, \quad i,j = 1,...,n,\ \mu = 0,...,d-1,$$

or applying Lemma 1,

$$\sum_{q\in Q'}C^{(q)\mu}{}_{k}\frac{\partial\mathscr{L}}{\partial A^{(q)}} = 0, \quad k = 1,...,n,\ \mu = 0,...,d-1. \tag{3.23}$$

Combining (3.23) with (3.7) we get

$$\frac{\partial\mathscr{L}}{\partial\chi^{A}_{,\mu}}(T_i)^{A}{}_{B}\chi^{B} + \sum_{q\in Q}\frac{\partial\mathscr{L}}{\partial A^{(q)J}{}_{,\mu}}(U_i^{(q)})^{J}{}_{K}A^{(q)K}$$
$$+\sum_{q\in Q}C^{(q)J,\mu}{}_{i}\frac{\partial\mathscr{L}}{\partial A^{(q)J}} = 0, \quad i = 1,...,n,\ \mu = 0,...,d-1. \tag{3.24}$$

Finally (3.8) becomes

$$\frac{\partial\mathscr{L}}{\partial\chi^{A}_{,\mu}}(T_i)^{A}{}_{B}\chi^{B}_{,\mu} + \frac{\partial\mathscr{L}}{\partial\chi^{A}}(T_i)^{A}{}_{B}\chi^{B} + \sum_{q\in Q}\frac{\partial\mathscr{L}}{\partial A^{(q)J}{}_{,\mu}}$$
$$\times(U_i^{(q)})^{J}{}_{K}A^{(q)K}{}_{,\mu} + \sum_{q\in Q}\frac{\partial\mathscr{L}}{\partial A^{(q)J}}(U_i^{(q)})^{J}{}_{K}A^{(q)K}{}_{,\mu} = 0,$$
$$i = 1,...,n. \tag{3.25}$$

Thus the system (3.6)–(3.8) is equivalent to (3.21)–(3.25). It is easy to see that the last one is complete.

We analyze now in detail the relations (3.19) and (3.20).

(1) We need the following lemma.

Lemma 3: Given $U_i, i = 1,...,n$ an irreducible representation of a simple Lie algebra $\mathscr{G}$ of dimension n, and n vectors C_i in the space of the representation, if the relation

$$U_iC_j = \mathscr{F}^{\ k}_{ij}C_k, \quad i,j = 1,...,n$$

holds, then we have the following two possibilities:

(a) C_i are LI,

(b) $C_i = 0$ for all i.

(See the Appendix for the proof.)

Applying this lemma to (3.19) one obtains that the indices μ can be split in to disjoint sets $\mathscr{A}$ and $\mathscr{B}$ so that if $\mu\in\mathscr{A}$, $C^{(q)\mu}{}_{i}$ are LI, dim $V_q = n$ (V_q cannot be larger because it would contradict the irreducibility of $U_i^{(q)}$), and if $\mu\in\mathscr{B}$, $C^{(q)\mu}{}_{i} = 0$ for all i. We cannot have $\mathscr{A}\neq\varnothing$ because in this case $A^{(q)}$ would have a transformation law of the first kind, contrary to the assumption b. If $\mu\in\mathscr{A}$, then in the basis $C^{(q)\mu}{}_{i}$, the matrices $U_i^{(q)}$ have the form

$$(U_i^{(q)})^{k}{}_{j} = \mathscr{F}^{\ k}_{ij}, \tag{3.26}$$

i.e., the adjoint representation.

If $\mu\neq\mu', \mu,\mu'\in\mathscr{A}$, then $U_i^{(q)}$ has the same form in the two bases $C^{(q)\mu}{}_{i}$ and $C^{(q)\mu'}{}_{i}$ so it commutes with the transformation matrix between the two bases. Because $U_i^{(q)}$ is irreducible, Schur's lemma tells us that the transformation matrix is proportional to the unit matrix in V_q, so we have in fact

$$C^{(q)\mu}{}_{i} = \lambda^{(q)\mu}C_i^{(q)}, \quad i = 1,...,n;\mu\in\mathscr{A}. \tag{3.27}$$

The vector $\lambda^{(q)}\in\mathbb{R}^d$ is different from zero (this is equivalent with $\mathscr{A}\neq\varnothing$) and $C_i^{(q)}, i = 1,...,n$ are LI and form a basis in V_q. The relation (3.5) gives for the fields $A^{(q)}$ the following transformation law:

$$\delta A^{(q)J} = \xi^{i}(U_i^{(q)})^{J}{}_{K}A^{(q)K} + C^{(q)J,\mu}{}_{i}\xi^{\ i}_{,\mu}.$$

Using (3.27) we get

$$\delta A^{(q)J} = \xi^{i}(U_i^{(q)})^{J}{}_{K}A^{(q)K} + \lambda^{(q)\mu}C_i^{(q)}\xi^{\ i}_{,\mu}. \tag{3.28}$$

Because $C_i^{(q)}$ are LI in V_q, the $n\times n$ matrix $C^{(q)}$ of elements $C^{(q)J}{}_{i}$ is invertible, i.e., $\exists\ C^{(q)-1}{}^{i}{}_{J}$ so that

$$C^{(q)-1}{}^{i}{}_{J}C^{(q)J}{}_{j} = \delta^{i}_{j}, \quad C^{(q)J}{}_{i}C^{(q)-1}{}^{i}{}_{K} = \delta^{J}_{K}. \tag{3.29}$$

We introduce instead of $A^{(q)J}$ some suitable LI combinations:

$$A^{(q)i} \equiv C^{(q)-1}{}^{i}{}_{J}A^{(q)J}. \tag{3.30}$$

Then the transformation law for $A^{(q)i}$ is

$$\delta A^{(q)k} = \xi^{i}\mathscr{F}^{\ k}_{ij}A^{(q)j} + \lambda^{(q)\mu}\xi^{\ k}_{,\mu}. \tag{3.31}$$

The vectors $\lambda^{(q)}\in\mathbb{R}^d$ are LI. [In fact, let us suppose that $\lambda^{(q)}$ are not LI. Then (3.27) tells us that $C^{(q)\mu}{}_{i}$ are not LI, which contradicts (b).] Choosing convenient new combinations $\tilde{A}^{(q)i} = \Sigma_{q'\in Q}\mathscr{D}^{qq'}A^{(q')i}$ (det $\mathscr{D}\neq 0$) we can arrange for the $\lambda^{(q)}$'s to be among the basis vectors of $\mathbb{R}^d$: $e_\mu, \mu = 0,...,d-1$, where $(e_\mu)^{\nu} = \delta^{\nu}_{\mu}$. Let us define the sets $\mathscr{C} = \{\mu|\exists\ q$ so that $\lambda^{(q)} = e_\mu\}$, and $\mathscr{C}' = \{0,...,d-1\}\setminus\mathscr{C}$. Then $\mathscr{C}\neq\varnothing$ is equivalent with $Q\neq\varnothing$ and we put by definition also

$$A^{\ i}_{\mu} \equiv A^{(q)i} \quad \text{if} \quad \lambda^{(q)} = e_\mu, \quad \mu\in\mathscr{C}. \tag{3.32}$$

Then, the transformation law (3.31) becomes

$$\delta A^{\ k}_{\mu} = \xi^{i}\mathscr{F}^{\ k}_{ij}A^{\ j}_{\mu} + \xi^{\ k}_{,\mu}, \quad \mu\in\mathscr{C},\ k = 1,...,n. \tag{3.33}$$

(2) For $q\in Q'$, (3.5) tells us (remember that dim $V_q = 1$)

$$\delta A^{(q)} = C^{(q)\mu}{}_{i}\xi^{\ i}_{,\mu}. \tag{3.34}$$

Now, the vectors $C^{(q)}\in\mathbb{R}^d\otimes\mathbb{R}^n$ of components $C^{(q)\mu}{}_{i}$ are LI, because otherwise (b) is contradicted. We can choose $C^{(q)}$ among the basis vectors of $\mathbb{R}^d\otimes\mathbb{R}^n$: $e_\mu\otimes E^{\ i}_{,\mu} = 0,...,d-1$, $i = 1,...,n$, where e_μ are defined at (1) and $E^{\ i}$ are given similarly by $(E^{\ i})_j = \delta^{i}_{j}$. Define the sets

$$\mathscr{D} = \{\mu|\exists\ C^{(q)}, \exists\ i \text{ so that } C^{(q)} = e_\mu\otimes E^{\ i}\},$$

and

$$\mathscr{D}_\mu = \{i|\exists\ C^{(q)} \text{ so that } C^{(q)} = e_\mu\otimes E^{\ i}\}.$$

If we denote

$$B^{\ i}_{\mu} \equiv A^{(q)} \quad \text{if} \quad C^{(q)} = e_\mu\otimes E^{\ i}\ (\mu\in\mathscr{D}, i\in\mathscr{D}_\mu), \tag{3.35}$$

then the transformation law of the B fields can be inferred from (3.34):

$$\delta B^{\ i}_{\mu} = \xi^{\ i}_{,\mu}, \quad \mu\in\mathscr{D},\ i\in\mathscr{D}_\mu. \tag{3.36}$$

We make the important observation that $\mathscr{C}\cap\mathscr{D} = \varnothing$. [In fact, if $\mu\in\mathscr{C}\cap\mathscr{D}$ then the fields $\tilde{A}^{\ i}_{\mu} \equiv A^{\ i}_{\mu} - B^{\ i}_{\mu}$ for $i\in\mathscr{D}_\mu$ are seen to have a first kind transformation law which contradicts (b).] So we have $\mathscr{D}\subseteq\mathscr{C}'$.

From now on we study separately the situation $d = 1$ and $d > 1$. The case $d = 1$ is studied for the sake of completeness.

A. $d = 1$. Because $\mathscr{C} \neq \varnothing$, we have $\mathscr{C}' = \varnothing$ and $\mathscr{D} = \varnothing$. Denote $A^i_0 \equiv A^i$, $A^i_{0,0} \equiv \dot{A}^i, \chi^A_{,0} \equiv \dot{\chi}^A$. Then (3.21), (3.24), and (3.25) become

$$\frac{\partial \mathscr{L}}{\partial \dot{A}_i} = 0, \tag{3.37}$$

$$\frac{\partial \mathscr{L}}{\partial \dot{\chi}^A}(T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial \dot{A}^k} \mathscr{F}^k_{ij} A^j + \frac{\partial \mathscr{L}}{\partial A^i} = 0, \tag{3.38}$$

$$\frac{\partial \mathscr{L}}{\partial \dot{\chi}^A}(T_i)^A{}_B \dot{\chi}^B + \frac{\partial \mathscr{L}}{\partial \chi^A}(T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial \dot{A}^k} \mathscr{F}^k_{ij} \dot{A}^j + \frac{\partial \mathscr{L}}{\partial A^k} \mathscr{F}^k_{ij} A^j = 0. \tag{3.39}$$

The relation (3.37) tells us that in fact $\mathscr{L} = \mathscr{L}(\chi^A, \dot{\chi}^A, A^i)$. We rewrite $\mathscr{L}$ as $\mathscr{L}'(\chi^A, \bar{\chi}^A, A^i)$ where

$$\bar{\chi}^A \equiv \dot{\chi}^A - A^i (T_i)^A{}_B \chi^B. \tag{3.40}$$

Then (3.38) tells us that $\partial \mathscr{L}'/\partial A^i = 0$ so that $\mathscr{L} = \mathscr{L}'(\chi^A, \bar{\chi}^A)$. After some computations (3.39) becomes

$$\frac{\partial \mathscr{L}'}{\partial \bar{\chi}^A}(T_i)^A{}_B \bar{\chi}^B + \frac{\partial \mathscr{L}'}{\partial \chi^A}(T_i)^A{}_B \chi^B = 0, \tag{3.41}$$

which is of the general form (2.1) with $X_\alpha{}^a$ of the form (2.8).

B. $d > 1$. After some computations, (3.21) and (3.24) can be brought to the form

$$\sum_{\rho \in \mathscr{C}} (e_\rho)^\mu \frac{\partial \mathscr{L}}{\partial A^i_{\rho,\nu}} + (\mu \leftrightarrow \nu) = 0, \quad i = 1,\ldots,n, \ \mu,\nu = 0,\ldots,d-1 \tag{3.42}$$

and

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}}(T_i)^A{}_B \chi^B + \sum_{\rho \in \mathscr{C}} \frac{\partial \mathscr{L}}{\partial A^k_{\rho,\mu}} \mathscr{F}^k_{ij} A^j_\rho + \sum_{\rho \in \mathscr{C}} (e_\rho)^\mu \frac{\partial \mathscr{L}}{\partial A^i_\rho} = 0,$$

$$i = 1,\ldots,n, \ \mu = 0,\ldots,d-1. \tag{3.43}$$

Choosing in (3.42) $\mu \in \mathscr{C}$, $\nu \in \mathscr{C}'$ we get $\partial \mathscr{L}/\partial A^i_{\mu,\nu} = 0$. Choosing in (3.43) $\mu \in \mathscr{C}'$ we get identically

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}}(T_i)^A{}_B \chi^B = 0, \quad \mu \in \mathscr{C}', \ i = 1,\ldots,n, \tag{3.44}$$

i.e.,

$$J^\mu_i \equiv 0, \quad i = 1,\ldots,n, \ \mu \in \mathscr{C}'. \tag{3.45}$$

Remember now that we have started to solve our problem to avoid the unusual situation $J^\mu_i \equiv 0$, $i = 1,\ldots,n$ and for *all* μ. We think that (3.45) is sufficiently unusual to restrict ourselves to the situation $\mathscr{C}' = \varnothing$. (This is the strong form of the current condition alluded to before.) An important by-product of this supposition is $\mathscr{D} = \varnothing$, so we get rid of the unwanted B-type fields. This is an important point so we remark that it will be desirable to find a stronger motivation for eliminating the B fields. Some propositions will be made a little later.

The conclusion is that our problem has nontrivial solutions if and only if the second type of fields are LI combinations of the A^i_μ, $i = 1,\ldots,n$, $\mu = 0,\ldots,d-1$. These are the famous Yang–Mills potentials. Remember that from condition (b) we have $dn \geqslant M$. Now we see that we have in fact $dn = M$. Because $\mathscr{C}' = \varnothing$, (3.33) becomes

$$\delta A^k_\mu = \xi^i \mathscr{F}^k_{ij} A^j_\mu + \xi^k_{,\mu} \quad k = 1,\ldots,n, \ \mu = 0,\ldots,d-1. \tag{3.46}$$

One must not forget to verify that (3.46) is the infinitesimal form of a representation of the gauge group $\tilde{G}$. If U_i is a representation of $\mathscr{G}$, then define the matrices

$$A_\mu \equiv A^i_\mu U_i, \quad \mu = 0,\ldots,d-1. \tag{3.47}$$

It is easy to verify[2] that (3.46) is the infinitesimal form of

$$(\mathscr{U}(\tilde{g})A_\mu)(x) = U(\tilde{g}(x))A_\mu(x)U(\tilde{g}(x))^{-1} + [\partial_\mu U(\tilde{g}(x))]\,U(\tilde{g}(x))^{-1}, \tag{3.48}$$

which is indeed a representation of $\tilde{G}$.

From now on the analysis of the case $d > 1$ is greatly simplified by the fact that the structure on the A-type fields is completely determined as a LI combination of the Yang–Mills potentials. The technique to be used is essentially that of Ref. 3, adapted to a situation a little more complicated.

Now (3.42) and (3.43) are

$$\frac{\partial \mathscr{L}}{\partial A^i_{\mu,\nu}} + (\mu \leftrightarrow \nu) = 0, \quad i = 1,\ldots,n; \ \mu, \ \nu = 0,\ldots,d-1, \tag{3.49}$$

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}}(T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial A^k_{\rho,\mu}} \mathscr{F}^k_{ij} A^j_\rho + \frac{\partial \mathscr{L}}{\partial A^i_\mu} = 0, \quad i = 1,\ldots,n, \ \mu = 0,\ldots,d-1. \tag{3.50}$$

Equation (3.25) becomes, after some computations,

$$\frac{\partial \mathscr{L}}{\partial \chi^A_{,\mu}}(T_i)^A{}_B \chi^B_{,\mu} + \frac{\partial \mathscr{L}}{\partial A^k_{\rho,\mu}} \mathscr{F}^k_{ij} A^j_{\rho,\mu} + \frac{\partial \mathscr{L}}{\partial \chi^A}(T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}}{\partial A^k_\mu} \mathscr{F}^k_{ij} A^j_\mu = 0, \quad i = 1,\ldots,n. \tag{3.51}$$

The system (3.49)–(3.51) can be simplified *à la* Utiyama. We define the covariant derivative

$$\chi^A_{;\mu} \equiv \chi^A_{,\mu} - A^i_\mu (T_i)^A{}_B \chi^B \tag{3.52}$$

and the Yang–Mills field

$$F^k_{\mu\nu} \equiv A^k_{\mu,\nu} - A^k_{\nu,\mu} + \mathscr{F}^k_{ij} A^i_\mu A^j_\nu, \quad k = 1,\ldots,n, \ \mu,\nu = 0,\ldots,d-1, \tag{3.53}$$

and we rewrite $\mathscr{L}$ as $\mathscr{L}'(\chi^A, \chi^A_{;\mu}, F^k_{\mu\nu}, A^k_{\{\mu,\nu\}}, A^k_\mu)$, where $A^k_{\{\mu,\nu\}} = A^k_{\mu,\nu} + (\mu \leftrightarrow \nu)$. Then (3.49) is equivalent to $\partial \mathscr{L}'/\partial A^i_{\{\mu,\nu\}} = 0$ and (3.50) to $\partial \mathscr{L}'/\partial A^i_\mu = 0$. So we have $\mathscr{L} = \mathscr{L}'(\chi^A, \chi^A_{;\mu}, F^k_{\mu\nu})$. After some calculation (3.51) gives

$$\frac{\partial \mathscr{L}'}{\partial \chi^A_{;\mu}}(T_i)^A{}_B \chi^B_{;\mu} + \frac{\partial \mathscr{L}'}{\partial \chi^A}(T_i)^A{}_B \chi^B + \frac{\partial \mathscr{L}'}{\partial F^k_{\mu\nu}} \mathscr{F}^k_{ij} F^j_{\mu\nu} = 0, \quad i = 1,\ldots,n. \tag{3.54}$$

This is our main result and now we make some comments.

(1) From (3.52) and (3.53) it is easy to obtain

$$\delta \chi^A_{;\mu} = \xi^i (T_i)^A{}_B \chi^B_{;\mu} \tag{3.55}$$

and

$$\delta F^k_{\mu\nu} = \xi^i \mathscr{F}^k_{ij} F^j_{\mu\nu}, \tag{3.56}$$

so the meaning of (3.54) is more transparent. It states that $\delta \mathscr{L}' = 0$.

(2) Let us introduce the vector

$$v^a = \begin{pmatrix} \chi^A \\ \chi^A_{,\mu} \\ F^k_{\mu\nu} \end{pmatrix}, \quad a = 1,...,(d+1)N + \frac{d(d-1)}{2}n.$$

The system (3.54) is of the general form (2.1) with the $X_\alpha{}^a$ of the form (2.8). The nonzero coefficients $Y_{(i)}{}^a{}_b$ are

$$Y_{(i)}{}^{\chi^A}{}_{\chi^B} = (T_i)^A{}_B,$$
$$Y_{(i)}{}^{\chi^A_{,\mu}}{}_{\chi^B_{,\nu}} = \delta^\nu_\mu (T_i)^A{}_B,$$
$$Y_{(i)}{}^{F^k_{\mu\nu}}{}_{F^j_{\rho\alpha}} = \tfrac{1}{2}(\delta^\rho_\mu \delta^\sigma_\nu - (\mu \leftrightarrow \nu))\mathscr{F}^k_{ij}.$$

It is easy to verify that

$$[Y_{(i)}, Y_{(j)}] = \mathscr{F}^k_{ij} Y_{(k)}, \tag{3.57}$$

so the system (3.56) is complete, in accordance with the completeness of the system (3.21)–(3.25). The equations of (3.54) are LI. (In fact suppose the contrary; then there are $Z^i, i = 1,...,n$ not all of them zero so that $Z^i \mathscr{F}^k_{ij} = 0$, i.e., the adjoint representation of G is not faithful which contradicts the simplicity of G.) Because $(d+1)N + \frac{1}{2}d(d-1)n > dn$ for $d > 1$, (3.54) has a nontrivial solution. Moreover, (3.54) is of the same type as (2.7), so by solving the system (2.1) and (2.8) one solves the global invariance problem and the local invariance problem as well. This problem is partially solved (see for instance Ref. 10 and references cited there).

(3) Let us suppose that the theory is also manifestly Lorentz invariant ($d = 4$). The no-go theorem[11] tells us that the theory is invariant under the direct product $G \times \mathscr{L}$. By Proposition III, Sec. 2.7, p. 56 of Ref. 9 we know that the irreducible representations of $G_1 \times G_2$ are a tensorial product of irreducible representations of G_1 and G_2, respectively. Because the index i of A^i_μ is carrying an irreducible representation of G (remember that G is simple so the adjoint representation is irreducible) the index μ must carry a representation of $\mathscr{L}$. This tells us that A^i_μ is a cuadrivector with respect to μ.

Let us note that the Lorentz invariance is a good argument for excluding the B-type fields for otherwise the manifest Lorentz covariance would conflict with the gauge transformation (3.46).

(4) The generalization to a semisimple group proceeds as follows. The arguments leading to (3.19)–(3.25) stay as they are because the Weyl theorem is valid for semisimple groups also. Now choosing convenient C_i's we can split (3.19)–(3.25) in a number of relations of the same structure, one set for every simple component of $\mathscr{G}$. The analysis given above applies for every simple component producing a Yang–Mills potential for each one. These can be recombined into a Yang–Mills potential for the whole algebra $\mathscr{G}$. From now on the analysis goes through unchanged.

(5) For an abelian group the result remains true but the proof is altered as follows. In (3.14)–(3.16) one must make $\mathscr{F}^k_{ij} = 0$. An important consequence is that (3.17) is no longer true so we are not prevented from having $U_i = 0, i = 1,...,n$. The relation (3.16) becomes

$$(U_i)^J{}_K C^{k,\mu}{}_j \frac{\partial \mathscr{L}}{\partial A^J} = 0, \quad i,j = 1,...,n, \ \mu = 0,...,d-1. \tag{3.58}$$

In analogy with the nonabelian case, we use Lemma 2 with $Z = U_i C^\mu_j$ and get

$$U_k U_i C^\mu_j = 0, \quad i,j,k = 1,...,n, \ \mu = 0,...,d-1. \tag{3.59}$$

Now we use the important result that the decomposition (3.18) remains true for abelian groups and, moreover, $\dim V_q = 1$. Equation (3.59) is then equivalent to

$$U^{(q)}_k U^{(q)}_i C^{(q)\mu}_j = 0, \quad i,j,k = 1,...,n, \ \mu = 0,...,d-1. \tag{3.60}$$

for all q. Let us prove that $U^{(q)}_i = 0, i = 1,...,n$ for all q. Suppose there is an index q so that $U^{(q)} \neq 0$. Because $U^{(q)}_i$ are numbers we get from (3.60),

$$C^{(q)\mu}{}_j = 0, \quad j = 1,...,n, \ \mu = 0,...,d-1,$$

i.e., $A^{(q)}$ has a first kind transformation law, which contradicts (b). We have proved that

$$U_i = 0, \quad i = 1,...,n. \tag{3.61}$$

In conclusion, if $U_i \neq 0$ our problem does not have solutions. In the following we suppose that (3.61) is true. The relations (3.14)–(3.16) become identities so the system is complete. The relation (3.5) is in this situation

$$\delta A^{(q)} = C^{(q)\mu}{}_i \xi^i{}_{,\mu}. \tag{3.62}$$

The analysis following relation (3.34) can be repeated with the result that $A^{(q)}$ are LI combinations of the Yang–Mills potential A^i_μ:

$$A^{(q)} \equiv A^i_\mu \quad \text{if} \quad \mu \in \mathscr{D}, \ i \in \mathscr{D}_\mu \quad \text{and} \quad C^{(q)} = e_\mu \otimes E^i. \tag{3.63}$$

The transformation law of A^i_μ is analogous to (3.36):

$$\delta A^i_\mu = \xi^i{}_{,\mu}, \quad \mu \in \mathscr{D}, \ i \in \mathscr{D}_\mu, \tag{3.64}$$

which is the infinitesimal form of the finite transformation

$$(\mathscr{U}(\tilde{g})A^i_\mu)(x) = A^i_\mu(x) + g^i(x)_{,\mu} \quad (g^i: \mathbb{R}^d \to \mathbb{R}). \tag{3.65}$$

[G is isomorphic with $\mathbb{R}^d$ and endowed with the additive composition law, so $\tilde{G} = \{g: \mathbb{R}^d \to \mathbb{R}^n | g \in \mathscr{C}^\infty(\mathbb{R}^d)\}$ is also endowed with the additive composition law.]

Now we can show that $\mathscr{D}_\mu = \{1,...,n\}$ and $\mathscr{D} = \{0,...,d-1\}$ analogously with the nonabelian case (in the opposite case, i.e., $\{1,...,n\} \setminus \mathscr{D}_\mu \neq \phi$ or $\{0,...,d-1\} \setminus \mathscr{D} \neq \phi$, we contradict the current condition).

From now on the analysis goes through unchanged (a minor modification is to put $\mathscr{F}^k_{ij} = 0$ everywhere).

It is easy to generalize now to $G = A \times S$ where A is abelian and S is semisimple. We have relaxed the condition as we have promised.

4. THE GROUP CONDITION

We know from the general theory of Lie groups (see for instance Ref. 12) that for a nonabelian Lie group, the commutator of two group transformations of parameters ξ^i_1 and ξ^j_2 must be a group transformation also of parameters $\mathscr{F}^k_{ij}\xi^i_1\xi^j_2$. We must expect that something analogous take place for gauge groups also (the group condition). Berg-

mann[13] has introduced this condition in the framework of general relativity. For gauge groups this was done by Schwinger[14] who supposed from the beginning that the A-type fields are LI combinations of fields A^i_μ, $i=1,...,n$, $\mu=0,...,d-1$ with the transformation law

$$A^k_\mu \to A'^k_\mu = A^k_\mu + \delta A^k_\mu, \quad \delta A^k_\mu = \xi^i (t_i)^k{}_j A^j_\mu + \xi^k{}_{,\mu},$$

and proved that $(t_i)^k{}_j = \mathscr{F}^k{}_{ij}$. His argument can be generalized as follows. Let us compute the commutator of two transformations of the form (3.5) of parameters $\xi^i{}_1$ and $\xi^j{}_2$. We get

$$\begin{aligned}(\delta_1\delta_2 - \delta_2\delta_1)A^J &= \mathscr{F}^k{}_{ij}\xi^i{}_1\xi^j{}_2(U_k)^J{}_K A^k \\ &\quad + C^{J,\mu}_k(\mathscr{F}^k{}_{ij}\xi^i{}_1\xi^j{}_2)_{,\mu} + (U_i C^\mu_j - \mathscr{F}^k{}_{ij}C^\mu_k)^J \\ &\quad \times(\xi^i{}_1\xi^j{}_2)_{,\mu} - (U_i C^\mu_j + U_j C^\mu_i)^J \xi^i{}_1\xi^j{}_{2,\mu}. \end{aligned} \tag{4.1}$$

The group condition can be fulfilled by choosing

$$U_i C^\mu_j = \mathscr{F}^k{}_{ij} C^\mu_k, \quad i,j=1,...,n,\ \mu=0,...,d-1 \tag{4.2}$$

[the last term from (4.1) is then automatically zero].

If (4.2) holds we can apply Lemma 3 and the analysis of Sec. 3 is greatly simplified. In the first place we get rid of the B-type fields without using the current condition. Secondly, we obtain that A^J are LI combinations of the Yang–Mills fields without using the invariance of the Lagrangian under (3.4) and (3.5). In this way we arrive directly at the system (3.49) and (3.51) which after introducing $\chi^A_{;\mu}$ and $F^k_{\mu\nu}$ is equivalent to the complete system (3.54).

This analysis seems to indicate that most of the structure of a local invariant theory can be obtained by a more careful analysis of the principal mathematical object—the gauge group. This analysis might require a more sophisticated mathematical apparatus, so we think that our elementary approach has the advantage of simplicity.

5. CONCLUSIONS

We have rederived in a more systematic fashion the result of Utiyama. Our proof disentangles the role of local gauge invariance and a space-time invariance (like Lorentz invariance) in the construction of the Yang–Mills theories so it can be used for constructing theories with other kinds of space-time symmetries. The "weak" point of our argument is the argument leading to the elimination of the B-type fields. We feel that this can be done by exploiting completely the group condition as indicated in Sec. 4. We think that it can be proved rigorously that there are only two types of representations of a gauge group: the "normal" type as for the χ fields [see condition (c) in Sec. 3] and the Yang–Mills type, i.e., (3.48). Of course the argument based on Lorentz invariance is solid enough but the point is to use only the gauge invariance.

The method can be used in other situations like gravity, supergravity, and Yang–Mills supersymmetry although there are some problems. For instance in Kibble's approach to gravity,[15] we explicitly checked that "the algebra does not close itself," i.e., the group condition is not fulfilled. In the last two cases the computation must take into account the appearance of anticommuting c-numbers.

ACKNOWLEDGMENT

The author wishes to thank Dr. V. Georgescu for reading carefully the manuscript and making many valuable suggestions.

APPENDIX

We want to prove Lemmas 1 and 3 from Sec. 3. It is convenient to recast them first into a basis independent form.

Lemma 1: Given $\mathscr{L}$, a simple Lie algebra, and $t:\mathscr{L}\to V$ a linear map so that $t([x,y])=0$ holds for every x, y from $\mathscr{L}$, then $t=0$.

Proof: Define $[\mathscr{L},\mathscr{L}] = \{z\in\mathscr{L} \mid \exists\ x,y\in\mathscr{L}$ so that $[x,y]=z\}$. It is easy to prove that this linear subspace of $\mathscr{L}$ is an ideal of $\mathscr{L}$, so from the simplicity of $\mathscr{L}$ we have $[\mathscr{L},\mathscr{L}]=\mathscr{L}$. So $\forall\ z\in\mathscr{L}, \exists\ x,y\in\mathscr{L}$ so that $z=[x,y]$. Then $t(z)=t([x,y])=0$, i.e., $t=0$. ■.

Lemma 2: Given $\mathscr{L}$ and t as before and ρ a representation of $\mathscr{L}$ in V so that $\rho(x)t(y)=t([x,y])$ holds for every x,y from $\mathscr{L}$, then t is either injective or $t=0$.

Proof: Define the linear subspace $\mathscr{L}'$ of $\mathscr{L}$; $\mathscr{L}' = \{x\in\mathscr{L} \mid t(x)=0\}$. Given $x\in\mathscr{L}$, $y\in\mathscr{L}'$ we have $t([x,y])=\rho(x)t(y)=0$ so $[x,y]\in\mathscr{L}'$. Then $\mathscr{L}'$ is an ideal of $\mathscr{L}$. From the simplicity of $\mathscr{L}$ we have $\mathscr{L}'=\mathscr{L}$ or $\mathscr{L}'=\{0\}$. ■.

[1] C. N. Yang and R. Mills, Phys. Rev. **96**, 191 (1954).
[2] E. Abers and B. W. Lee, Phys. Rep. **9**, 1 (1973).
[3] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[4] A. C. Davies, CERN Preprint TH. 3213, 1981.
[5] S. A. Bludman, Phys. Rev. **100**, 372 (1955).
[6] W. Pauli, Rev. Mod. Phys. **13**, 203 (1941).
[7] G. Chilov, Analyse mathématique, Fonctions de plusieurs variables réelles, (Éditions Mir, Moscow, 1975).
[8] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1957).
[9] M. Naïmark and A. Stern, Théorie des représentations des groupes, (Éditions Mir, Moscow, 1979).
[10] D. Krupka, Czech. J. Phys. B **29**, 300 (1979).
[11] L. O'Raifeartaigh, Phys. Rev. **139**, B 1052 (1965).
[12] G. Racah, "Lectures on Lie groups," in *Lectures of the Istanbul Summer School of Theoretical Physics*, edited by F. Gürsey. (Colorado Associated, Boulder, CO, 1964).
[13] P. Bergman, Phys. Rev. **75**, 680 (1949).
[14] J. Schwinger, Phys. Rev. **125**, 1043 (1962).
[15] T. W. B. Kibble, J. Math. Phys. **2**, 212 (1960).

Static solutions of the coupled Yang–Mills–Weyl equations

M. Magg
Institut für Theoretische Physik, RWTH Aachen, 51 Aachen, Federal Republic of Germany

(Received 26 April 1983; accepted for publication 2 September 1983)

It is shown that static, spherically symmetric solutions of the massless Yang–Mills–Dirac equation which carry finite energy are trivial.

PACS numbers: 11.15.Kc, 03.50.Kk

Static classical solutions with finite energy have led to exciting discoveries in the corresponding quantum field theory. A well-known example in particle physics is the magnetic monopole solution[1] and its fermionic fluctuation modes in nonabelian gauge theories.[2] It revealed the existence of magnetic monopoles as particles with half-integer fermion number in a number of gauge theories.

In this example one ignores the feedback of the spinor field onto the gauge fields because it is suppressed by the gauge coupling, which is assumed to be small in order to guarantee the validity of the semiclassical analysis. For the zero-energy modes this feedback happens to vanish exactly.

It has been argued that in cases where more fermion representations of the same type couple to the gauge field, the influence of the spinor fields upon the gauge field could become important and might give rise to nonperturbative solutions even in the absence of fundamental scalar fields.[3,4] It seems, therefore, worthwhile to investigate the possibility of static solutions with finite energy in the fully coupled system of Yang–Mills and Dirac equations. Some results are found in Refs. 5 and 6.

If fermion masses can be neglected, one can expect to ease this hard mathematical problem considerably. Scale invariance now becomes a strong tool which has been successfully applied to the same problem in classical field theories without spinors.[7] However, spinor fields present an obstacle for this way of reasoning because their classical field energy is an indefinite functional.

To overcome this difficulty we have had to restrict ourselves to spherically symmetric configurations. Under these assumptions we prove that there are no static solutions of the coupled Yang–Mills–Dirac equations which carry finite energy.

We study SU(2) Yang–Mills (YM) fields minimally coupled to a bispinor field with the Lagranian density

$$\mathscr{L} = \tfrac{1}{4}F_a^{\mu\nu}F_{a\mu\nu} + i\,\mathrm{tr}\{\chi^+\gamma^0\gamma^\mu(\partial_\mu + A_\mu)\chi\}. \tag{1}$$

The components A_a^μ of the gauge potential A^μ are given with respect to a standard basis of SU(2),

$$[T_a, T_b] = \epsilon_{abc}T_c, \quad 1\leqslant a,b,c\leqslant 3.$$

The spinor field χ is compactly written as a 4×2 matrix. The internal algebra acts on it as follows:

$$(T_a\chi)_{\alpha\beta} = -(i/2)(\chi\tilde{\sigma}_a)_{\alpha\beta}, \quad 1\leqslant\alpha\leqslant 4, \quad 1\leqslant\beta\leqslant 2.$$

The tilde indicates matrix transposition. The field strength tensor has the form

$$F_a^{\mu\nu} = \partial^\mu A_a^{\ \nu} - \partial^\nu A_a^{\ \mu} + \epsilon_{abc}A_b^{\ \mu}A_c^{\ \nu}.$$

The Dirac matrices are chosen in the chiral representation

$$\gamma^0 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \quad \gamma^0\gamma^i = \begin{pmatrix} \sigma^i & 0 \\ 0 & \sigma^i \end{pmatrix},$$

and the bispinor is given by its chiral components

$$\chi = \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix}, \quad \chi_\pm = 2\times 2 \text{ matrices}.$$

The Dirac equation which follows from (1) becomes actually a pair of Weyl equations

$$(\partial_0 + A_0)\chi_\pm = \mp\sigma^i(\partial_i + A_i)\chi_\pm. \tag{2}$$

The YM fields satisfy

$$\begin{aligned} \partial_j F_a^{j0} + \epsilon_{abc}A_{bj}F_c^{j0} &= i\,\mathrm{tr}\{\chi_+^+ T_a\chi_+\} \\ &\quad + i\,\mathrm{tr}\{\chi_-^+ T_a\chi_-\}, \\ \partial_\mu F^{\mu i} + \epsilon_{abc}A_{b\mu}F^{\mu i} &= i\,\mathrm{tr}\{\chi_+^+\sigma^i T_a\chi_+\} \\ &\quad - i\,\mathrm{tr}\{\chi_-^+\sigma^i T_a\chi_-\}. \end{aligned} \tag{3}$$

We have not bothered to write an explicit gauge coupling parameter in (1). In classical physics it can be absorbed by the normalization of the gauge potential and the spinor field.

A static solution of (2) and (3) is one where all gauge-invariant quantities are time independent. There is then a gauge with time-independent gauge potentials and the spinor field has at most a stationary time dependence. As we announced at the beginning we concentrate on spherically symmetric solutions. Again this is understood in the sense that all gauge-invariant fields show this symmetry. It can be shown that the whole set of symmetric configurations is represented by the ansatz[2,8]

$$A_a^0(\mathbf{x}) = \hat{x}^a\frac{H(r)}{r}, \tag{4}$$

$$A_a^i(\mathbf{x}) = -\epsilon_{aij}\hat{x}^j\frac{1-K(r)}{r} + \hat{x}^a\hat{x}^i\frac{M(r)}{r} + (\delta^{ai} - \hat{x}^a\hat{x}^i)\frac{N(r)}{r}, \tag{5}$$

$$\chi_\pm(\mathbf{x}, t) = e^{iEt}\cdot r^{-1}[g_\pm(r)\mathbb{1} + p_\pm(r)\hat{x}\cdot\boldsymbol{\sigma}]. \tag{6}$$

The additional case

$$A_a^i = 0, \quad A_a^0(\mathbf{x}) = \phi_a(r),$$

which appears in the classification of spherically symmetric pure-gauge fields, spoils this symmetry in the presence of spinor fields which accentuate the radial direction in internal space. The ansatz (4)–(6) still admits some gauge freedom. A spherically symmetric gauge function

$$g(\mathbf{x}) = \mathbb{1}\cos\Lambda(r) + i\hat{x}\cdot\boldsymbol{\sigma}\sin\Lambda(r),$$

transforms the radial functions in (5) and (6) without chang-

 0022-2488/84/051539-03$02.50

ing the general form of the ansatz

$$\begin{aligned} &K \to K\cos 2\Lambda + N\sin 2\Lambda,\\ &N \to -K\sin 2\Lambda + N\cos 2\Lambda,\\ &M \to M + 2\Lambda',\\ &g_\pm \to g_\pm \cos\Lambda - ip_\pm \sin\Lambda,\\ &p_\pm \to -ig_\pm \sin\Lambda + p_\pm \cos\Lambda. \end{aligned} \tag{7}$$

This is used to choose the gauge condition

$$N = 0. \tag{8}$$

Let us now recall the standard argument which forbids static solutions with finite energy in scale-invariant theories without spinor fields. In scale-invariant classical field theories there is a locally conserved current

$$\partial_\mu(x_\nu \theta^{\mu\nu}) = \theta^\mu{}_\mu = 0, \tag{9}$$

where $\theta^{\mu\nu}$ denotes the symmetrical gauge-invariant energy-momentum tensor. The resulting time independence of

$$\int dx^3\{t\theta^{00}(\mathbf{x}) + x_i\theta^{0i}(\mathbf{x})\} \tag{10}$$

can only hold if the energy vanishes,

$$\int dx^3\,\theta^{00}(\mathbf{x}) = 0, \tag{11}$$

provided the integrals in (10) converge. If the energy integral is positive definite this could hold only for the trivial solution.

In the theory we are considering, the energy-momentum tensor

$$\theta^{\mu\nu} = \theta^{\mu\nu}_{\rm YM} + \theta^{\mu\nu}_s, \tag{12}$$

$$\theta^{\mu\nu}_{\rm YM} = \tfrac{1}{4} g^{\mu\nu} F_{a\rho\sigma}F_a{}^{\rho\sigma} + F^{\mu\rho}_a F_{a\rho}{}^\nu, \tag{13}$$

$$\theta^{\mu\nu}_s = -\tfrac{1}{2}\mathrm{tr}\{\chi^+\gamma^0[\gamma^\mu(\partial^\nu + A^\nu) + \gamma^\nu(\gamma^\mu + A^\mu)]\chi\} \tag{14}$$

contains the indefinite part of the energy density

$$\theta^{00}_s = E\,\mathrm{tr}\{\chi^+\chi\} - iA^0_a\,\mathrm{tr}\{\chi^+ T_a\chi\} \tag{15}$$

due to the spinor fields. This forces us to go beyond mere symmetry arguments and have a look at the field equations.

Note that the energy density of the YM field,

$$\theta^{00}_{\rm YM} = \tfrac{1}{2}(F^{0i}_a)^2 + \tfrac{1}{4}(F^{ij}_a)^2, \tag{16}$$

is strictly positive for all gauge fields inequivalent to zero. For physical reasons we are interested only in solutions where the YM fields carry finite energy

$$\int dx^3\,\theta^{00}_{\rm YM}(\mathbf{x}) < \infty. \tag{17}$$

Together with the contribution of the interacting spinor field it adds up to zero.

It is straightforward to insert the ansatz (4)–(6) into the field equations (2)–(3). The Weyl equations become

$$[\mp H + 2ird/dr + 2iK]p_\pm = -(M \mp 2Er)g_\pm, \tag{18}$$

$$[\mp H + 2ird/dr - 2iK]g_\pm = -(M \mp 2Er)p_\pm, \tag{19}$$

and the YM equations reduce to

$$\begin{aligned} &-r^2K'' + K(K^2 + M^2 - H^2 - 1)\\ &\quad = -2ir[g_+p^*_+ - g_+p^*_+ - g_-p^*_- + g^*_-p_-], \end{aligned} \tag{20}$$

$$\begin{aligned} &r(K^2M)' - K^2M\\ &\quad = -rK[|g_+|^2 - |p_+|^2 - |g_-|^2 + |p_-|^2], \end{aligned} \tag{21}$$

$$2K^2M = -r[|g_+|^2 + |p_+|^2 - |g_-|^2 - |p_-|^2], \tag{22}$$

$$\begin{aligned} &r^2H'' - 2K^2H\\ &\quad = 2r\,[g_+p^*_+ + g^*_+p_+ + g_-p^*_- + g^*_-p_-]. \end{aligned} \tag{23}$$

At first glance the combined system of 12 real equations for 11 real radial functions seems to be overdetermined. However, Eq. (21) can be dropped since it follows from (22) and the Weyl equations (18) and (19). This redundancy reflects the gauge freedom which allowed for the gauge condition (8). Equation (22) eliminates the function M and one is left with 10 differential equations for 10 functions.

For any solution of the system (18)–(23) the substitution

$$\begin{aligned} &g_\pm \to g^*_\mp, \quad p_\pm \to p^*_\mp,\\ &K \to K, \quad M \to -M, \quad H \to H \end{aligned} \tag{24}$$

defines a new solution, replacing the eigenvalue E by $-E$. The YM energy

$$\begin{aligned} &4\pi\int_0^\infty dr\,r^{-2}[(KH)^2 + \tfrac{1}{2}(rH^{-1} - M)^2 + (rk^{-1})^2\\ &\quad + (KM)^2 + \tfrac{1}{2}(1-K)^2(1+K)^2] \end{aligned} \tag{25}$$

remains invariant under this transformation.

Using Eqs. (18) and (19) as well as their complex conjugate counterparts, one finds two first integrals of the system

$$g_\pm p^*_\pm + g^*_\pm p_\pm = C_\pm = \text{const.} \tag{26}$$

The YM energy (25) can be finite only if

$$C := C_+ + C_- = 0. \tag{27}$$

This follows from Eq. (23) which reads now

$$r^2H'' - 2K^2H = 2rC. \tag{28}$$

For the first derivative H' one gets

$$H'(r) = H'(r_0) + 2C\ln\frac{r}{r_0} + 2\int_{r_0}^r dr'\,r'^{-2}K^2H. \tag{29}$$

Convergence of the expression (25) requires that the integral in (29) converges for $r \to \infty$. This can be seen using the Cauchy–Schwarz inequality

$$\begin{aligned} \int_{r_0}^\infty dr\,r^{-2}K^2|H| \leqslant &\left(\int_{r_0}^\infty dr\,r^{-2}K^2H^2\right)^{1/2}\\ &\times\left(\int_{r_0}^\infty dr\,r^{-2}K^2\right)^{1/2}, \end{aligned}$$

combined with the estimate

$$\int_{r_0}^\infty dr\,r^{-2}K^2 \leqslant 2r_0^{-1}K^2(r_0) + 4\int_{r_0}^\infty dr\,K'^2.$$

With this observation one finds

$$H(r) - rH'(r) = -2Cr + o(r) \quad (r \to \infty). \tag{30}$$

This asymptotic behavior leads to a divergent YM energy

unless the constant C vanishes. The YM charge density vanishes because it is proportional to C,

$$i\,\mathrm{tr}\{\chi^+ T_a \chi\} = -2\hat{x}_a r^{-2} \cdot C.$$

The energy density (15) has now no explicit interaction term

$$\theta_s^{00} = E\,\mathrm{tr}\{\chi^+\chi\}. \tag{31}$$

For a solution with $E \geqslant 0$, Eq. (11) requires

$$\theta_{\mathrm{YM}}^{00} = 0. \tag{32}$$

The gauge field is then equivalent to zero and the spinor field is a free zero-energy mode. This means the spinor field is constant in the $A^\mu = 0$ gauge:

$$\chi_\pm(\mathbf{x}) = \epsilon_\pm \mathbf{1}, \quad |\epsilon_+| = |\epsilon_-|, \qquad A^\mu = 0. \tag{33}$$

The contributions which come from opposite chirality cancel each other in the YM current.

A solution with $E < 0$ can be transformed into one with $E > 0$ by using (24). Since this substitution does not change $\theta_{\mathrm{YM}}^{00}$, the new solution has to be trivial and so is the original one.

The result may be summarized as follows: all static, spherically symmetric solutions of the massless Yang–Mills–Dirac equations which carry finite energy are represented by (33). The whole analysis applies also to the more general situation where n chirality $+$ and m chirality $-$ isodoublets couple to SU(2) gauge fields. The 4-current on the right-hand side of the YM equations (3) represents now the contributions of all $n + m$ Weyl spinor fields. If one takes instead of SU(2) a gauge group of rank bigger than 1, then our analysis gets more complicated due to the increasing complexity of spherically symmetric fields.

[1] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974); A. M. Polyakov, JETP Lett. **20**, 194 (1974).
[2] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).
[3] R. Friedberg and T. D. Lee, Phys. Rev. D **13**, 3398 (1976).
[4] R. Akhoury and W. I. Weisberger, Phys. Rev. D **22** 959 (1980).
[5] R. Akhoury and W. I. Weisberger, Nucl. Phys. B **174**, 225 (1980).
[6] M. Basler, "Self-consistent spherically symmetric solutions of the Yang–Mills–Dirac equations," Preprint University Jena GDR (1983).
[7] S. Coleman, in *Proceedings of the 1975 International School of Subnuclear Physics*, edited by A. Zichichi, (Plenum, New York, 1977).
[8] E. Witten, Phys. Rev. Lett. **38**, 121 (1977); R. Jackiw, "Lecture at the Schladming Winter School 1980," Acta Phys. Austriaca Suppl. **XXII**, 383 (1980); Gu Chaohao, Phys. Rep. **80**, 251 (1981).

SO(4) reduction of the Yang–Mills equations for the classical gauge groups

M. Légaré and J. Harnad
Centre de recherche de mathématiques appliquées, Université de Montréal, C. P. 6128, Succursale A, Montréal, H3C 3J7, Canada

(Received 25 July 1983; accepted for publication 16 September 1983)

We show that the reduction of the SU(n), SO(n), and Sp(n) Yang–Mills equations in compactified Minkowski space by certain realizations of the SO(4) subgroup of the conformal group leads to systems of ordinary differential equations. The reduced systems are interpretable as Hamiltonian systems with symmetry constrained so that the momentum map equals zero. Explicit solutions for these systems are given.

PACS numbers: 11.15.Kc, 11.30. − j, 02.40. + m

1. INTRODUCTION

The geometric interpretation of invariant gauge fields under smooth group actions, studied by Harnad, Shnider, and Vinet[1] (hereafter noted HSV), has been used to derive the dimensional reduction procedure,[2] to solve pure-gauge[3] and matter-coupled-gauge systems,[4,5] and to determine invariant spinor fields with gauge freedom.[6] In this paper, we pursue this line of study with the investigation of solutions to the SU(n), SO(n), and Sp(n) Yang–Mills equations invariant under the SO(4) subgroup of the conformal group of space-time C(3,1). Such SO(4) invariant solutions have already been obtained for the SU(2) Yang–Mills equations by Luscher[7] and Schechter[8] and for Dirac spinors coupled to SU(2) gauge fields by Meetz[9] and Doneux, Saint-Aubin, and Vinet.[5]

The content of this paper is as follows. In Sec. 2, we present the calculation of SO(4) invariant Yang–Mills fields under certain realizations of the SO(4) subgroup $\subset$C(3,1) acting simultaneously on tensorial and gauge indices. These realizations derive from embeddings of the isotropy subgroup SU(2)$\subset$SO(4) in the classical groups. In Sec. 3, we derive the reduced Yang–Mills equations and Lagrangian density. It turns out that the reduced equations expressed in a convenient gauge are interpretable as Hamiltonian systems with large symmetry groups constrained by the condition that the solution be contained in the kernel of the associated momentum map. The residual symmetry allows us to integrate the system explicitly by reducing it to a set of decoupled anharmonic oscillators equivalent to the reduced systems for the SU(2) gauge group. In the summary, further possible extensions of this work are discussed.

2. SO(4) INVARIANT GAUGE FIELDS

We shall use in the following many of the conventions and notations of Refs. 1–3 which, for completeness, are summarized here. Let $\bar{M}\sim S^1\times S^3/Z_2$ be the conformally compactified Minkowski space, which is also identified with the group manifold U(2). For simplicity, we work on the twofold covering $S^1\times S^3$, which is in any case diffeomorphic to $\bar{M}$ and permits an identification with U(1)$\times$SU(2), with points designated as $p=(e^{i\psi},v)$ where $e^{i\psi}\in$U(1) and $v\in$SU(2). The relation to Cartesian coordinates is given by

$$x^\mu=\eta^\mu/(\eta^4+\eta^5)\quad(\mu=0,1,2,3) \tag{2.1}$$

with the introduction of a coordinate set $\{\eta^\mu,\eta^4,\eta^5\}$ corresponding to $e^{i\psi}=\eta^5+i\eta^0$ and $v=\eta^4-i\eta^i\sigma_i$ ($i=1,2,3$), where $\{\sigma_i\}$ are the Pauli matrices. We also define natural group actions on U(1)$\times$SU(2):

(a) left action of SU(2):

$$L_g\colon (e^{i\psi},v)\to(e^{i\psi},gv),\quad \text{where } g\in\text{SU(2)}; \tag{2.2}$$

(b) right action of SU(2):

$$R_g\colon (e^{i\psi},v)\to(e^{i\psi},vg),\quad \text{where } g\in\text{SU(2)}; \tag{2.3}$$

(c) right and left actions of U(1):

$$L_\phi(e^{i\psi},v)=R_\phi(e^{i\psi},v)=(e^{i(\psi+\phi)},v),\quad \text{where } e^{i\phi}\in\text{U(1)}; \tag{2.4}$$

(d) left action of SU(2)$_L\times$SU(2)$_R$:

$$L_{(g',g)}\colon (e^{i\psi},v)\to(e^{i\psi},g'vg^{-1}),\ \text{where } g,g'\in\text{SU(2)}; \tag{2.5}$$

(e) diagonal SU(2) subgroup action of SU(2)$_D\equiv$(SU(2)$_L\times$SU(2)$_R$)$_D$:

$$D_g\colon (e^{i\psi},v)\to(e^{i\psi},gvg^{-1}). \tag{2.6}$$

With the standard identification SU(2,2)/$Z_2\sim$SO$_0$(4,2), the subgroup SU(2)$_L\times$SU(2)$_R\subset$SU(2,2) is the twofold covering of the subgroup SO(4)$\subset$SO$_0$(4,2) acting on the $(\eta^1,\eta^2,\eta^3,\eta^4)$ subspace. The isotropy subgroup of SU(2)$_L\times$SU(2)$_R$ at the reference points $p_0=(e^{i\psi},1)\in$U(1)$\times$SU(2) is equal to SU(2)$_D$ and its orbits in $S^1\times S^3$ form one stratum (i.e., all isotropy subgroups are isomorphic), each being S^3, parametrized in the space of orbits by the S^1 coordinate. Thus, $\bar{M}\sim S^1\times$SU(2)$_L\times$SU(2)$_R$/SU(2) and following HSV, the SO(4) invariant gauge fields on $S^1\times S^3$ are determined by first specifying a group homomorphism, denoted λ, of the isotropy group SU(2)$_D$ into the classical gauge groups.

Since SU(2) is a simple group, the problem of the determination up to conjugacy of all its homomorphisms into SU($l+1$), SO($2l+1$), Sp(l), and SO($2l$) ($l\in$N) is related locally to the classification of the A_1 subalgebras of the respective simple Lie algebras A_l, B_l, C_l, and D_l. For any Lie group, this problem reduces to the classification of all the simple three-dimensional Lie subalgebras of the simple Lie

 0022-2488/84/051542-06$02.50

algebras (see Mal'cev[10] and Dynkin[11]). However, by considering the n-dimensional irreducible representation of A_l $(n = l + 1), B_l (n = 2l + 1), C_l$ and $D_l (n = 2l)$, Mal'cev[10] has found that each inner conjugacy class of simple subalgebras of the classical algebras corresponds in a one-to-one manner to a "mutually contragredient" system of highest weights of permissible subalgebra representations constrained so that the sum of their associated dimensions is less than or equal to $2n$, with the exception of D_l, where in general two classes are involved in this correspondence. The term permissible means that the unitary, orthogonal, or symplectic property of the n-dimensional irreducible representation must be respected by the system of subalgebra representations. But all the representations of SU(2) are self-contragredient and the irreducible representations are either symplectic, having half-integer weights, or orthogonal, with integer weights. Finally, we recall that because of the self-contragredience, it is possible to have SU(2) representations which are both orthogonal and symplectic when composed of two isomorphic representations (see Mal'cev, Ref. 10, Sec. 2, Theorem 4).

We do not study all possible classes of homomorphisms, limiting ourselves to the following cases:

(i) For SU(n): the classes of SU(2) embeddings given by the system of n/m irreducible representations $D^j(g)$ of highest weight $j = (m - 1)/2$, where m,n, and $n/m \in \mathbb{N}$, $g \in$ SU(2):

$$\lambda: (g,g) \in \mathrm{SU}(2)_D \to \mathbb{1}_{n/m} \otimes D^j(g) \in \mathrm{SU}(n). \tag{2.7}$$

(ii) For SO(n): the system is formed of n/m block-diagonal irreducible representations $D^j(g)$ with integer highest weight $j = (m - 1)/2$, where m,n, and $n/m \in \mathbb{N}$. We exclude the orthogonal representations involving symplectic irreducible representations of SU(2).

(iii) For Sp(n): we choose a system of $2n/m$ block-diagonal irreducible representations of highest weight $j = (m - 1)/2$. The above theorem of Mal'cev assures the symplectic property of the representations for odd m.

The Lie algebra so(4) $\sim$ su(2)$_L \oplus$ su(2)$_R$ admits a reductive decomposition: su(2)$_D$ + m, where su(2)$_D$ consists of elements of the form (ξ,ξ) with $\xi = \xi^i \tau_i \in$ su(2), relative to a basis with commutation rules $[\tau_i,\tau_j] = \varepsilon_{ijk}\tau_k$ and $\mathrm{m} = \{(\xi, -\xi) | \xi \in \mathrm{su}(2)\}$. Hence, for each class of homomorphisms λ, the most general invariant gauge field is expressed by the formula (see Ref. 2)

$$\omega = \lambda_* \circ \omega_L^D + \Phi_{S^1} \circ \omega_L^m + \omega_0 \theta^0, \tag{2.8}$$

where (i) $\omega_L^D = \theta^i(\tau_i,\tau_i)$ and $\omega_L^{\mathrm{m}} = \theta^i(\tau_i, -\tau_i)$ are the respective projections of the (su(2)$_D$ + m) decomposition of the pull-back of the Maurer–Cartan form on SO(4): $\omega_{\mathrm{MC}} = (v^{-1}\,dv, v'^{-1}\,dv')$ [with $v^{-1}\,dv = 2\theta^i\tau_i$, $v,v' \in$ SU(2)] under the map σ: su(2)$_D \to$ SO(4) given by $\sigma(v) = (v,e)$, which defines a convenient section of the principal bundle; (ii) λ_* represents the differential of the map λ at the identity taking su(2)$_D$ into the classical Lie algebras, and Φ_{S^1} denotes a smooth family of linear maps Φ: m$\to$classical Lie algebras, parametrized by S^1, which satisfy the condition

$$\Phi(\mathrm{Ad}_{g_0} X) = \mathrm{Ad}_{\lambda(g_0)} \Phi(X) \quad \forall\, X \in \mathrm{m} \text{ and } \forall\, g_0 \in \mathrm{SU}(2)_D. \tag{2.9}$$

(iii) θ^0 is a 1-form on S^1 and ω_0 a smooth function on S^1 with values in the centralizer c^λ of the image λ (SU(2)$_D$) in the gauge algebra (the residual gauge algebra). We can, moreover, choose a gauge where ω simplifies to

$$\omega = \lambda_* \circ \omega_L^D + \Phi_{S^1} \circ \omega_L^{\mathrm{m}}, \tag{2.10}$$

eliminating the S^1 part of the connection (2.8). We shall now determine the invariant gauge fields and their centralizer for each classical gauge group:

(i) SU(n)

The differential at the identity of the homomorphisms λ [(2.7)] is

$$\lambda_*: (\tau_i,\tau_i) \in \mathrm{su}(2)_D \to \mathbb{1}_{n/m} \otimes D^j(\tau_i) \in \mathrm{su}(n) \tag{2.11}$$

with $m = 2j + 1$ and $\tau_i \in$ su(2). Decomposing Φ as

$$\Phi \equiv \begin{bmatrix} \Phi_{11} & \cdots & \Phi_{1\frac{n}{m}} \\ \vdots & & \vdots \\ \Phi_{\frac{n}{m}1} & \cdots & \Phi_{\frac{n}{m}\frac{n}{m}} \end{bmatrix} \in \mathrm{su}(n), \tag{2.12}$$

where the $\Phi_{\mu\nu}$ are $m \times m$ matrices such that $\Phi_{\mu\nu} = -\Phi^\dagger_{\nu\mu}$, $\forall\, \mu,\nu = 1,\ldots,n/m$, and tr $\Phi = 0$, the condition (2.9) for Φ implies that

$$\Phi_{\mu\nu}(\mathrm{Ad}_{(g,g)} X) = D^j(g) \Phi_{\mu\nu}(X) D^j(g)^\dagger \tag{2.13}$$

$\forall\, \mu,\nu$, $g \in$ SU(2) and $\forall\, X \in$ m. These linear relations have the following solutions for $X = \xi^i(\tau_i, -\tau_i) \in$ m:

$$\Phi_{\mu\nu}(X) = \phi_{\mu\nu} \xi^i D^j(\tau_i), \tag{2.14}$$

$\phi_{\mu\nu}$ representing the (μ,ν) element of a constant Hermitian $(n/m) \times (n/m)$ matrix H. In tensor product notation,

$$\Phi: (\tau_i, -\tau_i) \in \mathrm{m} \to H \otimes D^j(\tau_i) \in \mathrm{su}(n), \tag{2.15}$$

where $H \in \mathbb{H}(n/m)$, the space of $(n/m) \times (n/m)$ Hermitian matrices. Accordingly, Eq. (2.10) gives the resulting SO(4) invariant SU(n) gauge fields for the homomorphisms λ:

$$\omega = H'(\psi) \otimes D^j(\tau_i) \theta^i, \tag{2.16}$$

where $\{\theta^i\}$ is the left-invariant coframe on S^3 given above, $H' \equiv H + \mathbb{1}_{n/m} \in \mathbb{H}(n/m)$, and ψ is the angular coordinate on S^1. We also compute the centralizer c^λ of the image λ_*(su(2)$_D$) in the Lie algebra su(n), consisting of elements $c \in$ su(n) satisfying

$$c = \mathrm{Ad}_{\lambda(g_0)} c \quad \forall\, g_0 \in \mathrm{SU}(2)_D. \tag{2.17}$$

It follows that for the homomorphism (2.7), $c^\lambda = \{c \in \mathrm{su}(n) | c = \Gamma \otimes \mathbb{1}_m, \Gamma \in \mathrm{su}(n/m)\}$.

(ii) SO(n)

From λ, we obtain the differential

$$\lambda_*: (\tau_i,\tau_i) \in \mathrm{su}(2)_D \to \mathbb{1}_{n/m} \otimes D^j(\tau_i) \in \mathrm{so}(n), \tag{2.18}$$

with $m = 2j + 1$ odd. Constructing Φ as in Eq. (2.12), with the conditions $\Phi^*_{\mu\nu} = \Phi_{\mu\nu}$, and $\Phi_{\mu\nu} = -\Phi^T_{\nu\mu}$, so that Φ be so(n) valued, we impose the further condition (2.9) and find

$$\Phi: (\tau_i, -\tau_i) \in \mathrm{m} \to S \otimes D^j(\tau_i), \tag{2.19}$$

where $S \in \mathbb{S}(n/m)$ is a real symmetric $(n/m) \times (n/m)$ matrix. The invariant SO(n) gauge fields are thus of the form

$$\omega = S'(\psi) \otimes D^j(\tau_i) \theta^i \tag{2.20}$$

with $S' \equiv S + \mathbf{1}_{n/m} \in \mathbb{S}(n/m)$. The associated centralizer c^λ in so(n) is $\{c = \Gamma \otimes \mathbf{1}_m \,|\, \Gamma \in \mathrm{so}(n/m)\}$. Since $\mathbb{H}(n/m) \sim \mathrm{u}(1) \oplus \mathrm{su}(n/m)$, let us remark that if we carry out the decomposition of $\mathbb{H}(n/m)$ into eigenspaces of the su(n/m) involutive automorphism of type AI (see Helgason[12]),

$$\theta(H) = H^*, \ \forall\, H \in \mathbb{H}(n/m) \tag{2.21}$$

(*: complex conjugation), we get the sum

$$\mathbb{H}(n/m) = i\{\mathrm{so}(n/m)\} \oplus \mathbb{S}(n/m). \tag{2.22}$$

As we can see, these two components are precisely the reduced gauge algebra c^λ and the linear space parametrizing the reduced gauge field.

(iii) Sp(n)

In order to perform this calculation, we distinguish between the even and odd m dimensions of the SU(2) representations in λ:

$$\lambda : (g,g) \in \mathrm{SU}(2)_D \to \mathbf{1}_{2n/m} \otimes D^j(g) \in \mathrm{Sp}(n) = \mathrm{U}(2n) \cap \mathrm{Sp}(n, \mathbb{C}),$$

where $g \in$ SU(2) and $m = 2j + 1$. The Sp(n) properties force λ to respect Eqs. (2.23) and (2.24).

$$\lambda^\dagger \lambda = \mathbf{1}_{2n}. \tag{2.23}$$

(a) even m: $\lambda^T J^e \lambda = J^e$, (2.24a)

J^e is the symplectic form matrix chosen for convenience to be

$$J^e = \mathbf{1}_{2n/m} \otimes D^j(U),$$

where $U = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in \mathrm{SU}(2)$ satisfies $g^T U g = U$.

(b) odd m: $\lambda^T J^0 \lambda = J^0$ (2.24b)

with the symplectic matrix

$$J^0 = \begin{bmatrix} 0 & \mathbf{1}_{n/m} \\ -\mathbf{1}_{n/m} & 0 \end{bmatrix} \otimes D^j(U).$$

We next apply condition (2.9) and require that Φ have values in the sp(n) algebra. This imposes the three restrictions (i) tr $\Phi = 0$, (ii) $\Phi^\dagger = -\Phi$, and (iii) $\Phi^T J^{e,0} + J^{e,0} \Phi = 0$, whether m is even or odd.

These restrictions reduce the maps Φ to

(a) For even m;

$$\Phi : (\tau_i, -\tau_i) \in \mathfrak{m} \to S \otimes D^j(\tau_i), \tag{2.25}$$

$S \in \mathbb{S}(2n/m)$. Hence, the invariant gauge fields become

$$\omega = S' \otimes D^j(\tau_i)\theta^i \quad (S' \equiv S + 1). \tag{2.26}$$

(b) For odd m;

$$\Phi : (\tau_i, -\tau_i) \in \mathfrak{m} \to \begin{bmatrix} H & A \\ -A^* & H^* \end{bmatrix} \otimes D^j(\tau_i), \tag{2.27}$$

$H \in \mathbb{H}(n/m)$ and $A \in \mathbb{A}^c(n/m)$ (antisymmetric complex matrices). The invariant gauge fields are written as

$$\omega = \begin{bmatrix} H' & A \\ -A^* & H'^* \end{bmatrix} \otimes D^j(\tau_i)\theta^i \quad (H' \equiv H + 1). \tag{2.28}$$

The computation of the centralizer c^λ yields

(a) For even m;

$$c_e^\lambda = \{c = \Gamma \otimes \mathbf{1}_m \,|\, \Gamma \in \mathrm{so}(2n/m)\}.$$

(b) For odd m;

$$c_0^\lambda = \left\{ c = \Gamma \otimes \mathbf{1}_m \,|\, \Gamma = \begin{bmatrix} \gamma_1 & \gamma_2 \\ -\gamma_2^* & \gamma_1^* \end{bmatrix} \in \mathrm{sp}\left(\frac{n}{m}\right); \right.$$

$$\left. \text{i.e., } \gamma_1, \gamma_2 \in \mathbb{C}^{n/m \times n/m}, \gamma_1^\dagger = -\gamma_1, \text{ and } \gamma_2^T = \gamma_2 \right\}.$$

In each case, there exists similarly to the SO(n) problem an involutive automorphism of su($2n/m$) dividing the space of Hermitian $(2n/m) \times (2n/m)$ matrices into eigenspaces associated with the centralizer (C^λ) or the invariant gauge field. Namely,

(a) A type AI involutive automorphism for even m;

$$\theta(H) = H^*, \tag{2.29}$$

giving

$$H\left(\frac{2n}{m}\right) = i\left\{\mathrm{so}\left(\frac{2n}{m}\right)\right\} \oplus \mathbb{S}\left(\frac{2n}{m}\right). \tag{2.30}$$

(b) A type AII involutive automorphism for odd m;

$$\theta(H) = JH^*J^{-1}, \quad J = \begin{bmatrix} 0 & \mathbf{1}_{n/m} \\ -\mathbf{1}_{n/m} & 0 \end{bmatrix} \tag{2.31}$$

giving

$$\mathbb{H}\left(\frac{2n}{m}\right) = i\left\{\mathrm{sp}\left(\frac{n}{m}\right)\right\} \oplus \left\{\begin{bmatrix} H & A \\ -A^* & H^* \end{bmatrix}\right\}. \tag{2.32}$$

3. YANG–MILLS EQUATIONS

As we have seen in the previous section, the SO(4) invariant gauge fields may be expressed in a convenient gauge as the tensor product of a matrix function on S^1: $M'(\psi)$ and a set of 1-forms on S^3 with values in an irreducible representation of the su(2) Lie algebra:

$$\omega = M'(\psi) \otimes D^j(\tau_i)\theta^i. \tag{3.1}$$

We have summarized in Table I the different $M'(\psi)$ for each classical group together with the associated centralizer c^λ and its group equivalent C^λ. Inserting these fields in the pure Yang–Mills equations,

$$D^* D\omega = 0, \tag{3.2}$$

we derive the following set of differential equations:

$$\ddot{M} = 2M(1 - M^2) \tag{3.3}$$

and

$$[M, \dot{M}] = 0, \tag{3.4}$$

where $\dot{M} \equiv dM/d\psi$ and $M' \equiv M + 1$. Alternatively, we can arrive at the same result by reducing the Lagrangian density $\mathcal{L}$ which equals

$$\mathcal{L} = k \,\mathrm{tr}(F \wedge {}^*F), \tag{3.5}$$

where k is a constant, $F = D\omega$ defines the curvature, ${}^*F$ its dual, and ω is the gauge potential. Choosing ω to be of the general reduced form (with residual gauge algebra c^λ):

TABLE I. Components of the reduced gauge potentials for the homomorphisms λ. m is the dimension of the irreducible SU(2) components of λ.

Gauge group	SU(n)	SO(n) (odd m)	Sp(n) (even m)	Sp(n) (odd m)
$M'(\psi)$	$H\in\mathbf{H}\left(\frac{n}{m}\right)$	$S\in\mathbf{S}\left(\frac{n}{m}\right)$	$S\in\mathbf{S}\left(\frac{2n}{m}\right)$	$\begin{bmatrix} H & A \\ -A^* & H^* \end{bmatrix}$; $H\in\mathbf{H}\left(\frac{n}{m}\right)$, $A\in\mathbf{A}^c\left(\frac{n}{m}\right)$
Algebra centralizer c^λ	$\left\{\Gamma\otimes\mathbf{1}_m \mid \Gamma\in\mathrm{su}\left(\frac{n}{m}\right)\right\}$	$\left\{\Gamma\otimes\mathbf{1}_m \mid \Gamma\in\mathrm{so}\left(\frac{n}{m}\right)\right\}$	$\left\{\Gamma\otimes\mathbf{1}_m \mid \Gamma\in\mathrm{so}\left(\frac{2n}{m}\right)\right\}$	$\left\{\Gamma\otimes\mathbf{1}_m \mid \Gamma\in\mathrm{sp}\left(\frac{n}{m}\right)\right\}$
Group centralizer C^λ	$\left\{\Lambda\otimes\mathbf{1}_m \mid \Lambda\in\mathrm{SU}\left(\frac{n}{m}\right)\right\}$	$\left\{\Lambda\otimes\mathbf{1}_m \mid \Lambda\in\mathrm{SO}\left(\frac{n}{m}\right)\right\}$	$\left\{\Lambda\otimes\mathbf{1}_m \mid \Lambda\in\mathrm{SO}\left(\frac{2n}{m}\right)\right\}$	$\left\{\Lambda\otimes\mathbf{1}_m \mid \Lambda\in\mathrm{Sp}\left(\frac{n}{m}\right)\right\}$

$$\omega = M'(\psi)\otimes D^j(\tau_i)\theta^i + \Gamma(\psi)\otimes\mathbf{1}_m d\psi, \tag{3.6}$$

this gives rise to a reduced Lagrangian density $\mathscr{L}_R$:

$$\mathscr{L}_R = \tfrac{1}{2}\,\mathrm{tr}((\mathscr{D}M)^2 - (\mathbf{1}-M^2)^2) \tag{3.7}$$

with the covariant derivative $\mathscr{D}M = \dot{M} + [\Gamma,M]$. The variational equations deduced from $\mathscr{L}_R$ are

$$\mathscr{D}(\mathscr{D}M) = 2M(\mathbf{1}-M^2) \tag{3.8}$$

and

$$[M,\mathscr{D}M] = 0. \tag{3.9}$$

If we make a gauge choice such that $\Gamma = 0$, we recover exactly the system (3.3) and (3.4) introduced earlier, and the residual gauge group consists of constant (type I) gauge transformations. Before solving these equations, we shall give an interpretation of this system in terms of a Hamiltonian system with symmetry.[13]

Let us consider for the gauge group SU(n) the symplectic manifold $T^*\mathbf{H}(n/m)\sim\{(H,P)\in\mathbf{H}(n/m)\times\mathbf{H}(n/m)\}$ by the identification $\mathbf{H}(n/m)\sim\mathbf{H}^*(n/m)$ given by the inner product $P\cdot H = \mathrm{tr}(PH)$ and symplectic form $\Omega = \mathrm{tr}(dH\wedge dP)$. As suggested by the reduced Lagrangian density, we consider the Hamiltonian

$$\mathfrak{H} = \tfrac{1}{2}\,\mathrm{tr}(P^2 + (\mathbf{1}_{n/m} - H^2)^2). \tag{3.10}$$

It then follows that Hamilton's equations

$$\dot{H} = P \quad\text{and}\quad \dot{P} = 2H(1-H^2) \tag{3.11}$$

are equivalent to the SO(4) reduced Yang–Mills equation (3.3). Futhermore, the symplectic action of SU(n/m) on $(T^*\mathbf{H}(n/m),\Omega)$: $(H,P)\mapsto(\Lambda H\Lambda^\dagger,\Lambda P\Lambda^\dagger)$, $\Lambda\in\mathrm{SU}(n/m)$, leaves invariant the Hamiltonian system. This symmetry group is exactly the reduced group of constant gauge transformations in C^λ. But a calculation shows that its "momentum" mapping[13] corresponds to

$$J(H,P) = [H,P]\in\mathrm{su}^*(n/m). \tag{3.12}$$

This means that the reduced Yang–Mills equation (3.4) fixes the momentum map value as zero,

$$[H,\dot{H}] = [H,P] = 0. \tag{3.13}$$

For the (i) SO(n), (ii) Sp(n) (even m), and (iii) Sp(n) (odd m) reductions, respectively, we restrict to the symplectic subspaces:

(i) $\{(S,P)\in\mathbf{S}(n/m)\times\mathbf{S}(n/m)\}$ with $\Omega = \mathrm{tr}(dS\wedge dP)$,
(ii) $\{(S,P)\in\mathbf{S}(2n/m)\times\mathbf{S}(2n/m)\}$ with $\Omega = \mathrm{tr}(dS\wedge dP)$,
(iii)

$$\left\{T=\begin{bmatrix} H & A \\ -A^* & H^* \end{bmatrix}, P=\begin{bmatrix} H' & A' \\ -A'^* & H'^* \end{bmatrix} \;\middle|\; H,H'\in\mathbf{H}\left(\frac{n}{m}\right),\quad A,A'\in\mathbf{A}^c\left(\frac{n}{m}\right)\right\}$$

with $\Omega = \mathrm{tr}(dT\wedge dP)$.

The Hamiltonian function (3.10) is defined on these subspaces and the same interpretation of the reduced Yang–Mills equations holds with regard to the corresponding symplectic action of the centralizers (i) SO(n/m), (ii) SO($2n/m$), and (iii) Sp(n/m).

Since there is a similar method of solution for each of the reduced systems studied, we only detail the procedure for the SU(n) Yang–Mills case and summarize the solutions for the remaining classical gauge groups.

In the Hamiltonian formulation of the SU(n) system, we have seen that H and P are commuting $(n/m)\times(n/m)$ Hermitian matrices. There therefore exists an element $U(\psi)\in\mathrm{SU}(n/m)$ which diagonalizes them simultaneously:

$$H = UH_DU^\dagger \quad\text{and}\quad P = UP_DU^\dagger. \tag{3.14}$$

It follows from Hamilton's equations (3.11) that the first derivatives of the diagonal matrices H_D and P_D satisfy

$$\dot{H}_D = -[u,H_D] + P_D \tag{3.15}$$

and

$$\dot{P}_D = -[u,P_D] + 2H_D(1-H_D^2), \tag{3.16}$$

where

$$\dot{U} = Uu, \tag{3.17}$$

$u\in u(n/m)$, the U(n/m) Lie algebra. Splitting these equations into two parts by orthogonal projection (under the Killing form) onto the stabilizer $\mathscr{G}_0\subset u(n/m)$ of (H_D,P_D) and its complement, we find that $u(\psi)\in\mathscr{G}_0$:

$$[u,H_D] = 0 \quad\text{and}\quad [u,P_D] = 0 \tag{3.18}$$

and

$$\dot{H}_D = P_D \quad\text{and}\quad \dot{P}_D = 2H_D(1-H_D^2). \tag{3.19}$$

Define now a function $g(\psi)$ with values in the common stabi-

lizer G_0 of H_D and P_D:

$$G_0 = \{g \in \mathrm{SU}(n/m) | H_D = \mathrm{Ad}_g H_D \text{ and } P_D = \mathrm{Ad}_g P_D\} \tag{3.20}$$

by the relation

$$\dot{g} = gu. \tag{3.21}$$

Generically, G_0 consists of the group $[\mathrm{U}(1)]^{(n/m)-1}$ of diagonal unitary matrices, although in principle, it may vary with ψ as various eigenvalue degeneracies occur. But for each value of ψ, $g(\psi)$ stabilizes $H_D(\psi)$ and $P_D(\psi)$ provided it does for some initial $\psi = \psi_0$.

If we next decompose $U(\psi)$ as

$$U(\psi) = V(\psi)g(\psi) \tag{3.22}$$

we deduce by differentiating (3.22) that V must be a constant element of $\mathrm{SU}(n/m)$. Hence $(H,P) \in T^*\mathbb{H}(n/m)$ is conjugate to (H_D,P_D) by the constant matrix V:

$$H = VH_DV^\dagger \quad \text{and} \quad P = VP_DV^\dagger. \tag{3.23}$$

Let us note that the above procedure is an application of the Marsden–Weinstein reduction scheme.[13,14] Since the value of the momentum map for this reduction is zero, its isotropy group is the entire $\mathrm{SU}(n/m)$, and quotienting the inverse image $J^{-1}(0)$ in $T^*\mathbb{H}(n/m) \sim \mathbb{H}(n/m) \times \mathbb{H}(n/m)$ leads to the reduced phase space defined by the pairs of diagonal matrices (H_D,P_D) and the reduced Hamilton equations (3.19). The Hamiltonian flow for the original problem is reconstructed from the reduced one by integrating (3.17) in the orbits $\mathrm{SU}(n/m)/G_0$. But interpreted on the coset space, this is just the zero vector field and hence the matrix V appearing in Eq. (3.23) may be taken as constant.

In second-order form, Eqs. (3.19) are

$$\ddot{H}_D = 2H_D(1 - H_D^2). \tag{3.24}$$

This involves a system of n/m uncoupled one-dimensional anharmonic oscillators $h_i(\psi) \in \mathbb{R}$ $(i = 1,...,n/m)$ described by

$$\ddot{h}_i = 2h_i(1 - h_i^2). \tag{3.25}$$

The solution to the second-order ordinary differential equation (3.25) is known in terms of elliptic functions and provides the solution to the reduced SU(2) problem.[7,8] We may thus express the solution to the reduced SU(n) Yang–Mills equations in the simple form:

$$M(\psi) = H(\psi) = Vd^{n/m}(\psi)V^\dagger, \tag{3.26}$$

where

$$d^l(\psi) \equiv \begin{bmatrix} & & 0 \\ & h_i(\psi) & \\ 0 & & \end{bmatrix} \quad (i = 1,...,l) \tag{3.27}$$

is a diagonal matrix of l solutions $(h_i(\psi))$ to Eq. (3.25) (see Refs. 7 and 8):

(a) if $E_i \equiv \frac{1}{2}(\dot{h}_i^2 + (1 - h_i^2)^2) \leqslant \frac{1}{2}$,

$$h_i(\psi) = \pm(1 + \sqrt{2E_i})^{1/2} dn\left[(1 + \sqrt{2E_i})^{1/2}(\psi - \psi_0); \alpha_i\right], \tag{3.28}$$

where

$$\alpha_i \equiv 2\sqrt{2E_i}/(1 + \sqrt{2E_i}).$$

(b) if $E_i > \frac{1}{2}$,

$$h_i(\psi) = -(1 + \sqrt{2E_i})^{1/2} cn\left[(8E_i)^{1/4}(\psi - \psi_0); \beta_i\right], \tag{3.29}$$

where

$$\beta_i \equiv (1 + 2E_i)/2\sqrt{2E_i}.$$

We finally write the solutions to the other reduced Yang–Mills systems which are obtained in the same manner:

(i) SO(n):

$$M(\psi) = Vd^{n/m}(\psi)V^T, \tag{3.30}$$

with a constant $V \in \mathrm{SO}(n/m)$,

(ii) Sp(n) (even m):

$$M(\psi) = Vd^{2n/m}(\psi)V^T, \tag{3.31}$$

with a constant $V \in \mathrm{SO}(2n/m)$;

(iii) Sp(n) (odd m):

$$M(\psi) = V\begin{bmatrix} d^{n/m}(\psi) & 0 \\ 0 & d^{n/m}(\psi) \end{bmatrix} V^\dagger, \tag{3.32}$$

with a constant $V \in \mathrm{Sp}(n/m)$.

4. SUMMARY

In this paper, we have explicitly reduced the SU(n), SO(n), and Sp(n) Yang–Mills equations on compactified Minkowski space $(S^1 \times S^3/Z_2)$ using SO(4) invariant gauge fields characterized by a particular set of homomorphisms λ of the isotropy subgroup $\mathrm{SU}(2)_D$ into the classical gauge groups. More generally, the homomorphisms of SU(2) into simple subgroups are classified by a set of highest weights of SU(2) irreducible representations, the sum of whose dimensions does not exceed the dimension of the smallest linear irreducible representation of the simple group.[10] Here, we have restricted ourselves to homomorphisms specified by irreducible representations of equal highest weights. For each embedding, the corresponding reduced Yang–Mills equations consist of matricial ordinary differential equations in the variable $\psi \in S^1$, which can be interpreted as a Hamiltonian system with symmetry constrained so that the associated momentum map has value zero. We have found that for this type of SO(4) action, the invariant solutions may be given in the form of a diagonal set of solutions to decoupled one-dimensional anharmonic oscillators, conjugated by an arbitrary constant element of the symmetry group.

In order to treat other SO(4) reductions which are not just uncoupled direct sums of the type obtained above, it is necessary to consider embeddings in the classical groups constructed with consecutive sequences $\{j, j+1,..., j+k\}$ of highest weights, which reduces to the study of adjacent pairs of weights $(j, j+1)$. It is very likely that the resulting reduced systems have a similar interpretation in terms of Hamiltonian systems with symmetry, and this will be the subject of further study.

Other possible further developments of this work would be the study of SO(4) invariant reduced systems involving spinors[5,6] or scalars[4] minimally coupled to gauge fields, reductions of gauge systems under other symmetry groups, and the study of the various corresponding semiclassical and quantum problems, generalizing the work done for the SU(2) gauge group.[15]

ACKNOWLEDGMENTS

One of the authors (M. L.) gratefully acknowledges scholarships from the Natural Science and Engineering Research Council of Canada, the fonds FCAC (Government du Québec), and the Université de Montréal.

[1]J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980).
[2]J. Harnad, S. Shnider, and J. Tafel, Lett. Math. Phys. **4**, 107 (1980).
[3]J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979).
[4]L. Vinet, Phys. Rev. D **24**, 3179 (1981).
[5]J. Doneux, Y. Saint-Aubin, and L. Vinet, Phys. Rev. D **25**, 484 (1982).
[6]M. Légaré, J. Math. Phys. **24**, 1219 (1983).
[7]M. Luscher, Phys. Lett. **70B**, 321 (1977).
[8]B. M. Schechter, Phys. Rev. D **16**, 3015 (1977).
[9]K. Meetz, Z. Phys. C **6**, 41 (1980).
[10]A. I. Mal'cev, Am. Math. Soc. Transl. Ser. 1 **9**, 173 (1962) or **33** (1950).
[11]E. B. Dynkin, Am. Math. Soc. Transl. Ser. 2 **6**, 111 (1965).
[12]S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978), p. 451.
[13]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), Chap. 4.
[14]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[15]R. Jackiw, C. Nohl, and C. Rebbi, in *Particles and Fields*, edited by D. Boal and A. Kamal (Plenum, New York, 1978).

Leading and misleading logs in perturbative QCD

M. R. Pennington
Department of Physics, University of Durham, Durham, United Kingdom

(Received 10 June 1983; accepted for publication 26 August 1983)

Misleading results can be obtained by applying a leading logarithm approximation where it is not valid. However obvious this fact, its neglect is shown to have led to misleading and even erroneous conclusions in several recent practical applications of perturbative QCD to the problem of summing soft gluons.

PACS numbers: 12.35.Cn, 12.35.Eq

1. THE APPROXIMATION

The leading logarithm approximation is natural in theoretical considerations of perturbative QCD.[1] There variables can be readily taken to infinity with impunity. However, in phenomenological applications, care must be taken not to misapply such results in kinematic regimes where the supposed leading log is not leading at all. For example, summing the leading double logarithms of multigluon emissions gives the well-known Sudakov-like quark form factor.[2–5] Though this exponential suppression is true asymptotically, it is readily masked by nonleading contributions at all foreseeable energies.[4,5] Thus, while of little concern to theorists calculating only asymptotic results, an understanding of the range of validity of these results is paramount in any practical evaluation, hence Ref. 6. Though this is self-evident both to those motivated to calculate analytically beyond the leading order (as pointed out in Ref. 7) and to some of those who compute numerically,[8,9] it has been missed in much recent phenomenology, leading to misleading, and often erroneous, conclusions.

The prospect of forthcoming experimental results has prompted a number of predictions for the transverse momentum distribution of hadrons in e^+e^- annihilation and of the Z^0 in $\bar{p}p$ interactions.[4,5,10–15] Many of these attempt to sum the effects of very soft gluons in a useful phenomenological fashion and it is in this context that our general point is made.

In Sec. 2, we consider a simple illustration of the effect of a misleading logarithm approximation. In Sec. 3, we discuss the way phenomenological calculations deal with the cancellation of infrared divergent logarithms and how neglected constant terms again affect the result. Sections 4 and 5 are devoted to studying the evolution of fragmentation functions summing soft gluon emissions and the misapplication of a leading logarithm approximation of Bassetto, Ciafaloni, and Marchesini[16] in recent phenomenology. In Sec. 6, the conclusions are given.

2. WHEN IS A LOGARITHM LEADING?

In many papers the leading logarithm approximation (LLA) and the full $O(\alpha_s)$ result of a perturbative calculation are differentiated. This differentiation is clearly artificial: only the latter is unique. Consider either the process $\gamma^* \to q\bar{q}g$ relevant for jet production or $q\bar{q} \to Zg$ as applicable for Z production by the Drell–Yan mechanism. Then when the gluon has small relative transverse momentum k_T (more precisely defined below), these differential cross sections for single gluon emission $\sigma^{(1)}$ behave like

$$\frac{d\sigma^{(1)}}{dk_T^2} \sim \frac{|\ln k_T^2|}{k_T^2} \tag{1}$$

in the LLA. However, some mass sets the scale for k_T^2 in the logarithm and as soon as we ask what this is, our discussion goes beyond the LLA. The LLA only strictly applies when $\ln 1/k_T^2$ is larger than any constant. Phenomenology in the LLA only has a meaning if one chooses a specific scale. At first sight, what scale appears merely a matter of choice. This is at least how it has been treated. However, the range of validity of any LLA is wholly determined by the size of the nonleading corrections. Since most phenomenological applications want to cover the whole range of permissible k_T, the scale of the logarithm must be chosen to mimic the full $O(\alpha_s)$ result, which is the only result applicable in the whole kinematic regime. Any distinction between full $O(\alpha_s)$ and the LLA is a mere artifact.

The theoretical advantage of the LLA is that the leading logs are generated in a universal, process-independent way and so amenable to a general treatment, while the terms below the leading depend on the process. In phenomenological applications it is important to recognize this distinction. For illustrative purposes we will specifically discuss e^+e^- annihilation, though, of course, the points are more general. In studying $e^+e^- \to \gamma^* \to q\bar{q}g$, it is natural to work in the Dalitz plot variables x_i $(i = 1,2,3)$, which are the fraction of the maximum available energy carried by $q,\bar{q}$ and g, respectively (Fig. 1), in the virtual photon rest frame. Energy conservation relates $x_1 + x_2 + x_3 = 2$. With Q the mass of the photon, it is useful to define x_T to be the corresponding fraction of the maximum transverse momentum relative to some axis (Fig. 1), so that $x_T = 2k_T/Q$ as $x_i = 2E_i/Q$. As is well-known,[17] the differential cross section for single gluon emission at $x_T \neq 0$ is

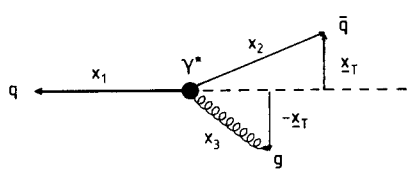


FIG. 1. The variables used to study $\gamma^* \to q\bar{q}g$ in the virtual photon rest frame as fractions of half the photon mass, so that $x_1 + x_2 + x_3 = 2$.

 0022-2488/84/051548-07$02.50

$$\frac{1}{\sigma_0}\frac{d^2\sigma^{(1)}}{dx_1 dx_2}=\frac{\alpha_s}{2\pi}C_F\frac{x_1^2+x_2^2}{(1-x_1)(1-x_2)}, \tag{2}$$

where the color Casimir $C_F=\frac{4}{3}$ and σ_0 is the parton cross section $4\pi\alpha^2 N_c e_i^2/3Q^2$ for each quark of flavor i. Virtual gluon corrections, which we discuss in more detail in Sec. 3, render the integrated cross section finite. With x_T measured relative to the x_1 axis we find that

$$\frac{1}{\sigma_0}\frac{d^2\sigma^{(1)}}{dx_1 dx_T^2}=\frac{\alpha_s}{2\pi}\frac{C_F}{x_T^2}\left(\frac{1+x_1^2}{1-x_1}-\frac{x_T^2 x_1(1-\frac{1}{4}x_1)}{(1-x_1)^2}\right)\times\left(1-\frac{x_T^2}{1-x_1}\right)^{-1/2}. \tag{3}$$

The leading logarithm result for $d\sigma^{(1)}/dx_T^2$ comes from the range of x_1 as close to the pole as possible at fixed x_T. A curve of fixed x_T is shown in the Dalitz plot [Fig. 2(a)]. Integrating over all x_1 from 0 to $(x_1)_{\max}=1-x_T^2$ we obtain analytically[6,18]

$$\frac{1}{\sigma_0}\frac{d\sigma^{(1)}}{dx_T^2}=\frac{2\alpha_s}{\pi}\frac{C_F}{x_T^2}\left(\left(1-\frac{x_T^2}{4}+\frac{x_T^4}{4}\right)\ln\left(\frac{1+(1-x_T^2)^{1/2}}{x_T}\right)-\frac{1}{4}\sqrt{1-x_T^2}(3-x_T^2)\right). \tag{4}$$

Now taking the LLA when $x_T\ll 1$, we see

$$\frac{1}{\sigma_0}\frac{d\sigma^{(1)}}{dx_T^2}=\frac{\alpha_s}{\pi}\frac{C_F}{x_T^2}\ln(1/x_T^2), \tag{5}$$

though from Eq. (4) we note that for $x_T\ll 1$ the argument of the logarithm is really $4/x_T^2$. Either of course is true in the LLA. However, they will have quite different ranges of validity. In phenomenological applications we want to apply these formulas to regimes where the scale is obviously relevant. There is then only one LLA which has the maximum range of validity, namely that for which the so-called LLA and the full $O(\alpha_s)$ are minimally different.

As noted in Ref. 6, for small x_T the form $\ln 1/x_T^2$ is in fact more appropriate, since the terms in the bracket of Eq. (4) are dominated by

$$\tfrac{1}{2}(\ln(1/x_T^2)+\ln 4-\tfrac{3}{2}), \tag{6}$$

where the last two terms almost exactly cancel. Thus, if, for example, in Ref. 11, $\ln M_z^2/4k_T^2$ had been used to define the LLA for Z production, their results for LLA and full $O(\alpha_s)$ would become indistinguishable. This has the important implication that, as far as this effect is concerned, it is only the narrower prediction of Ref. 11 for the transverse momentum distribution for W production that is relevant as a probe of the structure of perturbative QCD.

This provides a simple illustration of a misleading log approximation, which we will meet again in Sec. 5. In the next sections more crucial examples will be given.

In jet production in e^+e^- annihilation it is experimentally more feasible[19] to define the transverse momentum relative to the "parton" axis with the maximum energy, i.e., with largest x_i. The Dalitz plot for three jet production is then divided into three regions with q, $\bar{q}$, and g as thrust axes. In Fig. 2(b), we show a curve at fixed x_T in each of these regions. The differential cross section for one gluon emission is then given by

$$\frac{1}{\sigma_0}\frac{d\sigma^{(1)}}{dx_T^2}=\frac{2\alpha_s}{\pi}\frac{C_F}{x_T^2}\left(\left(1-\frac{x_T^2}{4}+\frac{x_T^4}{4}\right)\ln\left(\sqrt{1-T_m}\,\frac{(1+\xi)}{x_T}\right)-\frac{\xi}{8}(6-2T_m-T_m^2)+\frac{\xi}{4}x_T^2(1-T_m)\right) \tag{7}$$

for q or $\bar{q}$ as the thrust axis, and

$$\frac{1}{\sigma_0}\frac{d\sigma^{(1)}}{dx_T^2}=\frac{\alpha_s}{\pi}\frac{C_F}{x_T^2}\left(\frac{2}{\sqrt{1-x_T^2}}\ln\left(\frac{\sqrt{1-x_T^2}+\xi}{\sqrt{1-x_T^2}-\xi}\right)-4\left(1+\frac{x_T^2}{2}+\frac{x_T^4}{16}\right)\ln\left(\sqrt{1-T_m}\,\frac{(1+\xi)}{x_T}\right)-\frac{\xi}{2}(1-T_m)(3-T_m)-\frac{\xi}{4}x_T^2(1-T_m)\right) \tag{8}$$

with the gluon as the thrust axis, where

$$\xi^2=1-x_T^2/(1-T_m),$$

with T_m the minimum value of the thrust variable at a given x_T, i.e., it is the root of the cubic

$$x_T^2 T_m^2=4(1-T_m)^2(2T_m-1) \tag{9}$$

with $\frac{2}{3}\leqslant T_m\leqslant 1$. The kinematically allowed range of x_T is then $0\leqslant x_T^2\leqslant\frac{1}{3}$.

As seen from a comparison of Figs. 2(a) and (b) the behavior of the cross section for very small k_T is the same whether or not we define a thrust axis or not. Note that for $x_T\ll 1$, $T_m\simeq 1-\frac{1}{2}x_T$, $\xi\simeq 1-x_T$, so that the leading log behavior of Eq. (7) is just half Eq. (5), for each of $q,\bar{q}$ as the thrust axis. The cross section with the gluon as a thrust axis does not contribute, of course, to the LL behavior, since the poles at $x_1=1$ and $x_2=1$ of Eq. (2) do not occur in the region of gluon thrust [Fig. 2(b)]. Though the k_T's defined with the q as the thrust axis and with $\bar{q}$ as the thrust axis are distinct, amusingly the corrections to the LLA, which determine its range of validity are also simply related to that when no thrust axis is defined, viz. for small x_T, Eq. (7) gives just half of Eq. (6) too for each of q and $\bar{q}$. Brown and Ellis[20] find an identical form to Eq. (6) for the energy weighted back-to-back correlation too. This means that the energy weighting

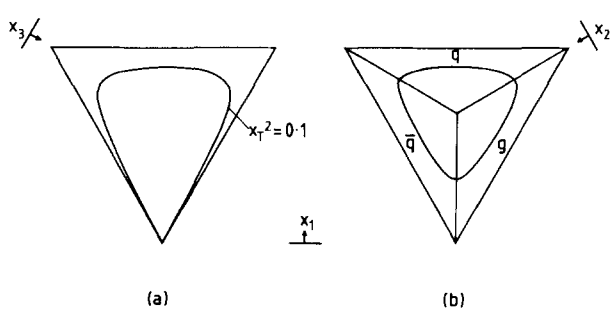


FIG. 2. The Dalitz plot for $\gamma^*\to q\bar{q}g$: (a) shows a contour at fixed x_T, where x_T is defined relative to the x_1 axis as in Fig. 1; (b) shows contours at fixed x_T, where x_T is defined relative to the appropriate thrust axis (i.e., that with the largest x_i) in each region.

factor x_1, being less than 1 for small x_T (see Fig. 2), only affects terms beyond the constant in powers of x_T^2, just as the difference between thrust axes for defining x_T does here. Thus the LLA has a similar range of validity in each case.

3. CANCELLING INFRARED DIVERGENCES BEYOND THE LEADING LOGS

As displayed in Eqs. (2)–(5), the cross section for real gluon emission is singular at $x_T = 0$. It is well-known that this is regularized by the virtual gluon graphs to give a finite $O(\alpha_s)$ cross section $\Sigma^{(1)}$. Remembering the virtual graphs only contribute at $Q_T = 0$, Parisi and Petronzio[4] give the following elegant prescription for defining this regularized cross section:

$$\frac{d\Sigma^{(1)}}{dQ_T^2} = \frac{d\sigma^{(1)}}{dQ_T^2} - \delta(Q_T^2)\int dk_T^2 \frac{d\sigma^{(1)}}{dk_T^2}. \tag{10}$$

The beauty of this regularization is especially apparent if the b-space transform is considered,

$$B(b) = \int_0^1 \frac{dx_T^2}{\sigma_0}\frac{d\Sigma^{(1)}}{dx_T^2} J_0\left(\frac{1}{2}x_T Qb\right) \tag{11}$$

$$= \int_0^1 \frac{dx_T^2}{\sigma_0}\frac{d\sigma^{(1)}}{dx_T^2}\left(J_0\left(\frac{1}{2}x_T Qb\right) - 1\right), \tag{12}$$

which is trivially finite. Moreover, all soft gluons can then be simply summed, assuming they are independently emitted, by[4]

$$\frac{1}{\sigma_0}\frac{d\sigma}{dQ_T^2} = \frac{1}{2}\int_0^\infty db\, b \exp(B(b))J_0(Q_T b). \tag{13}$$

However, this regularization, Eq. (10), is only appropriate for leading logs and not when superficially subleading terms are important. This is seen by noting that the virtual gluon contributions included there not only make the one gluon emission (and hence the n-gluon) cross section finite, but in fact zero. Asymptotically this means the all gluon and "no" gluon cross sections are equal[11] even without asymptotic freedom, viz.

$$\int dQ_T^2 \frac{d\sigma}{dQ_T^2} = \sigma_0. \tag{14}$$

This is, of course, an oversimplification of the Kinoshita–Lee–Nauenberg theorem. For though the virtual gluon graphs make the integrated cross section finite to any order, they do not make it vanish. Indeed, for the integrated one-gluon cross section the answer is well known for e^+e^- annihilation[21]

$$\Sigma^{(1)} = \tfrac{3}{4}C_F(\alpha_s(Q^2)/\pi)\sigma_0. \tag{15}$$

To recall how this arises, let us give the gluon a nonzero mass m_g so that the poles of Eq. (2) at $x_1 = 1$ and $x_2 = 1$, when the q and $\bar{q}$ are, respectively on shell, are no longer in the physical region. By this device the real and virtual gluon cross sections are both finite with cancelling double and single logarithms $\ln^2 Q^2/m_g^2 - 3\ln Q^2/m_g^2$. The Parisi–Petronzio prescription is constructed to ensure the cancellation of these logarithmic terms. However, a finite, process-dependent part remains, e.g., Eq. (15); a part that may be relevant at finite energies. Thus in e^+e^- annihilation the regularized one-gluon cross section $\Sigma^{(1)}$ should be given beyond the LLA at $O(\alpha_s)$ by

$$\frac{d\Sigma^{(1)}}{dQ_T^2} = \frac{d\sigma^{(1)}}{dQ_T^2} - \delta(Q_T^2)\left(\int dk_T^2 \frac{d\sigma^{(1)}}{dk_T^2} - \frac{3}{4}C_F\frac{|\alpha_s(Q^2)|}{\pi}\right) \tag{16}$$

to reproduce Eq. (15). Calling the Bessel transform defined in Eqs. (11) and (12) $B_0(b)$, the full transform, i.e., that of Eq. (16), is simply

$$B(b) = B_0(b) + \tfrac{3}{4}C_F|\alpha_s(Q^2)|/\pi \tag{17}$$

at fixed Q^2: a significant shift (see Fig. 2 of Ref. 6). This then renormalizes the "all" gluon cross section to be [cf. Eq. (13)]

$$\frac{1}{\sigma_0}\frac{d\sigma}{dQ_T^2} = \exp\left(\frac{|\alpha_s(Q^2)|}{\pi}\right)\frac{1}{2}\int db\, b \exp(B_0(b))J_0(Q_T b). \tag{18}$$

For the Drell–Yan process, a numerically larger renormalization occurs complicated by the fact that the virtual and real graphs involve spacelike and timelike Q^2, respectively, so that while the log and double log still cancel, there is a π^2 left behind.[3] Assuming the independent emission of gluons, fundamental to these summations [Eq. (10)–(13)], the cross section is renormalized by $\exp(C_F(\alpha_s/2\pi)\pi^2)$: the famous K factor of 2-3[22] below 100 GeV^2 results. Its inclusion significantly increases the predictions for W,Z production beyond those of Ref. 11, even at the $\bar{p}p$ collider. This current experiment will check.

4. EVOLVING FRAGMENTATION FUNCTIONS

In this section, we discuss another example of a misleading logarithm, generated by the use of a small transverse momentum approximation in a regime where it is not valid. Above we have considered the emission of many soft gluons without regard to how they fragment into hadrons. If we want to compare with experiment, we must fold in the appropriate fragmentation functions. For just a quark, or antiquark, this is trivial, e.g., Ref. 23. However, when many gluons are emitted, we must consider the evolution of the fragmentation functions summing these multigluon effects. For the nonsinglet fragmentation function $D(Q^2,p_T,x)$, its evolution is given by[16]

$$Q^2\frac{\partial}{\partial Q^2}D(Q^2,p_T,x) = \int_x^1 \frac{dz}{z}\left(\frac{\alpha_s(Q^2)}{2\pi}P_{qq}(z)\right)_+ \times\int\frac{d^2k_T}{\pi}\delta(z(1-z)Q^2 - k_T^2) \times D\left(Q^2,p_T - \frac{x}{z}k_T,\frac{x}{z}\right), \tag{19}$$

where strictly speaking the argument of α_s is not Q^2 but $Q^2(1-z)$.[24] Though this has an appreciable effect, as discussed in Ref. 25, we shall, for simplicity, ignore this, as it just complicates the point we want to make. The derivation of this equation follows from Eq. (2) with the approximation $z_T \equiv 2k_T/Q \ll z$, when the general lowest-order splitting function $\mathcal{P}_{qq}(z,z_T)$ reduces to $P_{qq}(z)$. We will relax this approximation in the next section.

To solve this equation, it is convenient once again to transform to the impact parameter space conjugate to the transverse momentum. Then in terms of the Fourier transform

$$\mathscr{D}(Q^2,b,x)\equiv\int d^2p_T e^{-i\mathbf{b}\cdot\mathbf{p}_T/x}D(Q^2,p_T,x)\,, \tag{20}$$

Eq. (19) transforms to

$$Q^2\frac{\partial}{\partial Q^2}\mathscr{D}(Q^2,b,x)=\int_x^1\frac{dz}{z}\left(\frac{\alpha_s(Q^2)}{2\pi}P_{qq}(z)\right)_+ \times J_0\left(Qb\sqrt{\frac{1-z}{z}}\right)\mathscr{D}\left(Q^2,b,\frac{x}{z}\right). \tag{21}$$

Since a convenient infrared-finite quantity to calculate, which can be easily compared with experiment, is the energy weighted accolinearity,[19,26] we study the energy-weighted transform $\mathscr{D}(Q^2,b)=\int_0^1 dx\,x\,\mathscr{D}(Q^2,b,x)$ rather than the general equation (21). Thus we have just a function of two variables satisfying

$$Q^2\frac{\partial}{\partial Q^2}\mathscr{D}(Q^2,b)=\int_0^1 dz\,z\left(\frac{\alpha_s(Q^2)}{2\pi}P_{qq}(z)\right)_+ \times J_0\left(Qb\sqrt{\frac{1-z}{z}}\right)\mathscr{D}\left(Q^2,\frac{b}{z}\right). \tag{22}$$

Then knowing $\mathscr{D}(Q^2=Q_0{}^2,b)$ at some Q_0, we want to evolve it to some higher Q^2. To solve this equation, Bassetto *et al.*[16] suggest the following procedure: let us write the right-hand side of this equation as

$$\frac{\alpha_s(Q^2)}{2\pi}\int_0^1 dz\,z(P_{qq}(z))_+\mathscr{D}(Q^2,b) + \frac{\alpha_s(Q^2)}{2\pi}\int_0^1 dz\,z\,P_{qq}(z) \times\left(\mathscr{D}\left(Q^2,\frac{b}{z}\right)J_0\left(Qb\sqrt{\frac{1-z}{z}}\right)-\mathscr{D}(Q^2,b)\right). \tag{23}$$

Then, as discussed by Baier and Fey,[12] z is set equal to 1 in each function of the second term, except in the explicit $1-z$ factors in P_{qq} and in the argument of the Bessel function, so that Eq. (22) becomes

$$Q^2\frac{\partial}{\partial Q^2}\mathscr{D}(Q^2,b) =\frac{\alpha_s(Q^2)}{2\pi}\mathscr{D}(Q^2,b)\left(-\gamma_2^0+\int_0^1 dz\,\frac{2C_2(F)}{1-z} \times\left(J_0(Qb\sqrt{1-z})-1\right)\right), \tag{24}$$

where γ_2^0 is the zeroth order $n=2$ anomalous dimension, $\gamma_2^0=16C_2(F)/3$, with the standard nonsinglet exponent $d_2^{\rm NS}\equiv\gamma_2^0/2\beta_0=32/(99-6N_f)$. It is then straightforward to see that the solution is

$$\mathscr{D}(Q^2,b)=\mathscr{D}(Q_0^2,b)\left(\frac{\alpha_s(Q^2)}{\alpha_s(Q_0^2)}\right)^{d_2^{\rm NS}}F(Q^2,Q_0^2,b)\,, \tag{25}$$

where the form factor

$$F(Q^2,Q_0^2,b)=\exp\left(\frac{2C_2(F)}{\pi}\int_{Q_0^2}^{Q^2}dk^2\,\frac{\alpha_s(k^2)}{k^2} \times\int_0^{kb}\frac{dy}{y}\,(J_0(y)-1)\right) \tag{26}$$

is just that of Parisi and Petronzio,[4] cf. Eqs. (12) and (13). For large b, this form factor has the behavior given by Rakow and Webber.[5] The $z\to1$ approximation generates their leading double logarithm result. However, the approximation of Bassetto *et al.*[16] is, of course, an asymptotic one. Though perfectly suitable for strictly theoretical considerations, at finite energies its use is quite misplaced. Experience from evaluating Eq. (13)[4,6] tells us that it is relatively small b that is important in inverting such Bessel transforms, e.g., Eq. (13), even at small Q_T or θ. Then the terms that generate the leading double logarithm result, no longer dominate, considerably affecting the phenomenology of Refs. 12 and 13.

To illustrate this, let us recast Eq. (22), guided by Eq. (24), in terms of $\mathscr{E}$ defined by

$$\mathscr{D}(Q^2,b)=\mathscr{E}(Q^2,b)(\alpha(Q^2))^{d_2^{\rm NS}}$$

so that

$$Q^2\frac{\partial}{\partial Q^2}\mathscr{E}(Q^2,b)=-\frac{\alpha_s(Q^2)}{2\pi}\int_0^1 dz\,z\,P_{qq}(z)\left(\mathscr{E}(Q^2,b) -\mathscr{E}\left(Q^2,\frac{b}{z}\right)J_0\left(Qb\sqrt{\frac{1-z}{z}}\right)\right). \tag{27}$$

This separation has the virtue that the integrand is clearly nonsingular with only real gluon emissions: the virtual graphs just contributing to the anomalous dimension term. The approximation of Bassetto *et al.* is to write the integrand of this equation simply as

$$(2C_2(F)/(1-z))\mathscr{E}(b,Q^2)(1-J_0(Qb\sqrt{1-z}))\,. \tag{28}$$

This would be good if $z\to1$ dominated; but unless Qb is large, it does not.

To illustrate this at its most simple, consider $\mathscr{E}(b,Q^2)$ to be a constant in b, e.g., 1. With $b=0.2$ and 1 GeV^{-1} the exact and approximate integrands are shown (with $Q_0=6$ GeV, as an example) in Fig. 3(a). We see the approximate integrand of Bassetto *et al.*[16] is not particularly peaked at $z=1$, and from Eq. (27) we see it will give the derivative wrong by a factor of two or more, and so the evolution starts off incorrectly.

This is exemplified by the behavior at $b=0$ of the function $\mathscr{D}(Q^2,b)$. Setting $b=0$ in either Eq. (22), or its approximate Eq. (24), we see $\partial\mathscr{E}(Q^2,0)/\partial Q^2=0$, so that

$$\mathscr{D}(Q^2,0)=\mathscr{D}(Q_0^2,0)\left(\frac{\alpha(Q^2)}{\alpha(Q_0^2)}\right)^{d_2^{\rm NS}} \tag{29}$$

for both. Now, however, let us consider the derivative at $b=0$ of each equation. It is easily checked that the approximation of Bassetto *et al.* gives, Eq. (28), $\partial^2\mathscr{E}(Q^2,b)/\partial\ln Q^2\partial b\,|_{b=0}=0$, which means

$$\frac{\partial\mathscr{D}}{\partial b}(Q^2,0)=\frac{\partial\mathscr{D}}{\partial b}(Q_0^2,0)\left(\frac{\alpha_s(Q^2)}{\alpha_s(Q_0^2)}\right)^{d_2^{\rm NS}},$$

whereas the exact result [either Eq. (22) or Eq. (27)] gives $\partial^2\mathscr{D}(Q^2,b)/\partial\ln Q^2\,\partial b\,|_{b=0}=0$, i.e.,

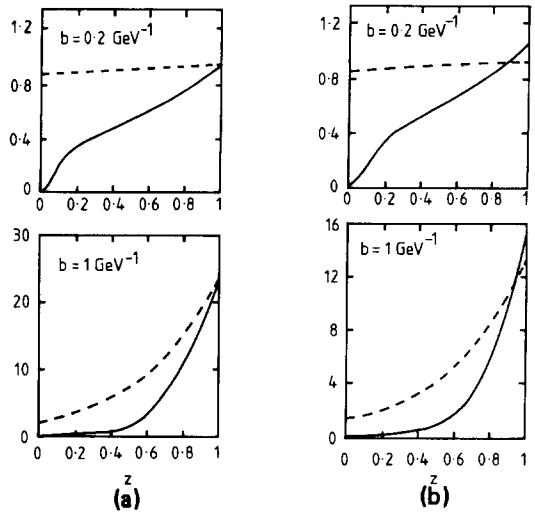


FIG. 3. The integrands of the evolution equation [Eq. (27), solid lines] and their approximation [Eq. (28), dashed lines] for (a) $\mathscr{E}(Q_0^2,b)=1$ and (b) $\mathscr{E}(Q_0^2,b)$ deduced from the phenomenological form Eq. (31). $Q_0=6$ GeV, $\Lambda=0.4$ GeV.

$$\frac{\partial \mathscr{D}}{\partial b}(Q^2,0)=\frac{\partial \mathscr{D}}{\partial b}(Q_0^2,0)\,. \tag{30}$$

Thus, just as Fig. 3(a) illustrates, the change with Q^2 is too great for the approximate solution for small b, which is the crucial region for determining the behavior of the QCD prediction.

Moreover, the difference shown in Fig. 3(a), between the exact and approximate integrands, is likely to be even greater in practice. At $z=1$, for example, the approximate integrand, Eq. (28), is just

$$\tfrac{1}{2}C_2(F)Q^2b^2\mathscr{E}(Q^2,b)\,,$$

while for the exact expression, Eq. (27), is

$$\frac{1}{2}C_2(F)\left(Q^2b^2\mathscr{E}(Q^2,b)-4b\frac{\partial}{\partial b}\mathscr{E}(Q^2,b)\right).$$

In the illustration of Fig. 3(a), with $\mathscr{E}=1$, this last derivative contributes nothing, but in general this will not be the case. Of course, if Qb is large enough, the integrands do peak at $z=1$, and the approximation of Bassetto *et al.*[16] is valid. However, we naturally want to evolve the fragmentation functions from some relatively low momentum, otherwise the predictions are already built into the starting function $\mathscr{D}(Q_0^2,b)$.

Since Eq. (19) is clearly tricky to solve analytically, we can evolve it numerically. As a suitable fragmentation function in the two jet regime, i.e., at $Q\sim 6$ GeV, we choose

$$D(Q^2,p_T,x)=(N/x)(1-x)^2\exp(-p_T^2/\langle p_T^2\rangle)\,. \tag{31}$$

Though such a form is expected for q or $\bar{q}$ separately and not for their nonsinglet combinations, it serves as a useful example. N then ensures momentum conservation, viz.

$$\int_0^1 dx\,x\int d^2p_T D(Q^2,p_T,x)=1\,, \tag{32}$$

which implies that the Fourier transform of the energy weighted function, viz. Eq. (22), is normalized so that $\mathscr{D}(Q^2,b=0)=1$. Note that the range of p_T [in Eqs. (31) and (32)] is limited by the energy fraction x so that[27,28] $x\geqslant x_T=2p_T/Q$.

With $\langle p_T^2\rangle=0.25$ GeV$/c^2$ at $Q_0=6$ GeV we have evolved the transform to $Q=30$ GeV and its square is shown in Fig. 4. As expected the true transform evolves more slowly than the approximation of Bassetto *et al.* In a practical calculation like that of the accolinearity $w(\theta)$, a coupled system of singlet quark and gluon fragmentation equations arises, with $w(\theta)$ given by[12]

$$\frac{dw}{d\sin^2\frac{1}{2}\theta}=\frac{1}{2}\int_0^\infty d(bQ)bQJ_0(bQ\sin\tfrac{1}{2}\theta)(\mathscr{D}(Q^2,b))^2\,, \tag{33}$$

where $\mathscr{D}$ is now a singlet quark distribution. Though this has zero anomalous dimension and so its value at $b=0$ is fixed with momentum [Eq. (29)], it evolves very similarly to that illustrated in Fig. 4 (the curves are just all normalized to one at $b=0$). Since the inversion, given by Eq. (33), needed to compare with experiment in the region $\cos\theta>0.97$,[26] picks out $b<0.3$ GeV^{-1}, we see this difference between the true and approximate transforms is crucial to a meaningful confrontation with data.

Though an analytic solution, like Eqs. (25) and (26), is not possible beyond the LLA, the numerical evaluation given above is straightforward. Moreover, the inclusion of the dominant higher-order terms in α_s by correcting the argument to k_T^2 from Q^2 in Eq. (19),[24] for which numerical solution is necessary anyway, produces no additional complications beyond those studied in Ref. 25. These effects are worthy of further study.[29]

5. RELAXING THE SMALL k_T APPROXIMATION

So far our discussion has been limited by the small transverse momentum approximation in which the full splitting function $\mathscr{P}_{qq}(x,x_T)$ is replaced by $P_{qq}(x)$. Kodaira and Trentadue[14,15] have attempted to go beyond this by including the terms in the splitting function that give more than the leading logarithm of Eq. (5), as well as higher-order terms $O(\alpha_s^2)$. By doing this Kodaira and Trentadue conclude that no agreement with experiment is possible. To see how they arrive at this conclusion by use of a misleading approximation let us drop the $O(\alpha_s^2)$ terms, which only complicate the discussion. To go beyond the leading logarithm approxima-

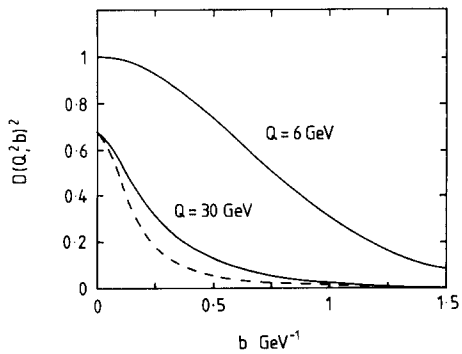


FIG. 4. The evolution of the Fourier transform of the energy-weighted fragmentation function $\mathscr{D}(Q^2,b)$ squared from $Q=6$–30 GeV according to Eq. (27) (solid curve) and its approximation Eq. (28) (dashed curve) as functions of b. $\Lambda=0.4$ GeV.

tion of Secs. 2 and 4, Kodaira and Trentadue include the dominant constant terms of Eq. (4). Thus they generalize Eq. (19) to the complete splitting function $\mathscr{P}(x,x_T)$, but approximate this to generate just the terms [Eq. (6)]

$$\ln(4/x_T^2) - \tfrac{3}{2}.$$

This is then transformed into the conjugate impact parameter space to give $T(b)$, the exponential of which generalizes the form factor of Eq. (26) on solving the resulting evolution equation by the method of Bassetto *et al.* This b-space form factor is expected to look rather like $\mathscr{E}(Q^2,b)^2$, cf. Fig. 4, which it generalizes. However, Kodaira and Trentadue[15] find the function drawn in Fig. 5, which is quite different. Notice that, while the form factor of Fig. 4 falls monotonically, that of Kodaira and Trentadue increases at small b and then falls. It is this strange behavior of the b-space form factor which gives the predictions that Kodaira and Trentadue find inconsistent with experiment. How does this happen?

The full splitting function $\mathscr{P}_{qq}(x,x_T)$, being proportional to a physical cross section, is positive definite (except at $x_T = 0$ where the virtual graphs contribute). As a consequence, the logarithm of the b-space form factor [Eq. (26), or $B(b)$ or Eq. (12), or $T(b)$ of Ref. 15] is negative definite, since $J_0(y) < 1 \forall y > 0$. However, the behavior in Fig. 5 means $T(b)$ of Kodaira and Trentadue has a positive excursion for $b < 0.2\ \text{GeV}^{-1}$. This is wholly due to the $-\frac{3}{2}$ of Eq. (6). This is indeed the next term in the expansion of the basic single gluon emission cross section, Eq. (4) or (7) for $k_T^2 \ll Q^2$. However, the terms which are negligible for k_T small ensure it remains positive as k_T increases. Thus the next term in Eq. (6) would give, from Eq. (4).

$$\frac{1}{2}\left(\left(1 - \frac{k_T^2}{Q^2} + 4\frac{k_T^4}{Q^4}\right)\ln\left(\frac{Q^2}{k_T^2}\right) - \frac{3}{2} + 3\frac{k_T^2}{Q^2} - 2\frac{k_T^4}{Q^4} - \frac{65}{3}\frac{k_T^6}{Q^6}\cdots\right).$$

Now it is the behavior of this cross section at large k_T that determines the Fourier transform at small Qb. There Kodaira and Trentadue[14,15] still use a small k_T approximation [Eq. (6)] where it surely cannot be valid. Not surprisingly, misleading conclusions then result.[30]

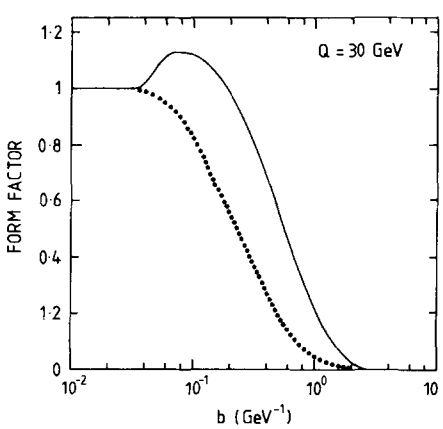


FIG. 5. The form factor $F(Q^2,Q_0^2,b)$ at $Q = 30$ GeV from Kodaira and Trentadue[15] (the curve shown solid), with $\Lambda = 0.2$ GeV, compared with that given by Eq. (26) (the dotted curve).

As already remarked, experience from evaluating Eqs. (12) and (26)[4,6] tells us small b is the dominant region. The behavior is largely determined there by the form of the single gluon cross section at large x_T. There perturbation theory straightforwardly applies[4] and the cross section, being physical, is positive definite. Consequently, the b-space form factor is definitely less than one. Predictions for the accolinearity, or energy-energy correlation, in e^+e^-,[19,26] depend on the details of how this form factor falls from one as b increases, e.g., Fig. 4. This should provide a valuable test of perturbative QCD.[4]

6. CONCLUSIONS

Leading logarithms in perturbative QCD have a universal, process-independent origin. Their study provides a most useful theoretical approximation yielding many important results.[1] However, in any confrontation of these with experiment, care must be taken to apply results so obtained only in kinematic regimes where such approximations are valid. To do otherwise is shown to lead to misleading results and produce erroneous conclusions on the ability of perturbative QCD to describe available data. In most present phenomenology, nonleading terms dominate and the leading logarithm approximation is quite misleading. The K-factor in Drell–Yan processes,[3,22] and its analog in other reactions (Sec. 3), dramatically affects their normalization at all foreseeable energies. The inclusion of nonleading terms usually means analytic solutions are no longer possible: for example, the fragmentation function formulas of Bassetto, Ciafaloni, and Marchesini.[16] In Sec. 4, we have shown how their leading logarithm approximation fails at finite momenta. Nevertheless, such evolution equations are readily solved numerically, generating quite different results.

In almost all physical problems, approximations are of course necessary; approximations that can be shown to be invalid in some regime. Nonetheless, however pathological the approximation there, it is often of no consequence for the original problem. The examples given here differ[31] in that they show how the leading logarithm approximation is dangerous in tackling presently relevant problems at presently accessible momenta. We must conclude that leading logs lead theoretically, but mislead phenomenology.

ACKNOWLEDGMENTS

It is a pleasure to thank Alan Martin for stimulating my interest in these questions, and other members of the Durham physics graduate workshop for discussions. I gratefully acknowledge the support of the U.K. Science and Engineering Research Council.

[1] E.g., Y. L. Dokshitser, D. I. D'yanakov, and S. I. Troyan, Phys. Rep. **58**, 269 (1980).
[2] J. M. Cornwall and G. Tiktopoulos, Phys. Rev. D **13**, 3370 (1976); G. Pancheri and Y. Srivastava, Phys. Rev. Lett. **43**, 11 (1979); G. Curci, M. Greco, and Y. Srivastava, Phys. Rev. Lett. B **43**, 834 (1978), Nucl. Phys. B **159**, 451 (1979); C. Y. Lo and J. D. Sullivan, Phys. Lett. B **86**, 327 (1979).

[3]G. Parisi, Phys. Lett. B **90**, 295 (1980).
[4]G. Parisi and R. Petronzio, Nucl. Phys. B **154**, 427 (1979).
[5]S. D. Ellis and W. J. Stirling, Phys. Rev. D **23**, 214 (1981); S. D. Ellis, N. Fleishon, and W. J. Stirling, Phys. Rev. D **24**, 1386 (1981); P. E. L. Rakow and B. R. Webber, Nucl. Phys. B **187**, 254 (1981); H. F. Jones and J. Wyndham, J. Phys. A **14**, 1457 (1981).
[6]M. R. Pennington, Nucl. Phys. B **204**, 189 (1982).
[7]R. P. Feynman, *Les Houches Lectures Session XXIX, 1976*, edited by R. Balian and C. H. Llewellyn-Smith (North-Holland, Amsterdam, 1977), p. 124.
[8]G. C. Fox and S. Wolfram, Nucl. Phys. B **168**, 285 (1980); S. Wolfram, *Proceedings of the Fifteenth Rencontre de Moriond* (March, 1980), edited by J. Tran Thanh Van; R. D. Field, *Proceedings of the Conference on Perturbative QCD*, Tallahassee, AIP Conference Proceedings, Vol. 74, edited by D. W. Duke and J. F. Owens (AIP, New York, 1981), p. 286; R. D. Field and S. Wolfram, University of Florida Preprint UFTP 82-12; R. D. Field and G. C. Fox, University of Florida Preprint UFTP 82-30.
[9]G. C. Fox, *Proceedings of the SLAC Summer Institute on Particle Physics, 1981*, edited by A. Mosher, SLAC Report No. 245, 141 (1982); P. Hoyer, P. Osland, H. G. Sander, T. F. Walsh, and P. M. Zerwas, Nucl. Phys. B **161**, 349 (1979).
[10]F. Halzen and D. M. Scott, Phys. Lett. B **78**, 318 (1978); C. L. Basham, L. S. Brown, S. D. Ellis, and S. T. Love, Phys. Lett. B **85**, 297 (1979), Phys. Rev. D **19**, 2019 (1979); B. Humpert and W. L. van Neerven, Phys. Lett B **85**, 293 (1979); J. Finjord, G. Girardi, and P. Sorba, Phys. Lett. B **89**, 99 (1979); J. Cleymans, M. Dechansreiter, F. Halzen, and D. M. Scott, Phys. Lett. B **89**, 403 (1980); R. W. Brown and K. O. Mikaelian, Phys. Rev. D **19**, 922 (1979); K. Konishi, A. Ukawa, and G. Veneziano, Phys. Lett. B **80**, 259 (1979); F. Rapuano, Lett. Nuovo Cimento **26**, 219 (1979); F. Halzen and D. M. Scott, Phys. Rev. D **21**, 131 (1980), *Proceedings of the International Symposium on Multiparticle Dynamics*, Bruges, 1980, edited by E. deWolf and F. Verbeure (University of Antwerp, Antwerp, 1980), p. 593; Phys. Lett. B **94**, 405 (1980); P. Aurenche and J. Lindfors, Phys. Lett. B **96**, 171 (1980); Nucl. Phys. B **185**, 301 (1981); B. Humpert and W. L. van Neerven, Phys. Lett. B **93**, 456 (1980); M. Chaichian, M. Hayashi, and M. Roos, Helsinki Preprint TFT 81 10, 1981; J. C. Collins and D. E. Soper, Nucl. Phys. B **193**, 381 (1981), University of Oregon Preprint OITS 178, Nov. 1981; F. Halzen, A. D. Martin, D. M. Scott, and M. Dechansreiter, Phys. Lett. B **106**, 147 (1981); P. Chiapetta and M. Greco, Phys. Lett. B **106**, 219 (1981); F. Halzen, A. D. Martin, and D. M. Scott, Phys. Rev. D **25**, 754 (1982); F. Halzen, A. D. Martin, D. M. Scott, and M. P. Tuite, Z. Phys. C **14**, 351 (1982).
[11]F. Halzen, A. D. Martin, and D. M. Scott, Phys. Lett. B **112**, 160 (1982).
[12]R. Baier and K. Fey, Nucl. Phys. B **179**, 49 (1981).
[13]J. Cleymans and M. Kuroda, Nucl. Phys. B **189**, 483 (1981).
[14]J. Kodaira and L. Trentadue, Phys. Lett. B **112**, 66 (1982).
[15]J. Kodaira and L. Trentadue, SLAC Preprints SLAC-PUB-2861, Dec. 1981 and SLAC-PUB-2934, June 1982.
[16]A. Bassetto, M. Ciafaloni, and G. Marchesini, Nucl. Phys. B **163**, 477 (1980).
[17]J. Ellis, M. K. Gaillard, and G. G. Ross, Nucl. Phys. B **111**, 253 (1976).
[18]We will consider below what happens if $k_T(x_T)$ is measured relative to some thrust axis.
[19]E.g., H. Meyer, *Proceedings of the Conference on Perturbative QCD*, Tallahassee, AIP Conference Proceeding, Vol. 74, edited by D. W. Duke and J. F. Owens (AIP, New York, 1981), p. 338.
[20]L. S. Brown and S. D. Ellis, Phys. Rev. D **24**, 2383 (1981).
[21]T. Appelquist and H. Georgi, Phys. Rev. D **8**, 4000 (1973); A. Zee, Phys. Rev. D **8**, 4038 (1973).
[22]C. H. Llewellyn-Smith, *Proceedings of the XXth International Conference on High Energy Physics*, Madison, AIP Conference Proceedings, Vol. 68 (AIP, New York, 1980), p. 1345.
[23]F. Halzen and D. M. Scott, Ann. Phys. (NY) **135**, 1 (1981).
[24]S. J. Brodsky and G. P. Lepage, *Proceedings of the SLAC Summer Institute on Particle Physics*, 1979, edited by A. Mosher, SLAC Report No. 224, 133, 1980; D. Amati, A. Bassetto, M. Ciafaloni, G. Marchesini, and G. Veneziano, Nucl. Phys. B **173**, 429 (1980).
[25]M. R. Pennington, R. G. Roberts, and G. G. Ross, Phys. Lett. B **120**, 204 (1983).
[26]C. Berger *et al.* (PLUTO Collaboration) Phys. Lett. B **99**, 292 (1981).
[27]There is a confusion in the literature about a factor of $\frac{1}{2}$ in the argument of the Bessel function J_0 in the b-space transform. This factor is seen in Eqs. (11) and (12) as it should consistently be, but to allow comparison with the works of Refs. 12–16 it is dropped here. This amounts to a trivial redefinition of b in Secs. 4 and 5 compared to Sec. 3. However, care must be taken in inversion, cf. Eq. (13).
[28]However, Baier and Fey[12] take $p_{T_{max}}$ to infinity regardless of the value of x. Again, this would be all right if $x \to 1$ dominated, but as seen from Eq. (32) it cannot. This inappropriate approximation leads to a spurious peaking of the Fourier transform at $b = 0$. The phenomenology of Baier and Fey unfortunately depends critically on this—in fact more than on the approximate evolution equation [Eq. (25)] they also use.
[29]This is being performed by P. J. Done, M.Sc. thesis submitted to University of Durham, U.K., 1983.
[30]The numerical techniques developed and exploited in Refs. 8 and 9, for example, avoid all these pitfalls and give phenomenologically meaningful results.
[31]I am grateful to Professor Roy Chisholm for encouraging me to emphasize this point.

Collectivity and geometry. I. General approach

Marcos Moshinsky[a)]
Instituto de Física, UNAM, Apdo. Postal 20-364, 01000 México, D.F., Mexico

(Received 3 March 1983; accepted for publication 22 July 1983)

In the last decade an extensive literature appeared in which the microscopic collective behavior of nuclei was associated with definite irreducible representations (irreps) of either the O(n) or Sp(6) groups, where $n = A - 1$ and A is the number of nucleons. It became clear that the two approaches are equivalent, as problems with $3n$ degrees of freedom are characterized by a definite irrep of the group Sp($6n$) and for its subgroup Sp(6)$\times$O(n) the irrep of O(n) determines that of Sp(6) and vice versa. Thus one can consider that collective effects appear when one introduces the constraint that in the many-body system the states are restricted to a definite irrep of O(n) [and thus also of Sp(6)] and the Hamiltonians are in the enveloping algebra of Sp(6) rather than in that of Sp($6n$). Once Sp(6) becomes the paramount group of collective motions, the problem is to determine the matrix elements of the generators of Sp(6) in a basis characterized by irreps of its subgroups. What subgroups to choose? Rowe and Rosensteel have taken Sp(6)$\supset$U(3) and Sp(6)$\supset$CM(3), where the latter has also been considered by Biedenharn *et al.* In the present series of papers we analyze the problem in the chain Sp(6)$\supset$Sp(2)$\times\mathscr{O}$(3), as we show that in the boson limit, i.e., when $n \gg 1$, the Casimir operator of Sp(2) goes into the Casimir operator of U(5), i.e., the corresponding chain is U(6)$\supset$U(5)$\supset\mathscr{O}$(3). In the case Sp(6)$\supset$U(3)$\supset\mathscr{O}$(3), the boson limit is U(6)$\supset$U(3)$\supset\mathscr{O}$(3). Thus in this series of papers we look at the microscopic collective model from what could be called the vibrational rather than the rotational point of view.

PACS numbers: 21.60.Ev, 21.60.Fw, 03.65.Ge

1. INTRODUCTION

In this, the first of a series of papers, we wish to discuss the relation between collective behavior in many-body systems and geometrical concepts. Specifically, we shall reexamine the microscopic description of nuclear collective motions and their relations with the symplectic geometry of the A nucleon system.

The first widely succesful way of introducing collective degrees of freedom in the nucleus was the liquid drop model of Niels Bohr.[1] In it the surface of the drop is given in spherical coordinates by $r = f(\theta,\varphi)$, and the right-hand side is developed in spherical harmonics with the coefficients becoming the collective degrees of freedom. In the more systematic approach developed years later by Bohr and Mottelson,[2] the drop is restricted to quadrupole deformations, and the fluid is taken as incompressible so that the equation of the surface becomes

$$r = r_0\left[1 + \sum_m \alpha^m Y_{2m}(\theta,\varphi)\right], \tag{1}$$

with $\alpha^m = (-1)^m \alpha_{-m}$, $m = 2,1,0,-1,-2$, being the collective degrees of freedom. The α_m are implicitly related[2,3] to the mass quadrupole of the many nucleon system defined by

$$q_m = \sum_{\tau,\tau'} \langle 1\tau,1\tau'|2m\rangle \sum_{s=1} x_{\tau s} x_{\tau' s}, \tag{2}$$

where $\langle\ |\ \rangle$ is a Clebsch–Gordan coefficient and $x_{\tau s}$, $\tau = 1,0,-1$, $s = 1,2,\ldots,n$, are the spherical components of the $n = A - 1$ relative Jacobi vectors of the A nucleon system as, from the beginning, we wish to eliminate the center-of-mass coordinate which is irrelevant for the analysis. It is immediately clear that the q_m defined by (2) is a scalar with respect to the O(n) group of rotations in the space associated with the index $s = 1,2,\ldots,n$. This point will be central to much of the following discussion.

In view of the great success of the Bohr-Mottelson model in explaining many of the features of nuclear structure, a very extensive literature developed trying to justify it microscopically. From the standpoint of this series of papers, we are concerned only with that part of the literature based on group-theoretical considerations, which we shall proceed to review briefly so as to place our approach in the context of the work done previously in this field.

Probably the first publications that develop the point of view in which we are interested are those of Goshen and Lipkin.[4] In the first one, that appeared in 1959, they consider an n-body system in one-dimensional space and indicate how from the symplectic $2n$-dimensional group Sp(2n) of the problem, one can pass to the group Sp(2)$\times$O(n), where the latter is the orthogonal group in n-dimensions. Collective excitations are then related to the Sp(2) group and, in particular, the collective Hamiltonian is associated with the O(2) subgroup of the latter. In a subsequent paper they extend the analysis to two space dimensions, i.e., Sp($4n$)$\supset$Sp(4)$\times$O(n). In several respects the papers of Goshen and Lipkin provide the foundations on which many of the group-theoretical microscopic collective theories were based, including the one to be presented in this series of papers. They did not provide, though, a detailed formalism for carrying out nuclear structure calculations.

By the early seventies a considerable push was given to the development of group-theoretical microscopic collective

[a)] Member of El Colegio Nacional and Instituto Nacional de Investigaciones Nucleares.

 0022-2488/84/051555-10$02.50

models through the work of Zickendraht[5] and Dzublik *et al.*[6] in which they introduced a transformations of coordinates that take us from the $3n$ variables x_{is}, $i = 1,2,3$, $s = 1,...,n$, to six collective degrees of freedom (three deformation parameters along the principal axis and three Euler angles) and $3n - 6$ variables associated with single-particle excitations. Almost immediately the group of Filippov[7] in Kiev and of Vanagas[8] in Vilnus realized that these transformations allowed one to discuss collective effects by restricting the states to a definite irreducible representation (irrep) of the $O(n)$ group which could be derived from shell model consideration.[7,8]

A problem of $3n$ degrees of freedom can be associated with a definite irrep of a dynamical group $Sp(6n)$, which among its subgroups has $Sp(6)\times O(n)$. It was shown by Moshinsky and Quesne[9] in 1971 that the irrep of the group $O(n)$ determines that of Sp(6), i.e., that they are "complementary,"[9] and thus the procedure of Filippov and Vanagas restricted to a specific irrep of $O(n)$ implied also a restriction to a definite irrep of Sp(6), as these authors realized in their later publications.[10,11]

By the middle seventies Rosensteel and Rowe[12] in Toronto and Bidenharn, Buck, Cusson, and Weaver[13] at Duke initiated an approach to the problem by first identifying the desired collective motions and then determining the operators that generate these motions as well as the Lie algebra that they satisfy. Thus appeared the collective motion Lie algebra[12,13] known as cm(3) as well as the explicit determination of the vortex spin operator which plays a very important role in coupling rotational motion with internal dynamics.[14]

Rosensteel and Rowe[15] and their collaborators quickly realized that the CM(3) group was a subgroup of Sp(6) [which they call $SP(3,R)$] and thus the basis for the irreps of the latter were fundamental for implementing their views on the microscopic origin of collective motions. A natural chain for discussing these basis is $Sp(6)\supset U(3)$ where the latter is the unitary group in three dimensions introduced by Elliott.[16] Rowe and Rosensteel then proceeded to find the basis for the irreps $Sp(6)\supset U(3)$ both in an abstract fashion and as shell model states, and to carry an extensive program of calculations[17] using these basis as well as discussing the implications for the states characterized by irreps in the chain $Sp(6)\supset CM(3)$.[12,13]

The remark made previously[9] about the relation between the irreps of Sp(6) and $O(n)$, when $Sp(6)\times O(n)$ is considered as a subgroup of a definite irrep of $Sp(6n)$, immediately relates the work done in Toronto and Duke to that carried out in Vilnus and Kiev, as was quickly clear to all the authors involved.[10,11,14,15] Thus the Sp(6) group became the paramount one for the microscopic analysis of nuclear collective motions, and its seemed that the discussion of the previously mentioned authors, and, in particular, the very extensive analysis of Rowe and Rosensteel,[15] provided all the background necessary for calculations in nuclear structure.

The reason that, in the present series of papers, the author and his collaborators are looking again at the Sp(6) problem stems from their experience[18–20] both in the Bohr–Mottelson model[2] and the interacting boson approximation (IBA)[21] introduced by Arima and Iachello. In particular in the latter, where pairs of nucleons are associated with s- and d-bosons, the fundamental group is U(6) and it admits, among others, the chain of subgroup $U(6)\supset U(3)\supset \mathcal{O}(3)$ and $U(6)\supset U(5)\supset \mathcal{O}(3)$ associated, respectively, with rotational and vibrational limits.[21] Now Deenen and Quesne[22] showed that for closed shells or, equivalently, scalar representation of $O(n)$, the states that are basis for irreps of $Sp(6)\supset U(3)$ can be put into one-to-one correspondence with basis states characterized by irreps of $U(6)\supset U(3)$ in the boson approximation. A similar result was proved for open shells, where now there is vortex spin, by Castaños and Frank,[23] but only in the limit when the number of nucleons $A\to\infty$. Thus the procedures of Rosensteel and Rowe are closely correlated with the $U(6)\supset U(3)$ chain in the boson approximation, as they themselves have stressed.[24]

Is there something equivalent in the Sp(6) model of the $U(6)\supset U(5)\supset \mathcal{O}(3)$ chain in the boson model? We shall show in the present paper that this is the case if we consider the $Sp(6)\supset Sp(2)\times \mathcal{O}(3)$ chain which, to our knowledge, has not been fully discussed before. Thus it becomes interesting to analyze the basis for the irreps in the chain $Sp(6)\supset Sp(2)\times \mathcal{O}(3)$ as well as the matrix element of the generators of Sp(6) in this basis. This will be the main objective of the present series of papers.

Before presenting a summary of this and the following publications, we would like to stress that the developments outlined above for a group-theoretical approach to a microscopic theory of collective motions are not contradictory but complementary. As we indicated, all of them can be related to the Sp(6) group, though some analysis emphasize the complementary group $O(n)$,[10,11] while others stress particular subgroups of Sp(6) whose generators are relevant to collective operators.[12,14,15] Thus one could affirm that the bare bones ideas introduced almost a quarter of century ago[4] have been fleshed out in a variety of ways that illuminate the microscopic nature of nuclear collective motions and help us also to understand better macroscopic collective models such as the Bohr–Mottelson one and the interacting boson approximation.

2. SUMMARY

We start our analysis in Sec. 3 by briefly outlining an analogy in Euclidean geometry which we think is helpful in understanding the nature of the problem. In Sec. 4 we then review the symplectic geometry for a problem of $3n$ degrees of freedom whose dynamical group is $Sp(6n)$, emphasizing the subgroup $Sp(6)\times O(n)$. In Sec. 5 we introduce our definition of collectivity as a restriction to a definite irrep of $O(n)$ for the states and to expressions in the enveloping algebra of Sp(6) for the Hamiltonians. As discussed in the Introduction, this definition coincides with those of other authors,[7,8,15] but in Sec. 6 we indicate that our procedure for calculations is based on the chain of subgroups $Sp(6)\supset Sp(2)\times \mathcal{O}(3)$. After a brief detour to the one-dimensional problem in Sec. 7, we tackle in Sec. 8 the three-dimensional case in the limit when $n\to\infty$. We show that the Casimir operator of the Sp(2) subgroup goes then into a collective Hamiltonian of the Bohr–

Mottelson type. As the latter can be formally associated with d-bosons only, and thus, with a U(5) symmetry group, we see that $Sp(6) \supset Sp(2) \times \mathcal{O}(3)$ in the boson limit (i.e., when $n \gg 1$) corresponds to the $U(6) \supset U(5) \supset \mathcal{O}(3)$ chain of groups, rather than to $U(6) \supset U(3)$, which is the boson limit for the $Sp(6) \supset U(3)$ approach followed by Rosensteel and Rowe.[15,24] Also in Sec. 8 we use a complete set of states[25] for microscopic collective models (again when $n \gg 1$), to give the matrix elements of the generators of Sp(6), and thus fully implement our program when the number of nucleons is large. In Sec. 9 we outline the program for the determination of the states and matrix elements for arbitrary n (and not as before for $n \gg 1$ only), and this program will be implemented in the second and third papers of this series in spaces of two and three dimensions, i.e., for the chains of groups $Sp(4) \supset Sp(2) \times \mathcal{O}(2)$ and $Sp(6) \supset Sp(2) \times \mathcal{O}(3)$.

3. ANALOGY IN ELEMENTARY GEOMETRY

Let us now consider an elementary problem which will provide a useful analogy for our future discussions. If we look at the three-dimensional Euclidean space of coordinates (x, y, z) we have in it the simple geometrical concepts and theorems discussed in Euclid's book. If we introduce a constraint in this space by the relation $x^2 + y^2 + (z - a)^2 = a^2$, we have now the more complex geometry of a sphere despite the fact that we are dealing with a two- rather than a three-dimensional manifold. We recover the simple Euclidean geometry, but now in two rather than three dimensions, for the points on the sphere near the origin when $a \to \infty$, as then that part of the sphere becomes essentially the x–y plane.

4. SYMPLECTIC GEOMETRY

Let us turn now our attention to the $A = n + 1$ particle system in a d-dimensional space, which we will later particularize to $d = 1,2$ as well as the physical case $d = 3$. Again eliminating the center of mass, we have the Jacobi coordinates and momenta

$$x_{is}, p_{is}, \quad [x_{is}, p_{jt}] = i\delta_{ij}\delta_{st}, \quad i,j = 1,...,d, \; s,t = 1,...,n, \tag{3}$$

which are the elements of a dn-dimensional Weyl Lie algebra $W(dn)$. We shall use in this paper units in which $\hbar$, the mass of the particles, and an appropriate frequency take the value 1.

The Hermitian quadratic expressions in coordinates and momenta

$$x_{is}x_{jt}, \quad x_{is}p_{jt} + p_{jt}x_{is}, \quad p_{is}p_{jt}, \tag{4}$$

which close under commutation,[26] provide us with the $dn(2dn + 1)$ generators of the symplectic group $Sp(2dn)$.

We now indicate a, purely conceptual, way in which we could use the $Sp(2dn)$ group to solve the many-body problem. We first need a complete set of states which can be provided by the eigenstates of the Hamiltonian

$$H_0 = \sum_{s=1}^{n} \sum_{i=1}^{d} H_{is}, \quad H_{is} = \tfrac{1}{2}(p_{is}^2 + x_{is}^2), \tag{5}$$

where the dn operators H_{is} are particular combinations of the generators (4) of $Sp(2dn)$ that commute among themselves, i.e., $[H_{is}, H_{jt}] = 0$, thus giving us the weight generators[27] of this group. As H_0 is invariant under the permutation of the $A = n + 1$ nucleons, the eigenstates can be characterized further by the irreducible representation (irrep) of symmetric group S_{n+1} and its subgroups, i.e., by a partition $\{f\}$ of $n + 1$ and the Yamanouchi symbol (r). Furthermore, in this basis we can calculate the matrix elements of all the generators (4) of the $Sp(2dn)$ group.

Now an arbitrary Hamiltonian H involving central forces (to which we shall restrict ourselves for simplicity) is in the enveloping algebra of $Sp(2dn)$, as it will be a function of the quadratic expressions in (4) which are invariant under space reflections, rather than on the linear x_{is}, p_{is}, which are not. This also applies to other integrals of motion such as the components of the total angular momentum or functions of them. Thus, if we have the matrix elements of the generators of $Sp(2dn)$ in the basis of eigenstates of H_0, we know, in principle, the matrix representation of an arbitrary Hamiltonian H and its integrals of motion. The diagonalization of these matrices, which we make finite by taking an upper bound to the number of quanta in the eigenstates of H_0, provide us with energy levels characterized by eigenvalues of the integrals of motion.

The procedure indicated in the previous paragraph is, of course, highly impractical, but it provides us with a framework which we shall use later when we restrict ourselves to the collective degrees of freedom.

Let us now turn our attention to the dn-dimensional oscillator whose Hamiltonian is the H_0 of (5). Its dynamical group is, of course,[26] $Sp(2dn)$ whose generators are given by (4) and all the states of even (odd) number of quanta N belong to the irrep $[\frac{1}{2}^{dn}]$ $([\frac{1}{2}^{dn-1}\frac{3}{2}])$ of this group, as can be seen when we apply H_{is} of (5) to the lowest weight state, i.e., for even (odd) number of quanta the $N = 0$ (1) state.

The subgroups of the dynamical group $Sp(2dn)$ allow us to classify further the eigenstates of H_0. As we show later, the relevant subgroup turns out to be

$$Sp(2dn) \supset Sp(2d) \times O(n), \tag{6}$$

whose generators can be obtained from (4) if we contract with respect to the indices i or s. For $O(n)$ these generators[27] have the well-known form

$$\mathcal{L}_{st} = \sum_{i=1}^{d} (x_{is}p_{it} - x_{it}p_{is}), \tag{7}$$

and there are $(n/2)(n - 1)$ of them with the quadratic Casimir operator being given by $\mathcal{L}^2 = \frac{1}{2}\Sigma_{st}\mathcal{L}_{st}^2$. For $Sp(2d)$ there are $d(2d + 1)$ Hermitian generators of the form[26,15]

$$q_{ij} = \sum_{s=1}^{n} (x_{is}x_{js}), \tag{8a}$$

$$S_{ij} = \frac{1}{2}\sum_{s=1}^{n} (x_{is}p_{js} + x_{js}p_{is} + p_{js}x_{is} + p_{is}x_{js}) = i[H_0, q_{ij}], \tag{8b}$$

$$T_{ij} = \sum_{s=1}^{n} (p_{is}p_{js}) = q_{ij} - \tfrac{1}{2}[H_0, [H_0, q_{ij}]], \tag{8c}$$

$$L_{ij} = \sum_{s=1}^{n} (x_{is}p_{js} - x_{js}p_{is}), \tag{8d}$$

where from the commutation relation (3) we see that S_{ij}, T_{ij}; $i,j = 1...d$, can be expressed in terms of commutators of q_{ij} with H_0. The generators in (8d) are those of the $\mathcal{O}(d)$ subgroup of Sp($2d$).

As we are dealing only with the irreps $[\frac{1}{2}^{dn}]$ or $[\frac{1}{2}^{dn-1}\frac{3}{2}]$ of Sp($2dn$) it is clear[9,28] that the irreps of Sp($2d$) and O(n) are complementary, i.e., if we fix the irrep of O(n) that of Sp($2d$) is given or vice versa. This is due to the fact that the Casimir operators of Sp($2d$) can be expressed as polynomial functions of those of O(n) or vice versa.

We now look at the type of irreps that we can have for O(n) in the eigenstates of H_0. The symmetry group of H_0 is U(dn), and it admits the subgroup chain

$$\mathrm{U}(dn) \supset \mathcal{U}(d) \times \mathrm{U}(n), \tag{9}$$

where in turn O(n), whose generators are given by (7), is a subgroup of U(n). As all states will be characterized by symmetric irreps $[N]$ of N quanta for U(dn), $\mathcal{U}(d)$, and U(n) are also complementary.[9,28] Assuming $d < n$ (which will be the case for most situations of interest as in physical space $d = 3$), we have that the irreps of both $\mathcal{U}(d)$ and U(n) are characterized by partitions of N in d numbers, i.e., $[h_1,h_2\cdots h_d]$. As O(n) is a subgroup of U(n), its most general irrep will also be given by a partition in d numbers $(\omega_1,\omega_2\cdots\omega_d)$, where $\omega_i \leqslant h_i$, $i = 1,...,d$. Thus we have characterized the irrep of O(n) and, because of the complementary relations, also that of Sp($2d$).

The eigenstates of H_0 are determined by dn quantum numbers. We wish now to characterize these states by an irrep $(\omega_1\cdots\omega_d)$ of O(n) and thus also by a definite irrep of Sp($2d$). We first inquire how many of the dn quantum numbers are associated with subgroups of the O(n) group. For this purpose let us consider the chain O(n) $\supset$O($n-1$)$\supset\cdots\supset$O(2) whose irreps can be arranged in a Gel'fand pattern[29]

$$\begin{array}{ll}
\mathrm{O}(n) & \omega_{1,n}\ \omega_{2,n}\ \cdots\cdots\cdots\ \omega_{d,n}\ 0\ \cdots\cdots\ 0 \\
\mathrm{O}(n-1) & \omega_{1,n-1}\ \cdots\cdots\cdots\ \omega_{d,n-1}\ 0\ \cdots\ 0 \\
\cdots & \cdots\cdots\cdots\cdots \\
\mathrm{O}(2d) & \omega_{1,2d}\ \cdots\cdots\ \omega_{d,2d} \\
\mathrm{O}(2d-1) & \omega_{1,2d-1}\ \cdots\cdots\ \omega_{d-1,2d-1} \\
\mathrm{O}(2d-2) & \omega_{1,2d-2}\ \cdots\ \omega_{d-1,2d-2} \\
\cdots & \cdots\cdots\cdots \\
\mathrm{O}(3) & \omega_{13} \\
\mathrm{O}(2) & \omega_{12}
\end{array}, \tag{10}$$

where to the left we indicate the orthogonal subgroup to which the irrep in each row belongs and we have added the index n to the ω_i, $i = 1,2,\cdots,n$ of the irrep of O(n). Clearly the total of quantum numbers is given by

$$(n - 2d + 1)d + \sum_{m=1}^{d-1} m = nd - d^2. \tag{11}$$

Thus the irreps of the subgroups of Sp($2d$), plus the multiplicity indices, must supply the missing d^2 quantum numbers. In an independent fashion[15,30] it is also possible to see that the total of quantum numbers required to characterize the most general irrep of Sp($2d$) is d^2.

5. DEFINITION OF COLLECTIVITY

So far we have been talking about mathematical relations. We wish now to address ourselves to the physics and ask how collectivity appears within this framework. Our definition will be that collective states are restricted to a single irrep of O(n)—and thus also of Sp($2d$)—and that collective Hamiltonians will be defined in the enveloping algebra of Sp($2d$), rather than in that of Sp($2dn$), i.e., will be functions of the operators (8) which, furthermore, must be invariant under the subgroup $\mathcal{O}(d)$.

The reasons for this definition are the following: First we can assume an oscillator shell model as the starting point in our analysis. The ground state for A nucleons (for which now $d = 3$), is formed when we fill compactly the levels of the oscillator to get a state of definite number of quanta N. These states can be characterized by irreps $(\lambda\ \mu)$ of SU(3) which correspond to irreps of U(3), and thus also of U(n), given by

$$[h_1h_2h_3], \tag{12a}$$

where

$$h_1 = \tfrac{1}{3}[(2\lambda + \mu) + N],$$
$$h_2 = \tfrac{1}{3}[(\mu - \lambda) + N], \quad h_3 = \tfrac{1}{3}[-(\lambda + 2\mu) + N]. \tag{12b}$$

As the levels are compactly filled,[25] we have that the irreps of O(n), which for $d = 3$ are given by $(\omega_1\ \omega_2\ \omega_3)$, become $(\omega_1\ \omega_2\ \omega_3) = [h_1h_2h_3]$. Thus ^{4}He, ^{16}O, ^{20}Ne, for which, respectively, the N and the irreps $(\lambda\ \mu)$ of SU(3) are given by 0, 12, 20 and (0,0), (0,0), (8,0), correspond to irreps $(\omega_1,\ \omega_2\ \omega_3)$ of O(n) that are (0,0,0), (4,4,4), (12,4,4). Thus the shell model for the ground state fixes the irrep of O(n), to which we restrict ourselves in collective excitations.

Clearly this procedure for determining $(\omega_1\ \omega_2\ \omega_3)$ does not hold for heavy nuclei above the 2s, 1d shell where the strong spin–orbit coupling destroys the SU(3) symmetry. However, alternative proposals have been presented,[15,31] or may become available in the future, to determine the $(\omega_1\ \omega_2\ \omega_3)$ to which one is restricted in collective excitations.

We now turn our attention to why the collective Hamiltonian should be defined in the enveloping algebra of Sp($2d$). To begin with, we notice that the mass quadrupole of the A nucleon system, given by (2), and already related to collective behavior by Bohr and Mottelson,[2,3] is in fact given in Cartesian components by $q_{ij} - \frac{1}{3}\sum_{k=1}^{3} q_{kk}\,\delta_{ij}$, where q_{ij} was defined in (8a). Thus it is an element of the Sp($2d$) Lie algebra. More generally, Vanagas[11] has stressed that the collective component of a given many-body Hamiltonian is related to that part of it that is invariant under the O(n) group and thus can be expressed in the enveloping algebra of Sp($2d$).

Collective states and Hamiltonians are not then the most general wave functions and operators that can be defined for the many-body system, but in effect they obey a constraint; for the states that they belong to a given irrep $(\omega_1\cdots\omega_d)$ of O(n) and for the Hamiltonians that they are invariant under O(n) and $\mathcal{O}(d)$. This restriction in our symplectic space resembles the constraint in Euclidean space that we discussed in Sec. 3 of this note. As in the Euclidean problem, the constraint in the symplectic manifold drastically reduces

the dimensions of the Hilbert space as we pass from one characterized by dn quantum numbers, if the physical space has d dimensions, to one in which we have only d^2 quantum numbers. This is a tremendous reduction if we consider that $n = A - 1$ is of the order of a hundred in nuclei in which collective effects are of interest, while $d = 3$ in physical space.

On the other hand, the constraint makes the determination of states and matrix elements more difficult, in the same way as the constraint in Euclidean space leads to a geometry on a sphere which is more difficult than that of a plane.

6. PROCEDURE FOR CALCULATIONS

What procedure must we then follow for calculations in the collective model defined above? First we must characterize the states of H_0 not only by an irrep $(\omega_1\,\omega_2\cdots\omega_d)$ of O(n), and thus by an irrep of Sp($2d$), but also by subgroups of the latter. We found it convenient to consider the chain

$$\mathrm{Sp}(2d)\supset \mathrm{Sp}(2)\times\mathcal{O}(d), \tag{13}$$

where the generators of $\mathcal{O}(d)$ are given by (8d), while those of Sp(2) are the scalars of $\mathcal{O}(d)$ that we can form from the generators (8a)–(8c), i.e.,

$$I_1=\frac{1}{4}\sum_{i=1}^{d}(T_{ii}-q_{ii}),\quad I_2=\frac{1}{4}\sum_{i=1}^{d}(S_{ii}),$$
$$I_3=\frac{1}{4}\sum_{i=1}^{d}(T_{ii}+q_{ii}), \tag{14}$$

which satisfy[26] the commutation rules
$[I_1,I_2]=-iI_3,\quad [I_3,I_1]=iI_2,\quad [I_2,I_3]=iI_1.$

The eigenstates of H_0 are then characterized by the irrep $(\omega_1\cdots\omega_d)$ of O(n), and thus also of Sp($2d$), and by the irrep λ of Sp(2) and $(L_1,L_2,\ldots,L_{[d/2]})$ of $\mathcal{O}(d)$ (where $[d/2]$ equals $d/2$ if d is even and $(d-1)/2$ if d is odd) as well as by multiplicity indices in the chain (13). The irreps of the subgroups of O(n), also required in the characterization of the eigenstates of H_0, will be irrelevant in the calculation of the matrix elements of operators in the enveloping algebra of Sp($2d$).

Clearly, then, what we require are the matrix elements of the q_{ij} of (8a) in the basis of eigenstates of H_0 characterized by the chain of subgroups (6), (13), as from them we can immediately obtain those of S_{ij}, T_{ij} using the commutation relations in (8b), (8c), and thus finally the matrix elements of any operator in the enveloping algebra of Sp($2d$).

The determination of the eigenstates mentioned in the previous paragraph requires the use of an appropriate coordinate system introduced by Zickendraht[5] and by Dzublik *et al.*[6] in which the dn coordinates x_{is}, $i=1,\ldots,d$, $s=1,\ldots,n$, are expressed as

$$x_{is}=\sum_{k=1}^{d}\rho_k D^{1}_{ki}(\vartheta)D^{1}_{n-d+k,s}(\phi). \tag{15}$$

In (15) the matrix $\mathbf{D}^1=\|D^1_{ki}(\vartheta)\|$ is the defining representation (which is the reason for the 1 appearing as an upper index) of the $\mathcal{O}(d)$ group, in terms of its $(d/2)(d-1)$ Euler angles. The matrix $\|D^1_{ts}(\phi)\|$ has the same meaning for the O(n) group and it depends on $(n/2)(n-1)$ angular coordinates denoted by ϕ. As in (15), we need only the last d rows of this matrix, the number of angular coordinates required will be those of the O(n) minus those of the O($n-d$) subgroup associated with rows 1 to $n-d$, i.e.,

$$(n/2)(n-1)-[(n-d)/2)](n-d-1)$$
$$=nd-(d/2)(d+1). \tag{16}$$

If we add to the number (16) of ϕ's that of ϑ's and the ρ's, we obtain nd, i.e., the number of x_{is}, as we should.

Furthermore, we note that from the orthogonal properties of the matrices $\|D^1_{ts}(\phi)\|$ we have

$$\mathbf{q}=\|q_{ij}\|=\tilde{\mathbf{D}}^1(\vartheta)\|\rho_i^2\delta_{ij}\|\mathbf{D}^1(\vartheta), \tag{17}$$

where $\tilde{}$ indicates the transposed matrix. Thus the ρ_k's appearing in (15) are related to the diagonal quadrupole matrix $\|\rho_i^2\delta_{ij}\|$ while $\mathbf{D}^1(\vartheta)$ is the orthogonal matrix, function of the $(d/2)(d-1)$ Euler angles, that takes us from the frame of reference fixed in space to the one fixed in the body.

7. ONE-DIMENSIONAL CASE

We wish now to illustrate our analysis for $d=1$. The problem is trivial as we can suppress the index $i=1$ and have $x_s=\rho D^1_{ns}(\phi)$, i.e., and expansion in hyperspherical coordinates. The Hamiltonian H_0 of (5) for $d=1$, can be written as

$$H_0=\frac{1}{2}\left(-\frac{1}{\rho^{n-1}}\frac{\partial}{\partial\rho}\rho^{n-1}\frac{\partial}{\partial\rho}+\frac{\mathscr{L}^2}{\rho^2}+\rho^2\right), \tag{18}$$

where $\mathscr{L}^2$ is the Casimir operator of O(n), defined after Eq. (7), whose eigenvalue[27] is given by $\omega(\omega+n-2)$.

From (18) we see that the collective state can be written as

$$\left|\begin{matrix}\omega\\ N;\delta\{f\}(r)\end{matrix}\right\rangle=\rho^{-(n-1)/2}f^{\omega+(n-2)/2}_{(N-\omega)/2}(\rho)D^{\omega}_{0,\delta\{f\}(r)}(\phi), \tag{19}$$

where N is the total number of quanta and $D^{\omega}(\phi)$ is the irrep of O(n), characterized by the single value ω, whose row[25] becomes 0, while the column is characterized by the irreps of the symmetric subgroups $S_{n+1}\supset\cdots\supset S_2$ of O(n), i.e., by the partition $\{f\}$ and Yamanouchi symbol (r). The δ corresponds to the multiplicity indices necessary to distinguish between identical irreps $\{f\}$ of S_{n+1} contained in the given irrep ω of O(n). The radial part of the function is given by[32]

$$f^{\omega+(n-2)/2}_{(N-\omega)/2}(\rho)=\{2[(N-\omega)/2]!/\Gamma[(N+\omega+n)/2]\}^{1/2}$$
$$\times e^{-\rho^2/2}\rho^{\omega+(n-1)/2}L^{\omega+(n-2)/2}_{(N-\omega)/2}(\rho^2), \tag{20}$$

where $L^{\omega+(n-2)/2}_{(N-\omega)/2}(\rho^2)$ is a Laguerre polynomial.

The matrix elements that we require are then only those of

$$q_{11}=\rho^2=2(I_3-I_1), \tag{21}$$

with respect to the states (19), which implies that the $D^{(\omega)}(\phi)$ will be irrelevant for their evaluation. The part connected with the radial function $f(\rho)$, can be obtained from the fact[32] that I_3 applied to the ket (19) gives $(N/2)+(n/4)$, while[32]

$$\left\langle\begin{matrix}\omega\\ N+2;\delta\{f\}(r)\end{matrix}\right|I_{\pm}\left|\begin{matrix}\omega\\ N;\delta\{f\}(r)\end{matrix}\right\rangle$$
$$=\tfrac{1}{2}[(N+\omega+n-1\pm1)(N-\omega\pm1)]^{1/2}, \tag{22}$$

where $I_{\pm}=I_1\pm iI_2$.

8. AN EXAMPLE IN THREE DIMENSIONS

For $d = 3$ the situation is more complex.[25] We will formulate it in general but will be able to implement our program explicitly only in the limiting case when $n \to \infty$. It is first convenient to replace the three coordinates ρ_k, $k = 1,2,3$, by ρ,b,c through the definitions

$$\rho_k^2 = (\rho^2/3)\{1 + 2b\cos[c - (2\pi k/3)]\}, \quad k = 1,2,3. \tag{23}$$

As ρ is then an hyperradius in a $3n$-dimensional space, the Hamiltonian H_0 can be written as

$$H_0 = \frac{1}{2}\left(-\frac{1}{\rho^{3n-1}}\frac{\partial}{\partial\rho}\rho^{3n-1}\frac{\partial}{\partial\rho} + \frac{\Lambda^2}{\rho^2} + \rho^2\right), \tag{24}$$

where Λ^2 is the Casimir operator of the group O($3n$) which is function[25] of b, c and their derivatives as well as of $\mathscr{L}'_{uv}$, $u,v = 1,\ldots,n$ defined by[25]

$$\mathscr{L}'_{uv} = \sum_{s,t=1}^{n} D^{1}_{us}(\phi)D^{1}_{vt}(\phi)\mathscr{L}_{st}, \tag{25}$$

where $\mathscr{L}_{st}$ are the generators (7) of O(n).

We note that the symplectic group Sp($6n$), corresponding to $d = 3$, admits also the subgroup

$$\mathrm{Sp}(6n) \supset \mathrm{Sp}(2) \times \mathrm{O}(3n), \tag{26}$$

where the generators of Sp(2) are given by (14) and those of O($3n$) by $x_{is}p_{jt} - x_{jt}p_{is}$, $i,j = 1,2,3$ $s,t = 1,\ldots,n$.

As in the analysis following (6)–(8) in the irreps $[\frac{1}{2}^{3n}]$, $[\frac{1}{2}^{3n-1}\frac{3}{2}]$ of Sp($6n$) the irreps of Sp(2) and O($3n$) are complementary.[9,28] In fact, it is easily seen from the generators (14) of Sp(2), and those of O($3n$) given in the previous paragraph, that the Casimir operators of these two groups are related by

$$I^2 \equiv I_3^2 - I_1^2 - I_2^2 = \tfrac{1}{4}[\Lambda^2 + \tfrac{1}{4}(3n)^2 - (3n)]. \tag{27}$$

As the eigenvalues of I^2 and Λ^2 can be denoted, respectively, by $\lambda(\lambda - 1)$ and $l(l + 3n - 2)$ we obtain the relation

$$\lambda = l/2 + 3n/4. \tag{28}$$

The eigenstates can then be written as

$$\left|\begin{matrix}(\omega_1\,\omega_2\,\omega_3)\\ T\lambda NLM;\delta\{f\}(r)\end{matrix}\right\rangle$$
$$= \rho^{-(3n-1)/2} f^{2\lambda-1}_{(N/2)-\lambda+(3n/4)}(\rho)\sum_{K}\sum_{\tau_1\tau_2\xi_1} \times F^{(\omega_1\,\omega_2\,\omega_3)T\lambda NL}_{K;\tau_1\tau_2,\xi_1}(b,c)D^{L}_{KM}(\vartheta_i)D^{(\omega_1\,\omega_2\,\omega_3)}_{\tau_1\tau_2,\xi_1;\delta\{f\}(r)}(\phi), \tag{29}$$

where the quantum numbers are associated with the irreps of the following groups: $(\omega_1\,\omega_2\,\omega_3)$ with both O(n) and Sp(6), λ with Sp(2), L with $\mathscr{O}(3)$ and M with its subgroup $\mathscr{O}(2)$, while $\delta,\{f\},(r)$ have the same meaning as in the one-dimensional case. The total number of quanta, N, is related to the irrep of the O(2) subgroup of Sp(2) whose generator is $I_3 = (1/2)H_0$; it follows then that the eigenvalue of H_0 is given by $N + (3n/2)$. Furthermore, T represents multiplicity indices between Sp(6) and Sp(2)$\times\mathscr{O}(3)$ (in fact five of them are required).

From (24) we see that the function $f^{2\lambda-1}_{(N/2)-\lambda+(3n/4)}(\rho)$ is given by (20), while $D^{L}_{KM}(\vartheta_i)$ is the Wigner function for $\mathscr{O}(3)$ in terms of the Euler angles $\vartheta_1,\vartheta_2,\vartheta_3$ and $D^{(\omega_1\,\omega_2\,\omega_3)}(\phi)$ is the corresponding function of the group O(n), in which we note[25] that the row is characterized by just three quantum numbers τ_1,τ_2,ξ_1. The remaining function $F(b,c)$ is quite difficult to determine except for the physically interesting case when the number of nucleons is very large.[25]

As indicated in our previous discussion, the next step is to calculate the matrix element of q_{ij} with respect to the states (29). It is convenient first to change q_{ij} for Cartesian components $i,j = 1,2,3$ to the spherical ones $\tau = 1,0,-1$ and then to the irreducible form q_m of (2) combined with $\bar{q} = (1/\sqrt{3})\Sigma_\tau(-1)^\tau q_{\tau,-\tau}$. Using then (17) and (23), we obtain

$$q_m = \sqrt{2/3}\,\rho^2\{2^{-1/2}\,b\sin c \times [D^{2}_{2m}(\vartheta_i) + D^{2}_{-2m}(\vartheta_i)] + b\cos cD^{2}_{0m}(\vartheta_i)\}, \tag{30a}$$

$$\bar{q} = (1/\sqrt{3})\rho^2. \tag{30b}$$

The action of these operators on the states (29) does not affect the irrep $D^{(\omega_1\,\omega_2\,\omega_3)}(\phi)$ of O(n), but only the part depending on ρ,b,c. For actual calculations we shall only consider the case when $A \gg 1$, which corresponds in this problem to $a \gg 1$ in the geometrical example given in Sec. 3. As discussed in Ref. 25, we can then substitute

$$1 + 2(\beta^2/\sigma^2) = (1 - b^2)(1 - 3b^2 + 2b^3\cos 3c)^{-1}, \quad c = \gamma, \tag{31}$$

where $\sigma^2 = (\omega_1 + \omega_2 + \omega_3) + \frac{3}{2}(A - 5) \gg 1$, and if we renormalize the operator Λ^2 through the change of volume element in Eq. (3.8) of Ref. 25, we get for the term of highest order in σ that[25]

$$\bar{\Lambda}^2 \equiv (\rho_1\rho_2\rho_3)^{(n-3)/2}\rho^4\Lambda^2(\rho_1\rho_2\rho_3)^{-(n-3)/2}\rho^{-4} = 4\sigma^2(H'_{\mathrm{BM}} - \tfrac{5}{2}), \tag{32}$$

in which

$$H'_{\mathrm{BM}} = \frac{1}{2}\left(-\frac{1}{\beta^4}\frac{\partial}{\partial\beta}\beta^4\frac{\partial}{\partial\beta} - \frac{1}{\beta^2\sin 3\gamma}\frac{\partial}{\partial\gamma}\sin 3\gamma\frac{\partial}{\partial\gamma}\right) + \sum_{k=1}^{3}\{4\beta^2\sin^2[\gamma - (2\pi k/3)]\}^{-1} \times (L'_k + \mathscr{L}'_k)^2 + (\beta^2/2), \tag{33}$$

where L'_k are the components of the ordinary angular momentum in the frame of reference fixed in the body, while the $\mathscr{L}'_k$, associated with what is known as vortex spin,[14] are given by

$$\mathscr{L}'_k = \sum_{k'k''}\epsilon_{kk'k''}\,\mathscr{L}'_{n-3+k',n-3+k''}, \tag{34}$$

where $\mathscr{L}'_{uv}$ is defined in (25) while $\epsilon_{kk'k''}$ is the antisymmetric tensor. The notation H'_{BM} for the operator appearing in (33) is to indicate its relation to the Bohr–Mottelson (BM) vibrational Hamiltonian though now including also a vortex spin. From (27), (32) we see that, when $A \gg 1$, H'_{BM} becomes a function of the Casimir operator I^2 of Sp(2).

We need now to consider what happens to the $H_0, q_m, \bar{q}$ of (24) and (30) when A, and particularly σ^2, are much larger than 1. For this purpose we can replace Λ^2 in (24) by its value (32) and it is also convenient to write ρ in terms of a new variable $\bar{\alpha}$ through the relation

$$\rho = \sigma \exp(\bar{\alpha}/\sqrt{2}\sigma), \tag{35}$$

where as $0 \leqslant \rho \leqslant \infty$ we have that $-\infty \leqslant \bar{\alpha} \leqslant \infty$.

From the analysis of Ref. 25, we note that the H_0 of (24), again renormalized as in (32), takes to order $(1/\sigma)$ the following form:

$$\begin{aligned}\overline{H}_0 &\equiv (\rho_1\rho_2\rho_3)^{(n-3)/2}\rho^4 H_0 (\rho_1\rho_2\rho_3)^{-(n-3)/2}\rho^{-4} \\ &= \sigma^2 + 2\left[\tfrac{1}{2}(-\partial^2/\partial\bar{\alpha}^2 + \bar{\alpha}^2) + H'_{\mathrm{BM}}\right] \\ &= \sigma^2 + 2H'_{\mathrm{IBA}}.\end{aligned} \tag{36}$$

The part in the square bracket resembles the Hamiltonian used in Ref. 20 for discussion of the states in the IBA. We have indicated this at the right-hand side of the equation through the introduction of H'_{IBA}. We note though that from (36), (33) the H'_{IBA} is also dependent on the vortex spin.

Turning now our attention of $q_m, \bar{q}$, we see that, in view of the fact that $\sigma \gg 1$, we get from (31) that $b \simeq (\beta/\sigma)$. It is clear then that, up to terms of lowest order in σ, we obtain

$$q_m \simeq \sqrt{2/3}\,\alpha_m, \tag{37a}$$

where α_m is related to $\beta, \gamma, \vartheta_i$ in the standard Bohr–Mottelson[3] fashion. Furthermore, we also get

$$\bar{q} \simeq (1/\sqrt{3})\sigma^2 + \sqrt{2/3}\,\bar{\alpha}. \tag{37b}$$

In the limiting case $A \gg 1$—which implies $\sigma \gg 1$—our problem is the determination of $q_m, \bar{q}$ of (37) with respect to the eigenstates of the H_0 in (36). The latter have been derived explicitly in Eq. (5.30) of Ref. 25. To obtain the matrix elements of $q_m, \bar{q}$ with respect to these states, we first consider the situation when the irrep $(\omega_1\,\omega_2\,\omega_3)$ of O(n) becomes $(\omega\omega\omega)$ corresponding to the case of closed shells. The eigenstate of (37) can then be written as[25]

$$\begin{aligned}&\left|\begin{matrix}\omega\omega\omega \\ \mathcal{N}\nu\Lambda tLM;\delta\{f\}(r)\end{matrix}\right\rangle \\ &\quad = |\nu\Lambda tLM\rangle \psi_{\mathcal{N}-\nu}(\bar{\alpha}) D^{(\omega\omega\omega)}_{\omega\omega,\omega;\delta\{f\}(r)}(\phi).\end{aligned} \tag{38}$$

In (38) $\mathcal{N} = \frac{1}{2}(N - \omega_1 - \omega_2 - \omega_3) = (N - 3\omega)/2$, where N, as before, is the total number of quanta in H_0 and from the factor 2 in (36) it jumps by steps of 2 (starting from its lowest value $\omega_1 + \omega_2 + \omega_3 = 3\omega$). The $\psi_{\mathcal{N}-\nu}(\bar{\alpha})$ is a one-dimensional oscillator state of $\mathcal{N} - \nu$ quanta associated with the operator $\frac{1}{2}(-\partial^2/\partial\bar{\alpha}^2 + \bar{\alpha}^2)$. The $D^{(\omega\omega\omega)}_{\omega\omega,\omega;\delta\{f\}(r)}(\phi)$ is the irrep $(\omega\omega\omega)$ of O(n) given in (29) when the row[1] is $\omega\omega,\omega$ so that $\mathscr{L}'_k$ acting on it vanishes, in the same way as $L_k|lm\rangle$ vanishes if $l = 0$. The $|\nu\Lambda tLM\rangle$ are the eigenkets of the Hamiltonian H'_{BM} of (33) in which $\mathscr{L}'_k = 0$, i.e., the eigenstates of the standard Bohr–Mottelson Hamiltonian[2] that were completely determined in Refs. 18 and 19.

From (36) it is clear that we need only to calculate the matrix elements of $\bar{\alpha}, \alpha_m$ with respect to the states (38). For $\bar{\alpha}$ this is trivial as $\psi_{\mathcal{N}-\nu}(\bar{\alpha})$ is the wave function of a one-dimensional oscillator. For α_m we have the matrix elements

$$\langle \nu'\Lambda' t' L' M' | \alpha_m | \nu\Lambda tLM \rangle, \tag{39}$$

which were obtained explicitly in Ref. 19.

We briefly discuss (39) to provide the setting for the situation when we deal with the general irrep $(\omega_1\,\omega_2\,\omega_3)$ of O(n). The eigenket of the Bohr–Mottelson Hamiltonian can be expressed as

$$|\nu\Lambda tLM\rangle = F^{\Lambda}_{(\nu-\Lambda)/2}(\beta) \sum_K \Phi^{\Lambda tL}_K(\gamma) D^L_{KM}(\vartheta_i), \tag{40}$$

in which $F(\beta), \Phi(\gamma)$ are the the part of the eigenket depending on β,γ where the former is well known and the latter was determined explicitly in Ref. 19. We note here that ν corresponds to the total number of quanta of the BM Hamiltonian, i.e., in a five-dimensional oscillator. The Λ corresponds[20] to the irrep of O(5) and t is a multiplicity index[20] in the chain O(5) $\supset$ O(3).

As indicated in Ref. 19, we can write

$$\alpha_m = \beta \sum_{k=-2} \Phi^{112}_k(\gamma) D^2_{km}(\vartheta_i), \tag{41}$$

where

$$\begin{aligned}&\Phi^{112}_2 = \Phi^{112}_{-2} = (1/\sqrt{2})\sin\gamma, \\ &\Phi^{112}_1 = \Phi^{112}_{-1} = 0, \quad \Phi^{112}_0 = \cos\gamma.\end{aligned} \tag{42}$$

The matrix element (39) can then be written as

$$\begin{aligned}&\langle \nu'\Lambda' t' L' M' | \alpha_m | \nu\Lambda tLM \rangle \\ &\quad = \{\nu'\Lambda' | 1 | \nu\Lambda\}(112;\Lambda tL;\Lambda' t' L') 8\pi^2 (2L'+1)^{-1/2} \\ &\qquad \times \langle LM, 2m | L' M' \rangle,\end{aligned} \tag{43}$$

where

$$\{\nu'\Lambda' | p | \nu\Lambda\} = \int_0^\infty F^{\Lambda'}_{(\nu-\Lambda')/2}(\beta)\beta^p F^{\Lambda}_{(\nu-\Lambda)/2}(\beta)\beta^4\,d\beta, \tag{44}$$

is the radial integral and

$$\begin{aligned}&(\Lambda'' t'' L''; \Lambda tL; \Lambda' t' L') \\ &\quad = \int_0^\pi \sum_{K K'} [(2L'+1)^{-1/2} \langle LK, L''K'' | L' - K' \rangle \\ &\qquad \times (-1)^{L'} \Phi^{\Lambda'' t'' L''}_{K''}(\gamma) \Phi^{\Lambda tL}_K(\gamma) \Phi^{\Lambda' t' L'}_{K'}(\gamma)] \sin 3\gamma\,d\gamma,\end{aligned} \tag{45}$$

is the integral depending on γ which can be interpreted as a reduced $3j$ symbol in the O(5) $\supset$ O(3) chain of groups.[19] Note that we use the selection rule[19] $\Phi^{\Lambda tL}_K(\gamma) = 0$ if K is odd and $\Phi^{\Lambda tL}_K(\gamma)^* = (-1)^L \Phi^{\Lambda tL}_{-K}(\gamma)$ when K is even. Explicit expressions are available for the brackets (45) and they have also been tabulated numerically.[33]

Let us now turn our attention to the case where we still maintain $A \gg 1$ but we have a general irrep $(\omega_1\,\omega_2\,\omega_3)$ of O(n). Then, as shown in Eq. (5.30) of Ref. 25, the eigenket of H_0 in (36) takes the form

$$\begin{aligned}&\left|\begin{matrix}\omega_1\,\omega_2\,\omega_3 \\ \mathcal{N}\nu,\Lambda tJ,\, LM, \Omega\mathbb{L}; \delta\{f\}(r)\end{matrix}\right\rangle \\ &\quad = \psi_{\mathcal{N}-\nu}(\bar{\alpha}) F^{\Lambda}_{(\nu-\Lambda)/2}(\beta) \sum_{\mathscr{K}} \Bigg\{ \Phi^{\Lambda tJ}_{\mathscr{K}}(\gamma) \\ &\qquad \times \sum_{K\mathbb{K}} [\langle LK, \mathbb{L}\mathbb{K} | J\mathscr{K}\rangle D^L_{KM}(\vartheta_i) \Delta^{(\omega_1\omega_2\omega_3)}_{\Omega\mathbb{L}K;\delta\{f\}(r)}(\phi)] \Bigg\},\end{aligned} \tag{46}$$

where

$$\begin{aligned}&\Delta^{(\omega_1\omega_2\omega_3)}_{\Omega\mathbb{L}K;\delta\{f\}(r)}(\phi) \\ &\quad = \sum_{\tau_1\tau_2\zeta_1} \left(\begin{matrix}\omega_1\,\omega_2\,\omega_3 \\ \Omega\mathbb{L}K\end{matrix} \middle| \begin{matrix}\omega_1\,\omega_2\,\omega_3 \\ \tau_1\,\tau_2 \\ \zeta_1\end{matrix} \right) D^{(\omega_1\omega_2\omega_3)}_{\tau_1\tau_2\zeta_1;\delta\{f\}(r)}(\phi).\end{aligned} \tag{47}$$

All functions appearing in (46) have already been defined except of $\Delta(\phi)$. The latter is given by (47), in which the $D^{(\omega_1\omega_2\omega_3)}(\phi)$ is the irrep of O(n) that already appeared in the general case (29), multiplied by a transformation bracket relating the U(3) basis characterized by the chain U(3)⊃U(2)⊃U(1) with the one characterized by U(3)⊃O(3)⊃O(2). In the latter the O(3) is *not* connected with orbital angular momentum but is rather associated with the vortex spin mentioned above, and its magnitude and projection are indicated by the quantum numbers L, $\mathbb{K}$. As usual, we add the multiplicity index required between U(3) and O(3), which we designate by Ω. The orbital angular momentum and the vortex spin are coupled to a total J as indicated in (46).

To determine now the matrix elements of α_m with respect to the states (46), we need only to recall that the reduced matrix elements of a tensor $T_L(1)$ depending on system 1, with respect to states in which the angular momenta of systems 1 and 2 are coupled to a definite j, is given by[34]

$$\begin{aligned}(j'j_1'j_2'\|T_L(1)\|jj_1j_2)&=\delta_{j_2'j_2}(-1)^{j_2+L-j_1-j'}[(2j_1'+1)(2j+1)]^{1/2}\\&\times W(jj_1'j'j_1';Lj_2)(j_1'\|T_L(1)\|j_1),\end{aligned}\tag{48}$$

where W is a Racah coefficient.

The matrix $\bar{\alpha}$ with respect to the states (46) involves only the one-dimensional oscillator $\psi_{\nu'-\nu}(\bar{\alpha})$ and therefore is immediate. For α_m we have now, using (48) and the results discussed above, that the matrix element becomes

$$\begin{aligned}&\left\langle {}^{\omega_1\omega_2\omega_3}_{\mathcal{N}'\nu',\Lambda' t'J',L'M',\Omega'\mathbb{L}';\delta\{f\}(r)}\middle|\alpha_m\middle|{}^{\omega_1\omega_2\omega_3}_{\mathcal{N}\nu,\Lambda tJ,LM,\Omega\mathbb{L};\delta\{f\}(r)}\right\rangle\\&=\delta_{\Omega'\Omega}\delta_{\mathbb{L}'\mathbb{L}}\delta_{\Lambda'-\nu',\Lambda-\nu}\{\nu'\Lambda'|1|\nu\Lambda\}\\&\times(112;\Lambda tJ;\Lambda' t'J')(8\pi^2)(2L'+1)^{-1/2}\\&\times\langle LM,2m|L'M'\rangle(-1)^{\mathbb{L}-L-J'}\\&\times[(2J+1)(2J'+1)]^{1/2}W(LL'JJ';2\mathbb{L}).\end{aligned}\tag{49}$$

We have thus obtained explicitly the matrix elements of $q_m,\bar{q}$ with respect to the eigenstates of H_0 when A—and thus σ—become very large. As the other generators of Sp(6) are obtained from commutators and double commutators of $q_m,\bar{q}$ with H_0, we have from (8) that all matrix elements of the generators of Sp(6), in the limit when $A\to\infty$, are obtained when we consider $[H_0,\alpha_m]$, $[H_0,[H_0,\alpha_m]]$, and similarly for $\bar{\alpha}$. From this result we have, in principle, that all matrix elements of operators in the enveloping algebra of Sp(6) are available.

In fact we have even a direct way of dealing with elements in the enveloping algebra of Sp(6) that are invariant under O(3). We can consider, for example

$$[\alpha\times\alpha]_0^0=\beta^2/\sqrt{5},\tag{50a}$$

$$[[\alpha\times\alpha]^2\times\alpha]_0^0=-\sqrt{2/35}\beta^3\cos 3\gamma.\tag{50b}$$

From (46) we immediately see that matrix elements of β^{2r} reduce to the integrals in β of the form $\{\nu'\Lambda|2r|\nu\Lambda\}$ with all the other quantum number in bra and ket being equal. From matrix elements of powers of $\beta^3\cos 3\gamma$, the β part is trivial, and we can decompose $(\cos 3\gamma)^p$ in Legendre polynomials $P_s(\cos 3\gamma)$. As $P_s(\cos 3\gamma)=\Phi_0^{3s,1,0}(\gamma)$, we can immediately obtain that the matrix element of $\beta^{2r+3s}P_s(\cos 3\gamma)$ is given by

$$\begin{aligned}&\left\langle {}^{\omega_1\omega_2\omega_3}_{\mathcal{N}'\nu',\Lambda' t'J',L'M',\Omega'\mathbb{L}';\delta\{f\}(r)}\middle|\beta^{2r+3s}P_s(\cos 3\gamma)\middle|{}^{\omega_1\omega_2\omega_3}_{\mathcal{N}\nu,\Lambda tJ,LM,\Omega\mathbb{L};\delta\{f\}(r)}\right\rangle\\&=\delta_{\Lambda'-\nu',\Lambda-\nu}\delta_{\Omega'\Omega}\delta_{\mathbb{L}'\mathbb{L}}\delta_{L'L}\delta_{M'M}\delta_{J'J}\\&\times\{\nu'\Lambda'|2r+3s|\nu\Lambda\}\\&\times(3s,1,0;\Lambda tJ,\Lambda' t'J')\\&\times(8\pi^2)(2L+1)^{-1}(2J+1)^{1/2},\end{aligned}\tag{51}$$

where the brackets have the same meaning as before, but now with new parameters.

The matrix elements (51) would allow us to make calculations in what has been called[35] the "Frankfurt collective model" but now in the presence of vortex spin.

9. GENERAL APPROACH TO THE PROBLEM

Summarizing, we can say that the framework we have introduced for description of collective effects requires, when we are dealing with a d-dimensional space,the determination of the matrix elements for the generators q_{ij} of the Sp($2d$) group given in (8a), with respect to the states of the oscillator Hamiltonian H_0 of (5) characterized by the chain of groups

$$\begin{array}{llcl}&\lambda&&N\\(\omega_1\,\omega_2\cdots\omega_d)&\mathrm{Sp}(2)&\supset&\mathrm{O}(2)\\\mathrm{Sp}(2d)\supset&\times&&\\&\mathscr{U}(d)&\supset\cdots\supset&\mathscr{U}(2)\\&(L_1\cdots L_{[d/2]})&\cdots&M\end{array},\tag{52}$$

where above and below the groups we give the quantum numbers that characterize their irreps. This is as much a geometrical problem as the one in which we require the matrix elements of the generators L_i, $i=1,2,3$, of the O(3) group, with respect to states that are characterized by the chain of groups O(3)⊃O(2).

In the present paper our program was implemented fully when $d=1$ in Sec. 7. In Sec. 8 it was implemented when $d=3$ but with the restriction that $A\gg 1$, and more correctly only in the limit when $A\to\infty$, where the problem simplifies in a sense similar to the elementary geometrical example in Sec. 3 when the radius of the sphere $a\to\infty$. For $d=2$ and finite, though arbitrary, A the complete solution of the problem will be presented in a companion paper of Chacón, Hess, and Moshinksy.[36] For $d=3$ and arbitrary A, we have just started the problem using the same techniques as in Ref. 36, though possibly it may be attacked with the alternative procedures developed by other authors.[15,30]

As the last point in this section, we would like to determine the relation between the irreps of O(n) and Sp($2d$), which will be useful for the analysis carried out in the next papers of this series.

From (5) and (8) the weight generators of the Sp($2d$) group are the operators H_i, $i=1,2,\ldots,d$, defined by

$$H_i = \frac{1}{2}\sum_{s=1}^{n}(p_{is}^2 + x_{is}^2) = \mathscr{C}_{ii} + \frac{n}{2}, \tag{53}$$

where the eigenvalues of the operator

$$\mathscr{C}_{ii} = \sum_{s=1}^{n} \eta_{is}\xi_{is}, \tag{54a}$$

$$\eta_{is} = (1/\sqrt{2})(x_{is} - ip_{is}), \quad \xi_{is} = (1/\sqrt{2})(x_{is} + ip_{is}), \tag{54b}$$

give the number of quanta for the corresponding oscillator Hamiltonian H_i, $i = 1,2,...,d$.

As indicated in Sec. 4, the Hamiltonian $H_0 = \Sigma_{i=1}^{d} H_i$, admits the chain of symmetry groups

$$\mathrm{U}(dn) \supset \mathscr{U}(d) \times \mathrm{U}(n), \tag{55}$$

which can be complemented with the subgroups

$$\mathscr{U}(d) \supset \mathscr{U}(d-1) \supset \cdots \supset \mathscr{U}(1),$$
$$\mathrm{U}(n) \supset \mathrm{O}(n) \supset S_{n+1} \supset S_n \supset \cdots \supset S_2, \tag{56}$$

where S_m are symmetric groups in m particles. The eigenstates of H_0 can then be denoted by

$$\left.\begin{array}{lcl} h_{1d}h_{2d} \cdots h_{dd} & ; & [h_{1d}\cdots h_{dd}] \\ h_{1,d-1}\cdots h_{d-1,d-1} & & (\omega_1\cdots\omega_d) \\ \cdots & & \\ h_{11} & & \delta\{f\}(r) \end{array}\right\rangle, \tag{57}$$

where $[h_{1i},...,h_{ii}]$, $i = 1,2,...,d$, are the irreps of $\mathscr{U}(i)$, $(\omega_1\cdots\omega_d)$ the irrep of $\mathrm{O}(n)$ and $\{f\}(r)$ correspond, as indicated after Eq. (19), to the irreps of $S_{n+1} \supset \cdots \supset S_2$. The $\mathscr{C}_{ii}$ applied to the kets (57) give the eigenvalues

$$\sum_{j=1}^{i} h_{ji} - \sum_{j=1}^{i-1} h_{j\,i-1}. \tag{58}$$

The irrep of $\mathrm{Sp}(2d)$ is given by the eigenvalues of H_i when applied to the *lowest* weight state (57) consistent with the given irrep $(\omega_1\cdots\omega_d)$ of $\mathrm{O}(n)$. As indicated in Sec. 4 this implies that $h_{id} = \omega_i$, $i = 1,2,...,d$, and, furthermore, that in each row on the left-hand side of the ket $h_{ji} = h_{j+1,i+1}$, $i,j = 1,2,...,d$. Thus we have to apply H_i to the ket

$$\left.\begin{array}{lcl} \omega_1\omega_2\cdots\omega_d & & [\omega_1\omega_2\cdots\omega_d] \\ \omega_2\cdots\omega_d & ; & (\omega_1\omega_2\cdots\omega_d) \\ \cdots & & \\ \omega_d & & \delta\{f\}(r) \end{array}\right\rangle, \tag{59}$$

which, from (53) and (58), gives the eigenvalue $\omega_{d-i+1} + n/2$ for the operator H_i.

Thus to an irrep $(\omega_1\omega_2\cdots\omega_d)$ of $\mathrm{O}(n)$ corresponds an irrep of $\mathrm{Sp}(2d)$ given by

$$[(n/2) + \omega_d, (n/2) + \omega_{d-1},...,(n/2) + \omega_1]. \tag{60}$$

It is interesting to note that we do not need to consider the most general irrep of $\mathrm{O}(n)$ to get (60). In fact, if we restrict ourselves to the irrep $(\omega_1',...,\omega_{d-1}',0)$ of $\mathrm{O}(n')$ we get the most general irrep (60) of $\mathrm{Sp}(2d)$ if we take $n' = n + 2\omega_d$, $\omega_i' = \omega_i - \omega_d$.

10. CONCLUSION

We have given in Sec. 5 a definition of collectivity and then shown that it leads to a specific mathematical problem which we intend to solve in the present series of papers. In this concluding section we would like to examine critically how close our definition is to the actual physical situation of collective effects for systems for nucleons in nuclei.

We mentioned in Sec. 5 that we could use the shell model to obtain the irreps of $\mathrm{O}(n)$ to which we restrict ourselves. But in this case the oscillator Hamiltonian H_0 would be the shell model Hamiltonian and $\hbar\omega$ would be of the order of MeV. This implies that we have degenerate ground states characterized by an irrep $(\lambda\mu)$ of SU(3) related with irrep $(\omega_1\omega_2\omega_3)$ of $\mathrm{O}(n)$ as in the discussion following Eq. (12). Furthermore, the collective excitations build up by jumps $2\hbar\omega$ from the ground state as indicated in the discussion following Eq. (38). This seems to imply that the definition of collectivity introduced in this paper is more closely related to collective effects in giant quadrupole and monopole resonances than to those in low lying collective phenomena.

On the other hand, it is an interesting fact that, for $A \gg 1$, the Casimir operators of the Sp(2) and O(2) groups appearing in the chain (52), are related to the Bohr–Mottelson and IBA basic Hamiltonians, albeit with an extra vortex spin. In view of the successes of the BM and IBA models, it is interesting to speculate whether H_0 is not necessarily connected with the shell model, but just provides a convenient starting point for the discussion of collective effects from a many-body point of view. In this case, though, we have to try to eliminate in some way the vortex spin effects, which do not appear in the models mentioned.

In any case the author feels strongly that collectivity must be related to constraints in the symplectic geometry associated with the many-nucleon system, though future analysis may show that the constraints are not necessarily of the type proposed in the present article.

ACKNOWLEDGMENTS

The present paper puts into focus ideas developed during the last years by the author and his collaborators. He would like to thank O. Castaños, E. Chacón, A. Frank, P. Hess, and T. H. Seligman in Mexico for many discussions, and abroad to J. Deenen, P. Kramer, C. Quesne, R. T. Sharp, and V. Vanagas. He is also indebted to L. C. Biedenharn and D. Rowe for helpful comments.

[1] N. Bohr, Nature **137**, 344 (1936).
[2] A. Bohr and B. Mottelson, K. Dan. Vidensk. Selsk. Mat. Fys. Med. **27**, No. 16, (1953).
[3] J. M. Eisenberg and W. Greiner, *Nuclear Models* (North-Holland, Amsterdam, 1970), p. 56.
[4] S. Goshen and H. J. Lipkin, Ann. Phys. (N.Y.) **6**, 301 (1959), S. Goshen and H. J. Lipkin "On the Application of the Group Sp(4) or R(5) to Nuclear Structure" in *Spectroscopic and Group Theoretical Methods in Physics. Racah Memorial Volume*, edited by F. Bloch, S. G. Cohen, A. de Shalit, S. Sambursky, and I. Talmi (North-Holland, Amsterdam, 1968), pp. 245–73.
[5] W. Zickendraht, J. Math. Phys. **12**, 1663 (1971).
[6] A. Ya Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[7] G. F. Filippov, A. I. Steshenko, and V. I. Ovcharenko, Preprint ITF-72-66, Institute of Theoretical Physics, Ukranian Academy of Sciences, 1972; also in Isv. Akad. Nauk (Ser. Fiz.) **37**, 1613 (1973).
[8] V. V. Vanagas and R. K. Kalinauskas, Yad. Fiz. **18**, 768 (1973); V. Vanagas, E. Nadjakov, and P. Raychev, Bulg. J. Phys. **6**, 559 (1975).
[9] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).

[10]G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979); V. C. Vasilevsky Ya. F. Smirnov, and G. F. Filippov, Yad. Fiz. **32**, 987 (1980); G. F. Filippov, V. I. Ovcharenko, and Ya. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (Naukova Dumka, Kiev, 1981) (in Russian). Further references are given in this book.
[11]V. V. Vanagas, "The Microscopic Nuclear Theory," in *Lecture Notes in Physics* (University of Toronto Press, Toronto, 1977); V. Vanagas, "The Microscopic Theory of the Collective Motions in Nuclei," in *Group Theory and Its Applications in Physics—1980*, edited by T. H. Seligman (AIP, New York, 1980). Further references are given in these books.
[12]G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **96**, 1 (1976); **123**, 36 (1978); **126**, 198 (1980).
[13]O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976).
[14]B. Buck, L. C. Biedenharm, and R. Y. Cusson, Nucl. Phys. A **317**, 205 (1979).
[15]G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **38**, 10 (1977); Ann. Phys. (N.Y.) **126**, 343 (1980); G. Rosensteel, J. Math. Phys. **21**, 924 (1980).
[16]J. P. Elliott, Proc. Roy. Soc. (London) A **245**, 128 (1958).
[17]J. Carvalho, P. Park, D. J. Rowe, and G. Rosensteel, Phys. Lett. B **119**, 249 (1982); M. Vassanji and D. J. Rowe, Phys. Lett. B **115**, 77 (1982); and two other papers to be published.
[18]E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976)
[19]E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).
[20]O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[21]A. Arima and F. Iachello, Ann. Phys. (N.Y.) **94**, 253 (1976); **111**, 201 (1978); A. Arima, F. Iachello, and D. Scholten, Ann. Phys. (N.Y.) **115**, 325 (1978).
[22]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982).
[23]O. Castaños and A. Frank, J. Math. Phys. 25, 388 (1984).
[24]G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **47**, 223 (1981); D. J. Rowe and G. Rosensteel, Phys. Rev. C **25**, 3236 (1982).
[25]O. Castaños, E. Chacón, A. Frank, P. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[26]M. Moshinsky, SIAM, J. Appl. Math. **25**, 193 (1973).
[27]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968).
[28]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[29]I. M. Gel'fand and M. L. Zettin, Dokl. Akad. Nauk, SSSR **71**, 825 (1950).
[30]R. Gaskell, G. Rosensteel, and R. T. Sharp, J. Math. Phys. **22**, 2732 (1981).
[31]L. Sabaliauskas, Liet. Fiz. Rinkinys **19**, 5 (1979).
[32]M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[33]A. Frank, "Group Theory in the Interacting Boson Model," PhD thesis, University of Mexico, 1979; O. Castaños, "Collective Degrees of Freedom in Nuclear Reactions," PhD thesis, University of Mexico, 1979 (both in Spanish). P. O. Hess, "General Collective Model and the Periodic System of Potential Energy Surfaces," PhD thesis, University of Frankfurt, 1980 (in German).
[34]E. M. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1951), p. 119.
[35]G. Gneuss and W. Greiner, Nucl. Phys. **171**, 449 (1971); P. O. Hess, M. Seiwert, J. Maruhn, and W. Greiner, Z. Phys. A **269**, 147 (1980); P. O. Hess, J. Maruhn, and W. Greiner, J. Phys. **67**, 737 (1981).
[36]E. Chacón, P. O. Hess, and M. Moshinsky, J. Math. Phys. **25**, 1565 (1984).

Collectivity and geometry. II. The two-dimensional case

E. Chacón, P. Hess,[a] and M. Moshinsky[b]
Instituto de Física, UNAM, Apdo. Postal 20-364, 01000 México, D. F., Mexico

(Received 3 March 1983; accepted for publication 22 July 1983)

In a previous paper of this series we showed that nuclear collective behavior can be related to the symplectic geometry of the many-nucleon system, when we introduce in it appropriate constraints. We showed in that paper that a full discussion of collective behavior in the many-body system requires—in a space of d dimensions—the basis for the irreducible representations of the Sp($2d$) group in the chain Sp($2d$) $\supset$ Sp(2) $\times \mathcal{O}(d)$, as well as the matrix elements of the generators of Sp($2d$) in this basis. In the present paper we implement this program fully for $d = 2$, both because of the mathematical interest of the chain Sp(4) $\supset$ Sp(2) $\times \mathcal{O}(2)$ and as a stepping stone to the chain Sp(6) $\supset$ Sp(2) $\times \mathcal{O}(3)$ which is central to the discussion of collective motions.

PACS numbers: 21.60.Ev, 21.60.Fw, 03.65.Ge

1. INTRODUCTION

In a previous paper[1] under the same general title (to be referred as I and whose equations will be quoted by their number followed by I) we showed that certain type of collective effects in nuclei follow from many-body theory, if we establish appropriate constraints in the symplectic geometry of the A-nucleon system.

In the present paper we intend to discuss fully the problem for an arbitrary number of nucleons, A (and not only for $A \to \infty$ as in I but when our space has only two dimensions, i.e., $d = 2$). Of course, our analysis will have physical relevance only when $d = 3$, but, as the problem is quite a difficult one, we need first to understand it when $d = 2$. Furthermore, the results obtained in this paper are relevant to the mathematical question of the states that are basis for irreducible representations (irreps) of the Sp(4) group, whose rows are characterized by the Sp(2) $\times \mathcal{O}(2)$ subgroup, and also to the derivation of the matrix elements of the generators of Sp(4) with respect to this basis.

As indicated in I, we eliminate from the beginning the center of mass of the A-body system, and thus deal with $n = A - 1$ Jacobi coordinates and momenta in this two-dimensional space, i.e.,

$$x_{is}, p_{is}\,, \quad i = 1,2\,, \quad s = 1,2,\ldots,n\,. \tag{1.1}$$

As indicated in Sec. 9 of I, our problem is then to get the matrix elements of

$$q_{ij} = \sum_{s=1}^{n} x_{is} x_{js} \tag{1.2}$$

with respect to the eigenstates of

$$H_0 = \frac{1}{2} \sum_{s=1}^{n} \sum_{i=1}^{2} (p_{is}^2 + x_{is}^2)\,, \tag{1.3}$$

when the former are characterized by the chain of groups

$$\mathrm{Sp}(4n) \supset \overset{[n/2+\omega_2, n/2+\omega_1]}{\mathrm{Sp}(4)} \times \overset{(\omega_1,\omega_2)}{\mathrm{O}(n)}\,, \tag{1.4a}$$

$$\mathrm{Sp}(4) \supset \overset{\lambda}{\mathrm{Sp}(2)} \times \underset{M}{\mathcal{O}(2)}\,, \tag{1.4b}$$

where above or below the groups we have indicated the quantum numbers that characterize their irreps.

As q_{ij} is a scalar in O(n), we do not need to consider subgroups of this group to characterize the states of H_0 and, in most of our discussion, will assume that the states are of highest weight in the chain O(n) $\supset$ O($n-1$) $\supset \cdots \supset$ O(2).

We shall not go directly into our general problem but first illustrate the procedure in particular cases where the analysis is very simple. Thus in Sec. 2 we deal with the problem when $n = 2$, while in Sec. 3 we discuss the case when the irrep of O(n) is (0,0) and thus that of Sp(4) is $[n/2, n/2]$.

In Sec. 4 we determine the eigenstates of H_0 characterized by the chain of groups (1.4) when the irrep of O(n) is $(\omega,0)$. As indicated in Sec. 9 of I, this gives the most general irrep $[n/2, n/2 + \omega]$ of Sp(4) as it depends on two parameters n,ω.

In Sec. 5 we deal with the matrix elements of q_{ij} with respect to the states mentioned while in the concluding Sec. 6 we discuss the problem of shape and also what parts of our analysis are likely to help in the problem when $d = 3$.

Before starting our discussion we establish the notation that we will be employing in the following sections. To begin with, as in Sec. 6 of I, we will be using the Zickendraht[2] and Dzublik *et al.*[3] (ZD) coordinate system through the definitions

$$x_{is} = \sum_{k=1}^{2} \rho_k D^{\frac{1}{2}}_{ki}(\vartheta) D^{\frac{1}{2}}_{n-2+k,s}(\phi)\,, \tag{1.5}$$

where

$$\mathbf{D}^{\frac{1}{2}}(\vartheta) = \| D^{\frac{1}{2}}_{ki}(\vartheta) \| = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \tag{1.6}$$

[a] Present address: Institut für Theoretische Physik, Universität Frankfurt, West Germany.
[b] Member of El Colegio Nacional and Instituto Nacional de Investigaciones Nucleares.

 0022-2488/84/051565-12$02.50

is the defining irrep of $\mathcal{O}(2)$, while $\|D^{\frac{1}{2}}_{is}(\phi)\|$ is the defining irrep of O(n). From (1.2) and (1.5) it is then clear that

$$\mathbf{q} = \|q_{ij}\| = \tilde{\mathbf{D}}(\vartheta)\begin{bmatrix}\rho_1^2 & 0\\ 0 & \rho_2^2\end{bmatrix}\mathbf{D}^1(\vartheta)$$
$$= \begin{bmatrix}\frac{1}{2}(\rho_1^2+\rho_2^2)+\frac{1}{2}(\rho_1^2-\rho_2^2)\cos 2\vartheta & \frac{1}{2}(\rho_1^2-\rho_2^2)\sin 2\vartheta\\ \frac{1}{2}(\rho_1^2-\rho_2^2)\sin 2\vartheta & \frac{1}{2}(\rho_1^2+\rho_2^2)-\frac{1}{2}(\rho_1^2-\rho_2^2)\cos 2\vartheta\end{bmatrix}. \tag{1.7}$$

Introducing now the variables

$$\rho_1 = \rho\cos\gamma\,,\quad \rho_2 = \rho\sin\gamma \tag{1.8}$$

and

$$r = \rho^2/2\,,\quad \theta = 2\gamma + \pi/2\,,\quad \varphi = 2\vartheta\,, \tag{1.9}$$

we obtain

$$\mathbf{q} = \begin{bmatrix}r + r\sin\theta\cos\varphi & r\sin\theta\sin\varphi\\ r\sin\theta\sin\varphi & r - r\sin\theta\cos\varphi\end{bmatrix}. \tag{1.10}$$

Finally instead of $q_{ij} = q_{ji}$, $i,j = 1,2$, we introduce the three variables $\bar{q} = \frac{1}{2}\,\mathrm{Tr}\,\mathbf{q}$, $q_{\pm} = \frac{1}{2}(q_{11} - q_{22}) \pm iq_{12}$ to obtain

$$\bar{q} = r\,,\quad q_{\pm} = r\sin\theta\, e^{\pm i\varphi}\,, \tag{1.11}$$

which are the operators whose matrix elements we wish to calculate.

Besides the coordinates and momenta x_{is}, p_{is}, we shall also need the creation and annihilation operators η_{is}, ξ_{is} given by the standard definitions

$$\eta_{is} = (1/\sqrt{2})(x_{is} - ip_{is})\,,\quad \xi_{is} = (1/\sqrt{2})(x_{is} + ip_{is})\,. \tag{1.12}$$

Furthermore, instead of the components $i = 1,2$ of all the vectors, we shall use their polar form through the definition

$$x_{\pm,s} = (1/\sqrt{2})(x_{is} \pm ix_{2s})\,. \tag{1.13}$$

As having expressions with two indices like $x_{\pm,s}$ will prove cumbersome, we introduce the notation

$$x_s \equiv x_{+,s}\,,\quad \bar{x}_s \equiv x_{-,s} \tag{1.14}$$

and similar definitions for $p_s, \bar{p}_s, \eta_s, \bar{\eta}_s$, and $\xi_s, \bar{\xi}_s$.

For application to our states of the generators of the O(n) group, it will prove necessary not to speak of components $s = 1,2,\ldots,n$ of x_s, p_s, η_s, ξ_s and those with bars, but rather to introduce what we could call their "hyperspherical" form. For this purpose we define j as

$$2j + 1 = n\,, \tag{1.15}$$

(which implies that j is integer when n is odd and half-integer when n is even) and introduce the index

$$m = j, j-1, j-2, \ldots, -j\,, \tag{1.16}$$

which will replace the $s = 1,2,\ldots,n$. For n odd we define as

$$x_{\pm j} = (1/\sqrt{2})(x_1 \pm ix_2)\,,$$
$$x_{\pm(j-1)} = (1/\sqrt{2})(x_3 \pm ix_4)\,,\quad \ldots\,,$$
$$x_{\pm 1} = (1/\sqrt{2})(x_{n-2} \pm ix_{n-1})\,,\quad x_0 = x_n\,, \tag{1.17a}$$

while for n even we have

$$x_{\pm j} = 1/\sqrt{2}(x_1 \pm ix_2)\,,$$
$$x_{\pm(j-1)} = (1/\sqrt{2})(x_3 \pm ix_4)\,,\quad \ldots\,,$$
$$x_{\pm 1/2} = (1/\sqrt{2})(x_{n-1} \pm ix_n)\,, \tag{1.17b}$$

and similarly for $\eta_m, \bar{x}_m, \bar{\eta}_m$.

As we have the relation

$$\sum_{s=1}^{n} x_s^2 = \sum_{m=-j}^{j} x_m x_{-m}\,, \tag{1.18}$$

we see that in "hyperspherical" components the metric tensor[4] is

$$g_{mm'} = \delta_{m,-m'}\,. \tag{1.19}$$

The momenta and annihilation operators can, from their commutation relation with the coordinate and creation operators, be defined as

$$p^m = \frac{1}{i}\frac{\partial}{\partial x_m}\,,\quad \xi^m = \frac{\partial}{\partial \eta_m} \tag{1.20}$$

and thus have a contravariant[4] rather than covariant form. From the metric tensor (1.19) we conclude that

$$p_m = p^{-m}\,,\quad \xi_m = \xi^{-m} \tag{1.21}$$

and similarly for the barred expressions.

The generators of the unitary group U(n) associated with the particle index $s = 1,2,\ldots,n$ are given by[5]

$$\Gamma_{ss'} = \sum_{i=1}^{2} \eta_{is}\xi_{is'}\,, \tag{1.22}$$

and clearly in "hyperspherical" coordinates they become

$$\Gamma^{m'}_{m} = \eta_m\xi^{m'} + \bar{\eta}_m\bar{\xi}^{m'} \equiv \mathscr{C}^{m'}_{m} + \overline{\mathscr{C}}^{m'}_{m}\,. \tag{1.23}$$

If we take the covariant form of these generators, i.e., $\Gamma_{mm'}$, we can define their antisymmetric part by

$$\Lambda_{mm'} = (\Gamma_{mm'} - \Gamma_{m'm})\,, \tag{1.24}$$

which in mixed tensor form becomes

$$\Lambda^{m'}_{m} = (\Gamma^{m'}_{m} - \Gamma^{-m}_{-m'})\,. \tag{1.25}$$

It is clear that $\Lambda^{-m}_{m} = 0$ and $\Lambda^{m'}_{m} = -\Lambda^{-m}_{-m'}$, and thus there are only

$$j(2j+1) = (n/2)(n-1) \tag{1.26}$$

independent operators of this type. Furthermore, they close under commutation,[5] and we have that

$$\left[\Lambda^{m'}_{m}, \sum_{m''}\eta_{m''}\eta_{-m''}\right]$$
$$= \left[\Gamma^{m'}_{m} - \Gamma^{-m}_{-m'}, 2\eta_m\eta_{-m} + 2\eta_{m'}\eta_{-m'}\right] = 0\,. \tag{1.27}$$

Clearly then the $\Lambda^{m'}_{m}$ of (1.25), where $m, m' = j, j-1, \ldots, -j$ are the generators of an O($2j+1$) = O(n) group.

The generators $\Lambda^{m'}_{m}$, $m + m' > 0$, can be separated into three subsets that respectively raise, give, and lower the weight, i.e.,[5]

$$\Lambda^{m'}_{m},\ m > m'\,,\quad \Lambda^{m}_{m}\,,\quad \Lambda^{m'}_{m},\ m < m'\,. \tag{1.28}$$

The highest weight state in O(n) would then become a poly-

nomial $P(\eta_m,\bar{\eta}_m)$ acting on the ground state $|0\rangle$ such that[5]

$$\Lambda^{m'}_{m} P(\eta_m,\bar{\eta}_m)|0\rangle = 0 \quad \text{if } m > m' ,$$
$$\Lambda^{m}_{m} P(\eta_m,\bar{\eta}_m)|0\rangle = L_m P(\eta_m,\bar{\eta}_m)|0\rangle , \tag{1.29}$$

where, in L_m, $m = j, j-1, j-2, \ldots, 1$ if j is integer (n odd) or $m = j, j-1, j-2, \ldots, \frac{1}{2}$ if j is half-integer (n even). The weight of the state and thus the irrep of $O(n) = O(2j+1)$ is given by

$$(L_j, L_{j-1}, L_{j-2}, \cdots) . \tag{1.30}$$

We note furthermore that the generators $\Lambda^{m'}_{m}$ can be written as

$$\Lambda^{m'}_{m} = \mathscr{L}^{m'}_{m} + \bar{\mathscr{L}}^{m'}_{m} , \tag{1.31}$$

where

$$\mathscr{L}^{m'}_{m} = \mathscr{C}^{m'}_{m} - \mathscr{C}^{-m}_{-m'} , \tag{1.32}$$

and similarly for the barred expression.

Having established our notation we proceed now to discuss the problem when $n = 2$.

2. THE CASE WHEN $n = 2$

The problem of $A = 3$ particles in a space of two dimensions corresponds then to $n = A - 1 = 2$, and from (1.5) its coordinates x_{is}, $i = 1,2$, $s = 1,2$, can be expressed in terms of the ZD variables as

$$x_{is} = \sum_{k=1}^{2} \rho_k D^{1}_{ki}(\vartheta) D^{1}_{ks}(\phi) , \quad i,s = 1,2, \tag{2.1}$$

where $\mathbf{D}^1$ is given by (1.6) in terms of the angles indicated.

Writing ρ_1, ρ_2 in terms of ρ, γ as in (1.8), we then obtain[6]

$$x_{11} = \rho(dbc + trs) , \tag{2.2a}$$
$$x_{12} = \rho(-dbs + trc) , \tag{2.2b}$$
$$x_{21} = \rho(-drc + tbs) , \tag{2.2c}$$
$$x_{22} = \rho(drs + tbc) , \tag{2.2d}$$

where

$$b = \cos\vartheta , \quad r = \sin\vartheta , \quad c = \cos\phi ,$$
$$s = \sin\phi , \quad d = \cos\gamma , \quad t = \sin\gamma . \tag{2.3}$$

The Hamiltonian H_0 of the form

$$H_0 = \frac{1}{2} \sum_{i=1}^{2} \sum_{s=1}^{2} \left(-\frac{\partial^2}{\partial x_{is}^2} + x_{is}^2 \right) \tag{2.4}$$

then becomes[6]

$$H_0 = \frac{1}{2}\Big\{ -\frac{1}{\rho^3}\frac{\partial}{\partial\rho}\rho^3\frac{\partial}{\partial\rho} - \frac{1}{\rho^2}\Big[\frac{1}{\cos 2\gamma}\frac{\partial}{\partial\gamma}\cos 2\gamma\frac{\partial}{\partial\gamma} + \frac{1}{\cos^2 2\gamma}\Big(\frac{\partial^2}{\partial\vartheta^2} + \frac{\partial^2}{\partial\phi^2}\Big) + \frac{2\sin 2\gamma}{\cos^2 2\gamma}\frac{\partial^2}{\partial\vartheta\,\partial\phi}\Big] + \rho^2\Big\} . \tag{2.5}$$

Introducing now the variables r,θ,φ through the definition (1.9), plus a χ given by

$$\chi = 2\phi , \tag{2.6}$$

we obtain that

$$H_0 = r\Big(-\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1}{r^2}\mathbf{L}^2 + 1 \Big) , \tag{2.7}$$

where

$$\mathbf{L}^2 = -\Big[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\Big(\frac{\partial}{\partial\varphi^2} + \frac{\partial^2}{\partial\chi^2}\Big) - \frac{2\cos\theta}{\sin^2\theta}\frac{\partial^2}{\partial\varphi\,\partial\chi}\Big] \tag{2.8}$$

is the Hamiltonian of the spherical top.[7]

Note that $4\mathbf{L}^2$ is related to the Casimir operator of the O(4) group in the chain $Sp(8) \supset Sp(2) \times O(4)$. Thus the eigenvalue $2L(2L+2)$ of $4\mathbf{L}^2$ is related to the eigenvalue $\lambda(\lambda-1)$ of the Casimir operator of Sp(2). From a reasoning similar to that given in (27I) we conclude that $\lambda = L + 1$.

The eigenket of H_0 can be written as

$$|\nu KLM\rangle = r^{-1} F^{L+1/2}_{\nu}(r) \Big(\frac{2L+1}{8\pi^2}\Big)^{1/2} D^{L}_{KM}(\varphi,\theta,\chi) , \tag{2.9}$$

where D^{L}_{KM} is the Wigner irrep of O(3) related to the $\mathscr{D}^{L}_{KM}$ of Rose's book[8] through

$$D^{L}_{KM}(\varphi,\theta,\chi) = \mathscr{D}^{L*}_{MK}(\varphi,\theta,\chi) = e^{iM\varphi} d^{L}_{MK}(\theta) e^{iK\chi} . \tag{2.10}$$

The normalized radial function is given by[9]

$$F^{L+1/2}_{\nu}(r) = 2^{L+1/2}\Big[\frac{2(\nu!)}{\Gamma(\nu+2L+2)}\Big]^{1/2} \times r^{L+1} e^{-r} L^{2L+1}_{\nu}(2r) , \tag{2.11}$$

where L^{2L+1}_{ν} is a Laguerre polynomial and the volume element for the problem is

$$r\,dr\sin\theta\,d\theta\,d\varphi\,d\chi \tag{2.12}$$

with the variables being in the intervals

$$0 \leqslant r \leqslant \infty , \tag{2.13}$$
$$0 \leqslant \theta \leqslant \pi , \tag{2.14}$$
$$0 \leqslant \varphi, \chi \leqslant 2\pi . \tag{2.15}$$

As $n = 2$, $O(n) = O(2)$ and its only generator is $-i\,\partial/\partial\phi = -(i/2)\,\partial/\partial\chi$, so that the irrep of $O(n)$ is $(\omega,0) = (K/2,0)$. Therefore, the irrep of Sp(4) corresponding to $A = n + 1 = 3$ particles is then given by

$$[1, 1 + (K/2)] . \tag{2.16}$$

We need now to calculate the matrix elements of

$$q_{\pm} = r\sin\theta\, e^{\pm i\varphi} = \mp(8\pi/3)^{1/2} r Y_{1,\pm 1}(\theta,\varphi) = \mp\sqrt{2} r D^{1}_{0,\pm 1}(\varphi,\theta,\chi) , \tag{2.17a}$$
$$\bar{q} = r = r D^{0}_{00}(\varphi,\theta,\chi) \tag{2.17b}$$

with respect to the states $|\nu KLM\rangle$. As the integration involving three D functions is well known,[8] and that involving the radial integral can also be determined,[9,10] we obtain

$$\langle \nu' KL'M' | q_{\pm} | \nu KLM \rangle = \mp [2(2L+1)/(2L'+1)]^{1/2} \langle LM;1,\pm 1|L'M'\rangle \times \langle LK;10|L'K\rangle \int_0^{\infty} F^{L'+1/2}_{\nu'}(r) F^{L+1/2}_{\nu}(r)\,dr , \tag{2.18}$$

$$\langle \nu' KLM | \bar{q} | \nu KLM \rangle = \int_0^{\infty} F^{L+1/2}_{\nu'}(r) F^{L+1/2}_{\nu}(r)\,dr , \tag{2.19}$$

where the radial integrals will be denoted by

$$\langle \nu' L' | \nu L \rangle \equiv \int_0^{\infty} F^{L'+1/2}_{\nu'}(r) F^{L+1/2}_{\nu}(r)\,dr , \tag{2.20a}$$

which, from properties of the Laguerre polynomials,[10] take the values

$$\langle \nu' L \,|\nu L\rangle = -\tfrac{1}{2}[(\nu+1)(\nu+2L+2)]^{1/2}\delta_{\nu',\nu+1} + (\nu+L+1)\delta_{\nu'\nu} - \tfrac{1}{2}[\nu(\nu+2L+1)]^{1/2}\delta_{\nu',\nu-1}, \tag{2.20b}$$

$$\langle \nu' L+1|\nu L\rangle = \tfrac{1}{2}[(\nu+2L+3)(\nu+2L+2)]^{1/2}\delta_{\nu'\nu} - [\nu(\nu+2L+2)]^{1/2}\delta_{\nu',\nu-1} + \tfrac{1}{2}[\nu(\nu-1)]^{1/2}\delta_{\nu',\nu-2}, \tag{2.20c}$$

$$\langle \nu' L-1|\nu L\rangle = \tfrac{1}{2}[(\nu+2L+1)(\nu+2L)]^{1/2}\delta_{\nu'\nu} - [(\nu+2L+1)(\nu+1)]^{1/2}\delta_{\nu',\nu+1} + \tfrac{1}{2}[(\nu+2)(\nu+1)]^{1/2}\delta_{\nu',\nu+2}. \tag{2.20d}$$

We note furthermore that the eigenvalue of the operator H_0 of (2.7) is[9]

$$2(\nu+l+1) \tag{2.21}$$

while from (2.4) it is given by $N+2$, where N is the total number of quanta. Thus we have

$$N = 2(\nu+l). \tag{2.22}$$

From (8I) and (2.18)–(2.21) we see that for $n=2$ we have determined in a direct fashion the matrix elements of all the generators of Sp(4) in the basis (2.9) of eigenstates of H_0.

3. THE CASE OF THE IRREP [n/2, n/2] OF Sp(4)

We now turn our attention to the problem of an arbitrary number of particles $A = n+1$, but whose states are characterized by the scalar representation of O(n), i.e., they belong[1] to the irrep $[n/2, n/2]$ of Sp(4). Our program then requires that we write the Hamiltonian (1.3) in terms of the ZD coordinates (1.5), which gives H_0 as a function of $\rho_1, \rho_2, \vartheta$ and their derivatives as well as of $\mathscr{L}'_{st}$, $s<t=2,3,\ldots,n$, where the latter are the "body-fixed" generators of O(n) defined in Ref. 11 as well as in (25I). The $\mathscr{L}'_{st}$ acting on states belonging to the scalar representation of O(n) give zero and thus the Hamiltonian (1.3) projected on this scalar representation, which we designate by $\bar{H}_0$, takes the form[12]

$$\bar{H}_0 = \frac{1}{2}\left[-\frac{\partial^2}{\partial\rho_1^2} - \frac{\partial^2}{\partial\rho_2^2} - (n-2)\left(\frac{1}{\rho_1}\frac{\partial}{\partial\rho_1} + \frac{1}{\rho_2}\frac{\partial}{\partial\rho_2}\right) - \frac{2}{(\rho_1^2-\rho_2^2)}\left(\rho_1\frac{\partial}{\partial\rho_1} - \rho_2\frac{\partial}{\partial\rho_2}\right) - \frac{(\rho_1^2+\rho_2^2)}{(\rho_1^2-\rho_2^2)^2}\frac{\partial^2}{\partial\vartheta^2} + (\rho_1^2+\rho_2^2)\right]. \tag{3.1}$$

Denoting by $\Psi_E(\rho_1,\rho_2,\vartheta)$ the eigenstate of the Hamiltonian (3.1) associated with the eigenvalue E, we can write it as

$$\Psi_E = (\rho_1\rho_2)^{-(n-2)/2}\psi_E, \tag{3.2}$$

and, passing from $\rho_1, \rho_2, \vartheta$ to r,θ,φ, related to them by (1.9), we obtain for ψ_E the following equation[12]:

$$\bar{H}'_0\psi_E \equiv r\left[-\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + \frac{\Lambda^2}{r^2} + 1\right]\psi_E = E\psi_E, \tag{3.3}$$

where

$$\Lambda^2 = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} - \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2} + \frac{(n-2)(n-4)}{4\cos^2\theta}. \tag{3.4}$$

As shown in Ref. 12, the Λ^2 is actually the Casimir operator

$$\Lambda^2 = \mathscr{N}'^2 - K'^2_4 - N'^2_4 \tag{3.5}$$

of an Sp(2) group whose generators are $\mathscr{N}', K'_4, N'_4$ satisfying the commutation relations

$$[K'_4, N'_4] = -i\mathscr{N}', \quad [\mathscr{N}', K'_4] = iN'_4, \quad [N'_4, \mathscr{N}'] = iK'_4. \tag{3.6}$$

The explicit expression for these operators is[12]

$$\mathscr{N}' = \frac{1}{2}\bar{H}'_0, \quad K'_4 = \frac{1}{2}\bar{H}'_0 - r, \quad N'_4 = \frac{1}{i}\left(r\frac{\partial}{\partial r} + 1\right). \tag{3.7}$$

The eigenvalue of the operator Λ^2 can be denoted, as in Sec. 8 of I, by $\lambda(\lambda-1)$, and writing

$$\lambda = l + (n/2), \quad l = 0,1,2,\ldots, \tag{3.8}$$

we can express the eigenfunction $\Phi_{lm}(\theta,\varphi)$ of Λ^2, i.e.,

$$\Lambda^2\Phi_{lm}(\theta,\varphi) = (l+n/2)(l+n/2-1)\Phi_{lm}(\theta,\varphi), \tag{3.9}$$

in the form[12]

$$\Phi_{lm}(\theta,\varphi) = (\sin\theta)^{|m|}(\cos\theta)^{(n-2)/2} \times P^{(|m|,[n-3]/2)}_{(l-|m|)/2}(\cos 2\theta)e^{im\varphi}, \tag{3.10}$$

where P is a Jacobi polynomial[13] and $m = l, l-2, l-4, \ldots, -l$.

From (3.3) the radial part, which we can write as $r^{-1}R(r)$, satisfies the equation

$$\frac{1}{2}r\left\{-\frac{d^2}{dr^2} + \frac{[l+(n-1)/2]^2 - \frac{1}{4}}{r^2} + 1\right\}R(r) = \frac{1}{2}ER(r). \tag{3.11}$$

The energy E takes then the value[9]

$$E = 2(\nu+l) + n, \quad \nu = 0,1,2,\ldots, \tag{3.12}$$

and the radial part of the wave function becomes[9]

$$R_{\nu l} = r^{l+(n/2)}e^{-r}L_\nu^{2l+n-1}(2r), \tag{3.13}$$

where L_ν^{2l+n-1} is a Laguerre polynomial.

Finally the eigenket of $\bar{H}'_0$ of (3.3) can be written as

$$|n;\nu lm\rangle = B_{\nu lm}r^{-1}R_{\nu l}(r)\Phi_{lm}(\theta,\varphi), \tag{3.14}$$

where the normalization constant is given by[12]

$$B_{\nu lm} = (-)^{(m+|m|)/2}\left[\frac{\nu!2^{2l+n}[(l-|m|)/2]!(2l+n-1)\Gamma((l+|m|+n-1)/2)}{2\pi(\nu+2l+n-1)![(l+|m|)/2]!\Gamma((l-|m|+n-1)/2)}\right]^{1/2}. \tag{3.15}$$

The phase factor in (3.15) was chosen to guarantee that

$$|n;\nu lm\rangle^* = (-)^m |n;\nu l,-m\rangle\,, \tag{3.16}$$

where we put the extra parameter $n = A - 1$ in the ket to indicate the number of particles to which the state belongs and thus also the irrep $[n/2,n/2]$ of Sp(4). The volume element is $r\,dr \sin\theta\,d\theta\,d\varphi$.

We now need the matrix elements of $r\sin\theta\exp(\pm i\varphi)$ and r with respect to the states $|n;\nu lm\rangle$ of (3.14). As this was done in Ref. 12, we just quote the results here:

$$\begin{aligned}\langle n;\nu' l' m'|r|n;\nu lm\rangle = &\{(\nu + l + n/2)\delta_{\nu'\nu} \\ &- \tfrac{1}{2}[\nu(\nu + 2l + n - 1)]^{1/2}\delta_{\nu',\nu-1} \\ &- \tfrac{1}{2}[(\nu+1)(\nu + 2l + n)]^{1/2}\delta_{\nu',\nu+1}\} \\ &\times\delta_{l'l}\delta_{m'm}\,,\end{aligned} \tag{3.17a}$$

$$\begin{aligned}&\langle n;\nu' l' m'|r\sin\theta\, e^{\pm i\varphi}|n;\nu lm\rangle \\ &= \left[\frac{(l\pm m + n - 1)(l\pm m + 2)}{(2l + n - 1)(2l + n + 1)}\right]^{1/2}\delta_{l',l+1}\delta_{m',m\pm1} \\ &\quad\times\{\pm[\nu(\nu + 2l + n)]^{1/2}\delta_{\nu',\nu-1} \mp \tfrac{1}{2}[\nu(\nu-1)]^{1/2} \\ &\quad\times\delta_{\nu',\nu-2} \mp [(\nu + 2l + n)(\nu + 2l + n + 1)]^{1/2}\delta_{\nu'\nu}\} \\ &\quad+\left[\frac{(l\mp m + n - 3)(l\mp m)}{(2l + n - 3)(2l + n - 1)}\right]^{1/2}\delta_{l',l-1}\delta_{m',m\pm1} \\ &\quad\times\{\mp[(\nu+1)(\nu + 2l + n - 1)]^{1/2}\delta_{\nu',\nu+1} \\ &\quad\pm\tfrac{1}{2}[(\nu+1)(\nu+2)]^{1/2}\delta_{\nu',\nu+2} \\ &\quad\pm\tfrac{1}{2}[(\nu + 2l + n - 2)(\nu + 2l + n - 1)]^{1/2}\delta_{\nu'\nu}\}\,.\end{aligned} \tag{3.17b}$$

Thus we have been able to implement our program for the irrep $[n/2,n/2]$ of Sp(4), when n is arbitrary.

4. THE GENERAL CASE: DETERMINATION OF THE STATES

The Hamiltonian H_0 of (1.3) is a $2n$-dimensional oscillator which admits a symmetry group U($2n$) whose linear Casimir operator

$$\hat{N} = \sum_m (\eta_m\xi^m + \bar{\eta}_m\bar{\xi}^m) = \sum_m \Gamma^m_m \tag{4.1}$$

gives the number of quanta N when applied to the state

$$P(\eta_m,\bar{\eta}_m)|0\rangle\,, \tag{4.2}$$

where P is an homogeneous polynomial of degree N in the variables indicated and $|0\rangle$ is the ground state.

We would like now to characterize the state (4.2) by irreps of subgroups of U($2n$) that are also contained in the chain (1.4) that was relevant for our collective problem. Clearly we should then consider

$$\begin{array}{ccccc} \mathrm{U}(2n)\supset & \mathscr{U}(2) & \times & \mathrm{U}(n) \\ & \cup & & \cup \\ & \mathscr{O}(2) & & \mathrm{O}(n) \end{array} \tag{4.3}$$

as $\mathscr{O}(2)$, O(n) are in (1.4), and look first for states that correspond to definite irreps N of U($2n$), M of $\mathscr{O}(2)$, and $(\omega 0)$ of O(n). For this we require that (4.2) satisfy

$$\hat{N}P|0\rangle = NP|0\rangle\,, \tag{4.4a}$$

$$\hat{L}_0 P|0\rangle = MP|0\rangle\,, \tag{4.4b}$$

$$\Lambda^j_j P|0\rangle = \omega P|0\rangle\,, \tag{4.4c}$$

$$\Lambda^{|m|}_{|m|}P|0\rangle = 0\,,\quad m = j-1, j-2,\ldots,-j+1\,, \tag{4.4d}$$

$$\Lambda^{m'}_{m}P|0\rangle = 0\,,\quad m > m',\ m + m' > 0,\ m,m' = j,\ldots,-j\,, \tag{4.4e}$$

where

$$\begin{aligned}\hat{L}_0 &= \sum_{s=1}^{n}(x_{1s}p_{2s} - x_{2s}p_{1s}) = \sum_m(\eta_m\xi^m - \bar{\eta}_m\bar{\xi}^m) \\ &= \frac{1}{i}\frac{\partial}{\partial\vartheta}\end{aligned} \tag{4.5}$$

and $\Lambda^{m'}_m$ are the generators of $\mathrm{O}(n) = \mathrm{O}(2j+1)$ discussed in the Introduction. As we mentioned there, we can take our state to be of highest weight in the chain $\mathrm{O}(2j+1) \supset \mathrm{O}(2\,\mathrm{j})\supset\cdots\supset\mathrm{O}(2)$ which is what (4.4e) implies.

We proceed to derive explicitly the P that satisfies (4.4). Once we achieve this objective we will replace the η_m, $\bar{\eta}_m$ that are standard boson operators by traceless boson operators,[14,15] so that automatically we get eigenstates of O($2n$) rather than U($2n$). As Sp($4n$) is the dynamical group of the $2n$-dimensional oscillator, it can also have the subgroup

$$\mathrm{Sp}(4n)\supset\mathrm{Sp}(2)\times\mathrm{O}(2n)\,, \tag{4.6}$$

in which, as discussed in I, Sp(2) and O($2n$) are complementary.[16] Thus the irrep of O($2n$) determines that of Sp(2), and therefore our states are characterized by the irrep of another of the groups that appears in (1.4).

So far our analysis is in terms of polynomials of creation operators acting on $|0\rangle$, but, by Dragt's theorem[17], they can be converted into the same polynomials in the coordinates $P(x_m,\bar{x}_m)$ multiplied by $\exp(-\rho^2/2)$. To get then the most general state, we have just to multiply the $P(x_m,\bar{x}_m)$ by an appropriate Laguerre polynomial[15] in ρ^2. We get thus an explicit expression for the states of H_0 characterized by the irreps of the chain of groups (1.4). We will see though that these irreps are not sufficient to determine the state and that in fact we need in the chain $\mathrm{O}(2n)\supset\mathscr{O}(2)\times\mathrm{O}(n)$ one extra multiplicity index. Thus our states will not be orthonormal in the latter.

Returning now to our state (4.2), we clearly see from (4.4a), (4.4b) and (4.1), (4.5) that it is a homogeneous polynomial of degrees $(N+M)/2$ and $(N-M)/2$ in η_m and $\bar{\eta}_m$, respectively. Thus we can write

$$P(\eta_m,\bar{\eta}_m) = \eta_j^{(N+M)/2}\bar{\eta}^{(N-M)/2}P'(\eta_m/\eta_j,\bar{\eta}_m/\bar{\eta}_j)\,, \tag{4.7}$$

where $(N\pm M)$ must be even.

As a next step we denote the scalar products in O(n) of the creation operators by

$$\begin{aligned}&(\boldsymbol{\eta}\cdot\boldsymbol{\eta}) = \sum_{m=-j}^{j}\eta_m\eta_{-m}\,,\quad (\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}}) = \sum_{m=-j}^{j}\bar{\eta}_m\bar{\eta}_{-m}\,, \\ &(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}}) = \sum_{m=-j}^{j}\eta_m\bar{\eta}_{-m}\,,\end{aligned} \tag{4.8}$$

and note that

$$\frac{(\boldsymbol{\eta}\cdot\boldsymbol{\eta})}{\eta_j^2} = 2\frac{\eta_{-j}}{\eta_j} + \sum_{m=-j+1}^{j-1}\frac{\eta_m}{\eta_j}\frac{\eta_{-m}}{\eta_j} \tag{4.9}$$

and similarly for the corresponding barred expression. Thus in (4.7) we can replace

(η_{-j}/η_j), $(\bar\eta_{-j}/\bar\eta_j)$ by $(\boldsymbol\eta\cdot\boldsymbol\eta)/\eta_j^2$, $(\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta})/\bar\eta_j^2$ and be left with (η_m/η_j), $(\bar\eta_m/\bar\eta_j)$ with m in the range $m=-j+1,\ -j+2,...,j-1$. Furthermore, we can take sums and differences of the last ratios to finally write

$$P(\eta_m,\bar\eta_m)=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}\times P''\left(\frac{\eta_m}{\eta_j}-\frac{\bar\eta_m}{\bar\eta_j},\ \frac{\eta_m}{\eta_j}+\frac{\bar\eta_m}{\bar\eta_j},\ \frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2},\ \frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2}\right) \tag{4.10}$$

in which in P'', $m=j-1,j-2,...,-j+1$.

We now apply the operator Λ_j^m, $j>m$ to $P|0\rangle$ and require from (4.4e) that the result vanishes. We first see that $\Lambda_m^{m'}$, and thus also Λ_j^m, commute with scalars of O(n), i.e.,

$$[\Lambda_m^{m'},\boldsymbol\eta\cdot\boldsymbol\eta]=[\Lambda_m^{m'},\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}]=[\Lambda_m^{m'},\boldsymbol\eta\cdot\bar{\boldsymbol\eta}]=0\,. \tag{4.11}$$

We then note that, as

$$\Lambda_j^m=\Gamma_j^m-\Gamma_{-m}^{-j}\,,\quad m=j-1,j-2,...,-j+1\,, \tag{4.12}$$

and we eliminated expressions with index $m=-j$ in (4.10), the application of Λ_j^m reduces to that of Γ_j^m, i.e.,

$$\begin{aligned}\Lambda_j^m P|0\rangle&=(\mathscr{C}_j^m+\bar{\mathscr{C}}_j^m)P|0\rangle\\&=2\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}\left\{\frac{\partial P''}{\partial[(\eta_m/\eta_j)+(\bar\eta_m/\bar\eta_j)]}\right\}|0\rangle\,.\end{aligned} \tag{4.13}$$

Thus $[(\eta_m/\eta_j)+(\bar\eta_m/\bar\eta_j)]$, $m=j-1,j-2,...,-j+1$, cannot appear in P'' of (4.10) so our state takes the form

$$P|0\rangle=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}P'''\left(\frac{\eta_m}{\eta_j}-\frac{\bar\eta_m}{\bar\eta_j},\ \frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2},\ \frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\eta_j^2}\right), \tag{4.14}$$

in which in P''', $m=j-1,j-2,...,-j+1$.

We now wish to use the scalar product

$$\frac{(\boldsymbol\eta\cdot\bar{\boldsymbol\eta})}{\eta_j\bar\eta_j}=\sum_{m=-j}^{j}\frac{\eta_m}{\eta_j}\frac{\bar\eta_{-m}}{\bar\eta_j}\,, \tag{4.15}$$

to eliminate the index $m=-j+1$ in the expression (4.14). For this purpose we define

$$\begin{aligned}A_j&\equiv\frac{1}{2}\sum_{m=-j}^{j}\left(\frac{\eta_m}{\eta_j}-\frac{\bar\eta_m}{\bar\eta_j}\right)\left(\frac{\eta_{-m}}{\eta_j}-\frac{\bar\eta_{-m}}{\bar\eta_j}\right)\\&=\frac{1}{2}\left(\frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2}+\frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2}\right)-\left(\frac{\boldsymbol\eta\cdot\bar{\boldsymbol\eta}}{\eta_j\bar\eta_j}\right)=A_{j-1}\\&=\left(\frac{\eta_{j-1}}{\eta_j}-\frac{\bar\eta_{j-1}}{\bar\eta_j}\right)\left(\frac{\eta_{-j+1}}{\eta_j}-\frac{\bar\eta_{-j+1}}{\bar\eta_j}\right)+A_{j-2}\,,\end{aligned} \tag{4.16}$$

and see that we can write

$$\begin{aligned}&\left(\frac{\eta_{-j+1}}{\eta_j}-\frac{\bar\eta_{-j+1}}{\bar\eta_j}\right)\\&=\left(\frac{\eta_{j-1}}{\eta_j}-\frac{\bar\eta_{j-1}}{\bar\eta_j}\right)^{-1}\left[\frac{1}{2}\left(\frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2}+\frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2}\right)-\frac{\boldsymbol\eta\cdot\bar{\boldsymbol\eta}}{\eta_j\bar\eta_j}-A_{j-2}\right],\end{aligned} \tag{4.17}$$

where A_{j-2} contains factors

$$[(\eta_m/\eta_j)-(\bar\eta_m/\bar\eta_j)] \tag{4.18}$$

but only with values $m=j-2,j-3,...,-j+2$.

We can then write the state (4.14) as

$$\begin{aligned}P|0\rangle=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}&\times R\left(\frac{\eta_m}{\eta_j}-\frac{\bar\eta_m}{\bar\eta_j},\ \frac{\eta_{j-1}}{\eta_j}-\frac{\bar\eta_{j-1}}{\bar\eta_j},\right.\\&\left.\frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2},\ \frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2},\ \frac{\boldsymbol\eta\cdot\bar{\boldsymbol\eta}}{\eta_j\bar\eta_j}\right),\end{aligned} \tag{4.19}$$

where $m=j-2,j-3,...,-j+2$ and R is a rational function of the variables indicated, as from (4.17) it can have negative powers in $[(\eta_{j-1}/\eta_j)-(\bar\eta_{j-1}/\bar\eta_j)]$, though it contains nonnegative powers of all the other variables.

We now apply

$$\Lambda_{j-1}^m=\Gamma_{j-1}^m-\Gamma_{-m}^{-j+1}\,,\quad m=j-2,j-3,...,-j+2\,, \tag{4.20}$$

and, as $[(\eta_{-j+1}/\eta_j)-(\bar\eta_{-j+1}/\bar\eta_j)]$ was already eliminated through (4.17), we get

$$\begin{aligned}\Lambda_{j-1}^m P|0\rangle=(\mathscr{C}_{j-1}^m+\bar{\mathscr{C}}_{j-1}^m)P|0\rangle=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}&\times\left\{\frac{\partial R}{\partial[(\eta_m/\eta_j)-(\bar\eta_m/\bar\eta_j)]}\right\}\\&\times\left(\frac{\eta_{j-1}}{\eta_j}-\frac{\bar\eta_{j-1}}{\bar\eta_j}\right)|0\rangle=0\,.\end{aligned} \tag{4.21}$$

Thus we see that R cannot depend on $[(\eta_m/\eta_j)-(\bar\eta_m/\bar\eta_j)]$, $m=j-2,j-3,...,-j+2$ and the condition (4.4e) establishes that our polynomial has the form

$$P|0\rangle=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}\times R'\left(\frac{\eta_{j-1}}{\eta_j}-\frac{\bar\eta_{j-1}}{\bar\eta_j},\ \frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2},\ \frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2},\ \frac{\boldsymbol\eta\cdot\bar{\boldsymbol\eta}}{\eta_j\bar\eta_j}\right), \tag{4.22}$$

where R' is a rational function of the variables indicated as it can have negative powers of $[(\eta_{j-1}/\eta_j)-(\bar\eta_{j-1}/\bar\eta_j)]$ but positive of all the others.

We now apply Λ_{j-1}^{j-1} to $P|0\rangle$, which from (4.4d) must give zero, and obtain

$$\begin{aligned}\Lambda_{j-1}^{j-1}P|0\rangle=(\mathscr{C}_{j-1}^{j-1}+\bar{\mathscr{C}}_{j-1}^{j-1})P|0\rangle=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}&\cdot\{\partial R'/\partial[(\eta_{j-1}/\eta_j)-(\bar\eta_{j-1}/\bar\eta_j)]\}\\&\times[(\eta_{j-1}/\eta_j)-(\bar\eta_{j-1}/\bar\eta_j)]|0\rangle=0\end{aligned} \tag{4.23}$$

which implies that R' is independent of $[(\eta_{j-1}/\eta_j)-(\bar\eta_{j-1}/\bar\eta_j)]$ and thus is a polynomial in the other variables, i.e.,

$$\begin{aligned}P|0\rangle&=\eta_j^{(N+M)/2}\bar\eta_j^{(N-M)/2}P^{\mathrm{IV}}\left(\frac{\boldsymbol\eta\cdot\boldsymbol\eta}{\eta_j^2},\ \frac{\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta}}{\bar\eta_j^2},\ \frac{\boldsymbol\eta\cdot\bar{\boldsymbol\eta}}{\eta_j\bar\eta_j}\right)|0\rangle\\&=\sum_{l_i}a_{l_1l_2l_3l_4l_5}\,\eta_j^{l_1}\bar\eta_j^{l_2}(\boldsymbol\eta\cdot\boldsymbol\eta)^{l_3}(\bar{\boldsymbol\eta}\cdot\bar{\boldsymbol\eta})^{l_4}(\boldsymbol\eta\cdot\bar{\boldsymbol\eta})^{l_5}|0\rangle\,,\end{aligned} \tag{4.24}$$

where from the equations (4.4a)–(4.4c) we have that the exponents satisfy the relations

$$l_1+l_2+2l_3+2l_4+2l_5=N\,, \tag{4.25a}$$

$$l_1-l_2+2l_3-2l_4=M\,, \tag{4.25b}$$

$$l_1 + l_2 = \omega\,, \tag{4.25c}$$

with all the $l_i \geqslant 0$ as there are no divisibility problems.[15] Note that the step (4.23) does not apply if $j = 1$, i.e., $n = 3$, and so the above analysis is valid only for $n > 3$. For $n = 3$ a special discussion is necessary.[18]

Denoting now

$$l_4 = q\,, \quad l_5 = l \tag{4.26}$$

and expressing l_1, l_2, l_3 in terms of them with the help of (4.25), we arrive at the states

$$|NM\omega q l\rangle = \{(\eta_j)^{\omega-(N-M)/2+2q+l}(\bar{\eta}_j)^{(N-M)/2-2q-l} \times(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{(N-\omega)/2-q-l}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{l}\}|0\rangle\,, \tag{4.27}$$

where, as all the exponents must be nonnegative integers, we have the inequalities

$$l \geqslant 0, \quad q \geqslant 0\,, \quad \omega - \tfrac{1}{2}(N-M) + 2q + l \geqslant 0\,,$$
$$\tfrac{1}{2}(N-M) - 2q - l \geqslant 0\,, \quad \tfrac{1}{2}(N-\omega) - q - l \geqslant 0 \tag{4.28}$$

plus the fact that $N - M$ and $N - \omega$ must be even.

The state (4.27) is an eigenket of H_0 characterized by the irreps of the chain of groups

$$\mathrm{U}(2n) \supset \mathcal{O}(2) \times \mathrm{O}(n)\,. \tag{4.29}$$

Between the first and last groups we can fit though a subgroup O(2n), and we would like actually to have states characterized by the chain

$$\underset{N}{\mathrm{U}(2n)} \supset \underset{L}{\mathrm{O}(2n)} \supset \underset{M}{\mathcal{O}(2)} \times \underset{\omega}{\mathrm{O}(n)}\,, \tag{4.30}$$

where below each group we give the value of its irrep. Clearly, as U(2n) is characterized by the symmetric irrep N, the O(2n) will also be characterized by a partition with one row which we denote $L = N, N-2, N-4, \ldots, 1$ or 0. As indicated above, the importance of the irrep L of O(2n) for our problem is that it also implies states characterized by an irrep λ of the Sp(2) group in (1.4) where, by a reasoning similar to the one that led to (28I), but now for $d = 2$ instead of $d = 3$, we have

$$\lambda = (L + n)/2\,. \tag{4.31}$$

Two of the authors[15,19] had to face a similar problem to the one leading from (4.29) to (4.30) when they discussed the eigenstates of the Bohr–Mottelson vibrational Hamiltonian. There[15,19] they had to go from U(5)⊃O(3) to U(5)⊃O(5)⊃O(3), and they were able to achieve their objective by replacing the boson creation operators by traceless ones.[14]

We shall follow the steps of Refs. 15 and 19 but now for our present problem. We start by defining the traceless boson operators

$$a_m = \eta_m - 2(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})(2\hat{N} + 2n)^{-1}\bar{\xi}_m\,, \tag{4.32a}$$
$$\bar{a}_m = \bar{\eta}_m - 2(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})(2\hat{N} + 2n)^{-1}\xi_m\,, \tag{4.32b}$$

where $\hat{N}$ is given by (4.1).

We then put $l = 0$ in (4.27) to eliminate

$$(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}}) = \sum_{i=1}^{2}\sum_{s=1}^{n}\eta_{is}^2 = \rho^2 - H_0 - \rho\frac{\partial}{\partial\rho} - n \tag{4.33}$$

which is a scalar[15] of O(2n) that acts only on the radial part, and replace N by L and $\eta_m, \bar{\eta}_m$ by $a_m, \bar{a}_m$ to get a state $|LM\omega q\rangle$

$$= (a_j)^{\omega-(L-M)/2+2q}(\bar{a}_j)^{(L-M)/2-2q} \times(\mathbf{a}\cdot\mathbf{a})^{(L-\omega)/2-q}(\bar{\mathbf{a}}\cdot\bar{\mathbf{a}})^{q}|0\rangle\,, \tag{4.34}$$

which corresponds[15,19] to the irrep L of both U(2n) and O(2n) as well as to the irreps M, ω of $\mathcal{O}(2)$, O(n). It is now a question of rewriting the state (4.34) in terms of $\eta_m, \bar{\eta}_m$ and then, by Dragt's theorem,[17] in terms of $x_m, \bar{x}_m$.

We shall start by applying the last two terms in (4.34) to the ground state, i.e., by determining

$$P_0|0\rangle \equiv (\mathbf{a}\cdot\mathbf{a})^{(L-\omega)/2-q}(\bar{\mathbf{a}}\cdot\bar{\mathbf{a}})^{q}|0\rangle\,. \tag{4.35}$$

In the context of the present problem, the analysis of Ref. 19 indicates that $P_0|0\rangle$ can be expanded in terms of a polynomial in scalars of O(n) that can be formed from $\eta_m, \bar{\eta}_m$, acting on $|0\rangle$, i.e.,

$$P_0|0\rangle = \sum_{n_1 n_2 n_3} C_{n_1 n_2 n_3}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{n_1}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{n_2}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{n_3}|0\rangle\,, \tag{4.36}$$

subject to the following conditions.

(a) The total degree of the polynomial in the creation operators should be the same as in (4.35), i.e., $L - \omega$.

(b) The total angular momentum, i.e., the irrep of $\mathcal{O}(2)$, should be the same as in (4.35), i.e., $(L - \omega) - 4q$.

(c) The $P_0|0\rangle$ should satisfy the equation

$$\sum_{i=1}^{2}\sum_{s=1}^{n}\xi_{is}^2 P_0|0\rangle = (\boldsymbol{\xi}\cdot\bar{\boldsymbol{\xi}})P_0|0\rangle = 0\,. \tag{4.37}$$

From the first two conditions we get

$$2n_1 + 2n_2 + 2n_3 = L - \omega\,, \tag{4.38a}$$

$$2n_1 - 2n_2 = L - \omega - 4q\,, \tag{4.38b}$$

so that, denoting $n_3 = 2r$, we get

$$P_0|0\rangle = \sum_r C_r(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{(L-\omega)/2-q-r}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q-r}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{2r}|0\rangle\,, \tag{4.39}$$

where, as the exponents must all be nonnegative integers, the r is restricted to integer values.

We have now to apply the operator $(\boldsymbol{\xi}\cdot\bar{\boldsymbol{\xi}})$ to (4.39) as indicated in condition (c). This is done in Appendix A, where it is shown that it leads to a recurrence relation for C_r, from which, up to a multiplicative constant, it can be evaluated as

$$C_r = \{r!(q-r)!\,[\tfrac{1}{2}(L-\omega) - q - r]! \times(-\tfrac{1}{2}[L-\omega] - \tfrac{1}{2}[n-3])_r\}^{-1}\,, \tag{4.40}$$

where $(x)_r$ is a Pochammer symbol $(x)_r = x(x+1)\cdots(x+r-1)$.

We now have to apply

$$(a_j)^{\omega-(L-M)/2+2q}(\bar{a}_j)^{(L-M)/2-2q} \tag{4.41}$$

to the $P_0|0\rangle$ (4.39) in which C_r is given by (4.40). This is also done in Appendix A where we show that, up to a constant, the state (4.34) can be written as

$$|LM\omega q\rangle = \sum_{\sigma,\tau}\{C_{\sigma\tau}^{LM\omega q}(\eta_j)^{\omega-l+\sigma-\tau}(\bar{\eta}_j)^{l+\tau-\sigma} \times(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{(L-\omega)/2-q-\sigma}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q-\tau} \times(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{\sigma+\tau}\}|0\rangle , \tag{4.42}$$

where

$$l=\tfrac{1}{2}(L-M)-2q , \tag{4.43}$$

and the coefficient in the double summation takes the form

$$C_{\sigma\tau}^{LM\omega q}=\frac{(-)^{\sigma+\tau}2^{\sigma+\tau}(L+n-2-\sigma-\tau)!}{[(L-\omega)/2-q-\sigma]!(q-\tau)!} \times\sum_{r}\left\{2^{2r}r!\left(-\frac{[L-\omega]}{2}-\frac{[n-3]}{2}\right)_r \times(\sigma-r)!\left[\frac{(L-M)}{2}-2q-\sigma+r\right]! \times(L-\omega+n-2-2r)!(\tau-r)! \times\left[\omega-\frac{(L-M)}{2}+2q-\tau+r\right]!\right\}^{-1} . \tag{4.44}$$

So far our state is given as a polynomial in the creation operators acting on the ground state. But, as it corresponds to the irrep L of O(2n), it can, by Dragt's theorem,[17] be transformed into a wave function of the coordinates, i.e.,

$$|LM\omega q\rangle = P_{LM\omega q}(\eta_m,\bar{\eta}_m)|0\rangle = \pi^{-(n/2)}2^{L/2}e^{-\rho^2/2}P_{LM\omega q}(x_m,\bar{x}_m) . \tag{4.45}$$

In (4.45) the irrep of U(2n) is the same as that of O(2n) so that a more correct notation for this state, characterized by the irreps in the chain of groups (4.30) and the multiplicity index q, would be $|LLM\omega q\rangle$. The most general state of this type would be $|NLM\omega q\rangle$ where $N=L,L+2,\ldots$. To get this state, once $|LM\omega q\rangle$ of (4.45) is known, is quite simple. We consider the H_0 of (1.3) in the ZD coordinates (1.5) in which $\rho_1=\rho\cos\gamma, \rho_2=\rho\sin\gamma$ and get

$$\frac{1}{2}\left[-\frac{1}{\rho^{2n-1}}\frac{\partial}{\partial\rho}\rho^{2n-1}\frac{\partial}{\partial\rho}+\frac{\mathscr{L}^2}{\rho^2}+\rho^2\right]|NLM\omega q\rangle = (N+n)|NLM\omega q\rangle , \tag{4.46}$$

where $\mathscr{L}^2$ is the Casimir operator of O(2n). We now write

$$|NLM\omega q\rangle = \rho^{-n+1/2}R(\rho)(e^{\rho^2/2}\rho^{-L}|LM\omega q\rangle) , \tag{4.47}$$

where the right-hand side ket is given by (4.45). As $\rho^{-L}\exp(\rho^2/2)|LM\omega q\rangle$ is independent of ρ and satisfies

$$\mathscr{L}^2[\rho^{-L}e^{\rho^2/2}|LM\omega q\rangle] = L(L+2n-2)[\rho^{-L}e^{\rho^2/2}|LM\omega q\rangle] , \tag{4.48}$$

we see that $R(\rho)$ satisfies the equation

$$\frac{1}{2}\left\{-\frac{d^2}{d\rho^2}+\frac{(L+n-1)^2-\frac{1}{4}}{\rho^2}+\rho^2\right\}R(\rho) = (2\nu+L+n)R(\rho) , \tag{4.49}$$

where we have written

$$N=2\nu+L . \tag{4.50}$$

The normalized $R(\rho)$, which we characterize by the indices ν,L, can be expressed as

$$R_{\nu L}(\rho)=[2(\nu!)/\Gamma(\nu+L+n)]^{1/2} \times\rho^{L+n-1/2}e^{-\rho^2/2}L_{\nu}^{L+n-1}(\rho^2) , \tag{4.51}$$

where L_{ν}^{L+n-1} is a Laguerre polynomial.

We can then using (4.45) and (4.42) write our state as

$$|NLM\omega q\rangle = \pi^{-(n/2)}2^{L/2}\rho^{-(n+L)+1/2}R_{\nu L}(\rho) \times\sum_{\sigma,\tau}\{C_{\sigma\tau}^{LM\omega q}(x_j)^{\omega-l+\sigma+\tau}(\bar{x}_j)^{l+\tau-\sigma} \times(\mathbf{x}\cdot\mathbf{x})^{(L-\omega)/2-q-\sigma}(\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})^{q-\tau}(\mathbf{x}\cdot\bar{\mathbf{x}})^{\sigma+\tau}\} \equiv\rho^{-n+1/2}R_{\nu L}(\rho)|LM\omega q\} , \tag{4.52}$$

where l is given by (4.43) and $C_{\sigma\tau}^{LM\omega q}$ by (4.44) and $|LM\omega q\}$ is independent of ρ.

We have thus obtained explicitly the eigenstates of H_0 characterized by the irreps of the group chain (4.30). As the irrep L of O(2n) determines the irrep λ of Sp(2) through (4.31), and in view of the relation (4.50) between N and ν, we can also denote the state (4.52) as characterized by the irreps of the chain (1.4), i.e.,

$$|[n/2,n/2+\omega]\lambda M,\nu q\rangle = |NLM\omega q\rangle . \tag{4.53}$$

Thus we have the most general state associated with this chain and in the next section we proceed to discuss the matrix elements of the generators of Sp(4) with respect to the kets (4.53).

5. THE GENERAL CASE: ALGORITHM FOR THE MATRIX ELEMENTS

The state (4.52) is a rather complex one, and thus we do not expect a simple expression for the matrix elements of r, $r\sin\theta\exp(\pm i\varphi)$ with respect to it. Rather than a closed formula, we shall give an algorithm that permits the calculation of all required matrix elements. This algorithm can be easily programmed and eventually we plan to do this both for the $d=2$ case discussed here and the $d=3$ that we intend to analyze in a future publication.

We shall start with the calculation of r which, from (1.9), is $r=(\rho^2/2)$. It is clear therefore that only matrix elements different from zero are

$$\langle N'LM\omega q'|(\rho^2/2)|NLM\omega q\rangle = \langle\nu'L\,|(\rho^2/2)|\nu L\rangle\{LM\omega q'|LM\omega q\} , \tag{5.1}$$

where the radial part

$$\langle\nu'L\,|(\rho^2/2)|\nu L\rangle = \int_0^\infty R_{\nu'L}(\rho)(\rho^2/2)R_{\nu L}(\rho)\,d\rho \tag{5.2}$$

is evaluated in Sec. 7 of I.

There remains the scalar product

$$\{LM\omega q'|LM\omega q\} = 2[(L+n)!]^{-1}\langle LM\omega q'|LM\omega q\rangle , \tag{5.3}$$

where the radial factors in the angular kets are $\rho^L\exp(-\rho^2/2)$, and we used the fact that

$$\int_0^\infty \rho^{2L}e^{-\rho^2}\rho^{2n-1}\,d\rho=\tfrac{1}{2}(L+n)! . \tag{5.4}$$

The angular kets $|LM\omega q\rangle$ are given by (4.45), and in Appendix B we calculate the scalar product on the right-hand side of (5.3) to get

$$\langle LM\omega q'|LM\omega q\rangle = \frac{2^{(L-\omega)/2}(L+n-2)!(\omega+M+2q'+n-2)!!(L-M-2q'+n-2)!!}{(L-\omega+n-2)![(L-M)/2-2q']!}$$
$$\times\sum_{\sigma\tau\rho}\frac{C^{LM\omega q}_{\sigma\tau}(\sigma+\tau)![\omega-(L-m)/2+\sigma-\tau]!}{2^{\rho}\rho!(q-q'+\sigma-\rho)![\omega-(L-M)/2+q+q'-\tau+\rho]![q'-q+\tau-\rho)!}$$
$$\times\frac{[(L-M)/2-q-q'+\tau-\rho]!}{(2\omega-L+M+2q+2q'+n-2+2\sigma)!!(L-M-2q-2q'+n-2+2\tau-2\rho)!!}\,. \tag{5.5}$$

We now turn our attention to the matrix elements of

$$r\sin\theta\, e^{i\varphi} = (\mathbf{x}\cdot\mathbf{x})\,, \tag{5.6a}$$
$$r\sin\theta\, e^{-i\varphi} = (\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})\,. \tag{5.6b}$$

We shall only sketch here the algorithm to obtain the matrix elements of (5.6a) as a similar one holds for (5.6b).

As we already know how to calculate the matrix element of r, we can restrict ourselves to that of

$$\sin\theta\, e^{i\varphi} = [(\mathbf{x}\cdot\mathbf{x})/(\mathbf{x}\cdot\bar{\mathbf{x}})]\,, \tag{5.7}$$

for which we can restrict ourselves to the states with $N=L$, i.e., the $|LM\omega q\rangle$ of (4.45). We thus get

$$[(\mathbf{x}\cdot\mathbf{x})/(\mathbf{x}\cdot\bar{\mathbf{x}})]\,|LM\omega q\rangle$$
$$= (\pi^{-n/2}2^{L/2}e^{-\rho^2/2})\sum_{\sigma,\tau}C^{LM\omega q}_{\sigma\tau}(x_j)^{\omega-l+\sigma-\tau}(\bar{x}_j)^{l-\sigma+\tau}$$
$$\times(\mathbf{x}\cdot\mathbf{x})^{(L-\omega)/2-q-\sigma+1}(\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})^{q-\tau}(\mathbf{x}\cdot\bar{\mathbf{x}})^{\sigma+\tau-1}\,, \tag{5.8}$$

where $l=[(L-M)/2]-2q$.

We now use the same expression (5.8) for a particular ket $|L'M'\omega q'\rangle$ multiplied by $(\mathbf{x}\cdot\bar{\mathbf{x}})^{s+t-1}$, i.e.,

$$(\mathbf{x}\cdot\bar{\mathbf{x}})^{s+t-1}|L+2-2s-2t,M+2,\omega,q-t\rangle$$
$$= (\pi^{-n/2}2^{L/2}e^{-\rho^2/2})\sum_{\sigma,\tau}C^{L+2-2s-2t,M+2,\omega,q}_{\sigma-s,\tau-t}$$
$$\times(x_j)^{\omega-\bar{l}+\sigma-\tau-s+t}(\bar{x}_j)^{\bar{l}-\sigma+\tau+s-t}$$
$$\times(\mathbf{x}\cdot\mathbf{x})^{(L+2-2s-2t-\omega)/2-q+t-\sigma+s}$$
$$\times(\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})^{q-\tau}(\mathbf{x}\cdot\bar{\mathbf{x}})^{\sigma+\tau-1}\,, \tag{5.9}$$

where

$$\bar{l} = \tfrac{1}{2}(L+2-2s-2t)-\tfrac{1}{2}(M+2)-2(q-t) \tag{5.10}$$

and instead of writing the summation indices as σ,τ in (5.8) we write them as $\sigma-s,\tau-t$.

Using (5.10) and simplifying terms, we finally have

$$(\mathbf{x}\cdot\bar{\mathbf{x}})^{s+t-1}|L+2-2s-2t,M+2,\omega,q-t\rangle$$
$$= (\pi^{-n/2}2^{L/2}e^{-\rho^2/2})\sum_{\sigma,\tau}C^{L+2-2s-2t,M+2,\omega,q-t}_{\sigma-s,\tau-t}$$
$$\times(x_j)^{\omega-l+\sigma-\tau}(\bar{x}_j)^{l-\sigma+\tau}$$
$$\times(\mathbf{x}\cdot\mathbf{x})^{(L-\omega)/2-q-\sigma+1}(\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})^{q-\tau}(\mathbf{x}\cdot\bar{\mathbf{x}})^{\sigma+\tau-1}\,. \tag{5.11}$$

Comparing (5.8) and (5.11), we see that in them all the powers of x_j, $\bar{x}_j$, $(\mathbf{x}\cdot\mathbf{x})$, $(\bar{\mathbf{x}}\cdot\bar{\mathbf{x}})$, $(\mathbf{x}\cdot\bar{\mathbf{x}})$ are the same though the coefficients C are different. This clearly indicates that we can write

$$[(\mathbf{x}\cdot\mathbf{x})/(\mathbf{x}\cdot\bar{\mathbf{x}})]\,|LM\omega q\rangle$$
$$= \sum_{s,t}A_{st}(\mathbf{x}\cdot\bar{\mathbf{x}})^{s+t-1}|L+2-2s-2t,M+2,\omega,q-t\rangle\,, \tag{5.12}$$

and, comparing coefficients in (5.8) and (5.11), we get

$$C^{LM\omega q}_{\sigma\tau} = \sum_{s,t}A_{st}C^{L+2-2s-2t,M+2,\omega,q-t}_{\sigma-s,\tau-t}\,, \tag{5.13}$$

and thus we have an inhomogeneous system of linear algebraic equations to determine the A_{st}.

Because in (5.11) the indices $\sigma-s,\tau-t$ of the C's must be nonnegative, we have that if we make a choice of σ,τ, the s,t are restricted by $0\leqslant s\leqslant\sigma$, $0\leqslant t\leqslant\tau$. Now if, for example, we take $\sigma=0$, then $s=0$, and we get the system of equations

$$C^{LM\omega q}_{0\tau} = \sum_{t=0}^{\tau}A_{0t}C^{L+2-2t,M+2,\omega,q-t}_{0,\tau-t}\,. \tag{5.14}$$

The system of equations (5.14) in the unknowns A_{0t} has a triangular determinant so it is immediate to solve giving us A_{0t} for $t=0,1,\ldots,\tau_{\max}$. Once we have these A_{0t}, we can pass in (5.13) to $s=1$ where the only unknowns are A_{1t}, and again we have a system of linear equations also with a triangular determinant so we can obtain the A_{1t} immediately. We then pass to $s=2$ and so on up to $s=\sigma_{max}$ to get all the A_{st} as illustrated graphically in Fig. 1.

There remains the problem of finding $(\mathbf{x}\cdot\bar{\mathbf{x}})^{u}|LLM\omega q\rangle$ in terms of $|NLM\omega q\rangle$, but, as $(\mathbf{x}\cdot\bar{\mathbf{x}})=(\rho^2/2)$, this is achieved trivially with the help of the discussion given at the beginning of this section. Of course, eventually we want the matrix elements not with respect to the $|NLM\omega q\rangle$ but with respect to combinations of them in the q index that will be orthonormal. But this can be achieved in a standard way[20] with the help of the matrix elements (5.5).

Thus we obtained an algorithm for the determination of the matrix elements of $\sin\theta\exp(i\varphi)$, and, in a similar fashion, we can obtain those of $\sin\theta\exp(-i\varphi)$.

Because the generators of Sp(4) are given by the q_{ij} of

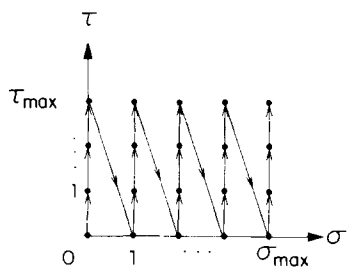


FIG. 1. The figure illustrates the successive steps for the determination of the coefficients A_{st} in (5.13). Each dot stands for an A_{st}. Setting $\sigma=0$ in (5.13) we obtain (5.14) from which all A_{0t} can be found, then set $\sigma=1$ and determine all A_{1t}; continue increasing σ up to $\sigma=\sigma_{max}$ which allows us to find the last unknowns $A_{\sigma_{max},t}$.

(1.10) and commutators of q_{ij} with H_0 as shown in (8I), it is clear that for $d = 2$ we are in a position to give all the matrix elements of the generators of Sp(4) with respect to the states (4.53) that are characterized by the irreps of the chain of groups (1.4).

6. CONCLUSION

In the previous sections we obtained in (4.52) and (4.53) the eigenstates of H_0 characterized by the chains of groups (1.4) and the matrix elements of q_{ij} of (1.2) with respect to them. We want first to indicate the information we can obtain from the latter on the shape of collective states for two-dimensional many-body systems and then outline the procedure for generalizing our results to three dimensions.

As the $q_{ij}, i,j = 1,2$, of (1.2) give the components of the quadrupole tensor in two-dimensional space, it is clear from (1.7) that ρ_1^2, ρ_2^2 give the components of the quadrupole tensor in the frame of reference fixed in the body.

From the discussion in I, the collective Hamiltonian H is in the enveloping algebra of Sp(4), and thus its eigenstates can be expressed as linear combinations of the kets (4.53) in which n,ω—and thus the irrep $[(n/2),(n/2)+\omega]$ of Sp(4)—are kept fixed and also M as the angular momentum is an integral of motion. The eigenstates of H can then be expressed as

$$|n\omega M) = \sum_{\lambda\nu q} A_{\lambda\nu q} \left| \left[\frac{n}{2}, \frac{n}{2} + \omega \right] \lambda M, \nu q \right\rangle . \tag{6.1}$$

To get information about the shape of these states, we would require the expectation values of ρ_1^2, ρ_2^2 with respect to the ket (6.1) and thus the matrix elements of ρ_1^2, ρ_2^2 with respect to the states (4.53). Alternatively as ρ_1^2, ρ_2^2 are roots of the characteristic equation

$$x^2 - (\rho_1^2 + \rho_2^2)x + \rho_1^2 \rho_2^2 = 0 , \tag{6.2}$$

we can calculate the matrix elements of

$$\rho_1^2 + \rho_2^2 = \rho^2 = 2r , \quad \rho_1^2 \rho_2^2 = \tfrac{1}{4}\rho^2 \sin^2 2\gamma = r^2 \cos^2\theta \tag{6.3}$$

and then solve Eq. (6.2) when we replace $2r, r^2\cos^2\theta$ by their expectation values with respect to the collective states (6.1).

To implement this program, we require the matrix elements of

$$r = (\mathbf{x} \cdot \bar{\mathbf{x}}) , \tag{6.4a}$$

$$r^2 \sin^2\theta = (\mathbf{x} \cdot \mathbf{x})(\bar{\mathbf{x}} \cdot \bar{\mathbf{x}}) \tag{6.4b}$$

with respect to the states (4.53). The one of (6.4a) was explicitly given in (5.1), while as an algorithm was obtained in Sec. 5 for $(\mathbf{x} \cdot \mathbf{x})$ and $(\bar{\mathbf{x}} \cdot \bar{\mathbf{x}})$, we can obtain from them one for (6.4b).

We do not give here explicit expressions for the expectation values of $r, r^2\cos^2\theta$ with respect to the states (6.1) as we prefer to discuss this problem in detail in another publication dealing with shape for systems of A particles both in two- and three-dimensional space, where, of course, the latter is the physically relevant case.

We want also to indicate briefly what will be the procedure for generalizing our analysis to $d = 3$. In this case we have the position coordinate x_{is}, $i = 1,2,3$, $s = 1,2,...,n$, which we can write in "hyperspherical" covariant form as

$$x_{qm} , \quad q = +,0,- , \quad m = j, j-1, ..., -j , \quad n = 2j+1 , \tag{6.5}$$

which are defined in a way similar to the analysis given between (1.13) and (1.17). Besides we have also the covariant creation operators η_{qm}, and the contravariant momenta and annihilation operators p^{qm}, ξ^{qm}, where

$$x_{qm} = (-)^q x_m^{-q} = (-)^q x^{-q,-m} . \tag{6.6}$$

What we are looking for are eigenstates of

$$H_0 = \frac{1}{2} \sum_{i=1}^{3} \sum_{s=1}^{n} (p_{is}^2 + x_{is}^2) \tag{6.7}$$

characterized by the chain of groups

$$\begin{array}{lcccc} \mathrm{Sp}(6n) & \supset & \mathrm{Sp}(6) & \times & \underset{(\omega_1\omega_2\omega_3)}{\mathrm{O}(n)} , \\ & & \overset{\lambda}{} & & \overset{\nu}{} \\ & \supset & \mathrm{Sp}(2) & \supset & \mathrm{O}(2) \\ \overset{[n/2+\omega_3, n/2+\omega_2, n/2+\omega_1]}{\mathrm{Sp}(6)} & & \times & & \end{array} \tag{6.8a}$$

$$\underset{L}{\mathcal{O}(3)} \supset \underset{M}{\mathcal{O}(2)} , \tag{6.8b}$$

where above and below the groups we give their irreps.

We also require the matrix elements of the operator

$$Q_q^{q'} = \sum_{m=-j}^{j} x_{qm} x^{q'm} \tag{6.9}$$

with respect to the states characterized by the chain of groups (6.8).

It is clear that much of the analysis carried out in Secs. 4 and 5 can be generalized to the case $d = 3$, though the algebraic work is likely to be much harder. From the discussion given in I we see that the determination of the basis for the irreducible representations of Sp(6) in the chain (6.8b) will be central to the discussion of collective behavior in many-body systems.

APPENDIX A: DERIVATION OF THE COEFFICIENTS C_r AND $C_{\sigma\tau}^{LM\omega q}$

In the first part of this appendix we discuss the derivation of the coefficients C_r appearing in the state $P_0|0\rangle$ of (4.39) which satisfies the condition

$$(\xi \cdot \bar{\xi}) P_0 |0\rangle = 0 . \tag{A1}$$

From (1.20) we see that essentially it is a matter of evaluating the expression

$$\sum_{m=j}^{-j} \frac{\partial^2}{\partial\bar{\eta}_{-m}\partial\eta_m} M_{n_1,n_2,2r} , \tag{A2}$$

where

$$M_{n_1,n_2,2r} \equiv (\eta \cdot \eta)^{n_1} (\bar{\eta} \cdot \bar{\eta})^{n_2} (\eta \cdot \bar{\eta})^{2r} \tag{A3}$$

and the scalar products are defined in Eqs. (4.8).

It is immediately deduced that

$$\sum_m \frac{\partial^2}{\partial\bar\eta_{-m}\partial\eta_m} M_{n_1,n_2,2r}$$

$$= \sum_m \frac{\partial}{\partial\bar\eta_{-m}} \{ 2n_1\eta_{-m} M_{n_1-1,n_2,2r} + 2r\bar\eta_{-m} M_{n_1,n_2,2r-1} \}$$

$$= \sum_m \{ 2n_1\cdot 2n_2\eta_{-m}\bar\eta_m M_{n_1-1,n_2-1,2r} + 2n_1\cdot 2r\eta_{-m}\eta_m M_{n_1-1,n_2,2r-1} + 2r\delta_{mm} M_{n_1,n_2,2r-1} + 2r\cdot 2n_2\bar\eta_{-m}\bar\eta_m M_{n_1,n_2-1,2r-1} + 2r(2r-1)\bar\eta_{-m}\eta_m M_{n_1,n_2,2r-2} \}$$

$$= 4n_1n_2 M_{n_1-1,n_2-1,2r+1} + 2r(2n_1+2n_2+2r-1+n) M_{n_1,n_2,2r-1} . \quad \text{(A4)}$$

Introducing the values

$$n_1 = \tfrac{1}{2}(L-\omega) - q - r , \quad n_2 = q - r \quad \text{(A5)}$$

appropriate for $P_0|0\rangle$ of (4.39) and making a shift of dummy index $r \to r-1$ in the term containing $(\eta\cdot\bar\eta)^{2r+1}$, we obtain

$$(\xi\cdot\bar\xi)P_0|0\rangle = 4\sum_r \left\{ \left(\frac{L-\omega}{2} - q + 1 - r\right)(q+1-r)C_{r-1} + r\left(\frac{L-\omega}{2} + \frac{n-1}{2} - r\right)C_r \right\} \times (\eta\cdot\eta)^{(L-\omega)/2-q-r}(\bar\eta\cdot\bar\eta)^{q-r}(\eta\cdot\bar\eta)^{2r-1}|0\rangle . \quad \text{(A6)}$$

Thus Eq. (A1) will be satisfied provided we choose the coefficients C_r such that

$$\frac{C_r}{C_{r-1}} = \frac{[(L-\omega)/2 - q + 1 - r](q+1-r)}{r[-(L-\omega)/2 - (n-1)/2 + r]} . \quad \text{(A7)}$$

The solution to this recursion relation is the C_r given in (4.40).

We turn now to the application of powers of the traceless boson operators $a_j,\bar a_j$ of (4.32) on the state $P_0|0\rangle$ of (4.39). Again, from (1.20) we see that this is equivalent to the application on the monomial $M_{n_1,n_2,2r}$ of (A3) of powers of the differential operators

$$a_j = \eta_j - 2(\eta\cdot\bar\eta)(2\hat N + 2n)^{-1}\frac{\partial}{\partial\bar\eta_{-j}} , \quad \text{(A8a)}$$

$$\bar a_j = \bar\eta_j - 2(\eta\cdot\bar\eta)(2\hat N + 2n)^{-1}\frac{\partial}{\partial\eta_{-j}} . \quad \text{(A8b)}$$

Applying first the $\bar a_j$, we find

$$(\bar a_j)^l M_{n_1,n_2,2r} = \frac{(2n_1+2n_2+4r+n-2)!n_1!}{(2n_1+2n_2+4r+n-2+l)!(2n_1+2n_2+2r+n-2)!} \times \sum_s \frac{(-2)^s l!(2n_1+2n_2+2r+n-2+l-s)!}{s!(l-s)!(n_1-s)!}(\eta_j)^s(\bar\eta_j)^{l-s} M_{n_1-s,n_2,2r+s} , \quad \text{(A9)}$$

a result which has been checked by the method of induction. By a similar procedure, applying the a_j on the previous polynomial, we find

$$(a_j)^{l'}(\eta_j)^s(\bar\eta_j)^{l-s} M_{n_1-s,n_2,2r+s} = \frac{(2n_1+2n_2+4r+l+n-2)!n_2!}{(2n_1+2n_2+4r+l+l'+n-2)!(2n_1+2n_2+2r-s+l+n-2)!} \times \sum_t \frac{(-2)^t l'!(2n_1+2n_2+2r-s+n-2+l+l'-t)!}{t!(l'-t)!(n_2-t)!}(\eta_j)^{l'+s-t}(\bar\eta_j)^{l+t-s} M_{n_1-s,n_2-t,2r+s+t} . \quad \text{(A10)}$$

The general state

$$|LM\omega q\rangle = (a_j)^{l'}(\bar a_j)^l P_0|0\rangle \quad \text{(A11)}$$

can be obtained by combining (A9), (A10) with (4.39), introducing as new summation indices $\sigma\equiv r+s$, $\tau\equiv r+t$ instead of s,t, and giving to n_1,n_2 the values in (A5), to l the value in (4.43), and setting $l' = \omega - l$. The result appears in Eqs. (4.42) and (4.44).

APPENDIX B: EVALUATION OF THE SCALAR PRODUCT (5.5)

The state $|LM\omega q\rangle$ of (4.42) was shown in Sec. 4 to belong to a basis for the irrep (L) of both U($2n$) and O($2n$); thus it satisfies the condition

$$(\xi\cdot\bar\xi)|LM\omega q\rangle = 0 , \quad \text{i.e.,} \quad (\eta\cdot\bar\eta)^\dagger|LM\omega q\rangle = 0 , \quad \text{(B1)}$$

where $\eta_m^\dagger$ means the Hermitian conjugate of η_m, and similarly for $\bar\eta_m^\dagger$. Therefore, when taking the scalar product of $|LM\omega q\rangle$ with $|LM\omega q'\rangle$, in the explicit realization of the latter state, Eq. (4.42), only the term in the double sum corresponding to $\sigma=\tau=0$ can give a nonvanishing contribution. As the scalar product is real, we can write

$$\langle LM\omega q'|LM\omega q\rangle = C_{00}^{LM\omega q'}\sum_{\sigma\tau} C_{\sigma\tau}^{LM\omega q} \times \langle 0|[(\eta_j)^{\omega-l+\sigma-\tau}(\bar\eta_j)^{l+\tau-\sigma}(\eta\cdot\eta)^{(L-\omega)/2-q-\sigma} \times (\bar\eta\cdot\bar\eta)^{q-\tau}(\eta\cdot\bar\eta)^{\sigma+\tau}]^\dagger(\eta_j)^{\omega-l'} \times (\bar\eta_j)^{l'}(\eta\cdot\eta)^{(L-\omega)/2-q'}(\bar\eta\cdot\bar\eta)^{q'}|0\rangle , \quad \text{(B2)}$$

where

$$l = (L-M)/2 - 2q , \quad l' = (L-M)/2 - 2q' \quad \text{(B3)}$$

and the C's are given in (4.44).

We now get rid of the powers of $(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{\dagger}$ in (B3) by using the fact that, from (1.20), $(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{\dagger}$ acting on $P(\eta)|0\rangle$ is equivalent to $\Sigma_m(\partial^2 P/\partial\eta_m\partial\eta_{-m})|0\rangle$, and

$$\left[\sum_m \frac{\partial^2}{\partial\eta_m\partial\eta_{-m}}\right]^{(L-\omega)/2-q-\sigma}(\eta_j)^{\omega-l'}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{(L-\omega)/2-q'} = A_{\sigma\tau}^{LM\omega qq'}(\eta_j)^{\omega-l'}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{q-q'+\sigma} \tag{B4}$$

with

$$A_{\sigma}^{LM\omega qq'} = \frac{2^{(L-\omega)/2-q-\sigma}[(L-\omega)/2-q']!(L+\omega-2l'-2q'+n-2)!!}{(q-q'+\sigma)!(2\omega-2l'-2q'+n-2+2\sigma+2q)!!}. \tag{B5}$$

Thus, using again the reality of the scalar product, we have

$$\langle LM\omega q'|LM\omega q\rangle = C_{00}^{LM\omega q'}\sum_{\sigma\tau} C_{\sigma\tau}^{LM\omega q}A_{\sigma}^{LM\omega qq'}\langle 0|\left[(\eta_j)^{\omega-l'}(\bar{\eta}_j)^{l'}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{q-q'+\sigma}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q'}\right]^{\dagger}$$
$$\times(\eta_j)^{\omega-l+\sigma-\tau}(\bar{\eta}_j)^{l+\tau-\sigma}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q-\tau}(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{\sigma+\tau}|0\rangle. \tag{B6}$$

By a similar procedure, we can now get rid of the powers of $(\boldsymbol{\eta}\cdot\boldsymbol{\eta})^{\dagger}$ in (B6) using the fact that

$$\left[\sum_m \frac{\partial^2}{\partial\eta_m\partial\eta_{-m}}\right]^{q-q'+\sigma}(\eta_j)^{\omega-l+\sigma-\tau}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{\sigma+\tau}$$
$$= \sum_{\rho} B_{\sigma\tau\rho}^{LM\omega qq'}(\eta_j)^{\omega-l+q'-q-\tau+\rho}(\bar{\eta}_j)^{q-q'+\sigma-\rho}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{\rho}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{q'-q+\tau-\rho} \tag{B7}$$

with

$$B_{\sigma\tau\rho}^{LM\omega qq'} = \frac{(\omega-l+\sigma-\tau)!(\sigma+\tau)!(q-q'+\sigma)!2^{q-q'+\sigma-\rho}}{\rho!(q-q'+\sigma-\rho)!(\omega-l-\tau-q+q'+\rho)!(q'-q+\tau-\rho)!}, \tag{B8}$$

a result which can be checked by the method of induction. Therefore, we have, at this stage,

$$\langle LM\omega q'|LM\omega q\rangle$$
$$= C_{00}^{LM\omega q'}\sum_{\sigma\tau\rho} C_{\sigma\tau}^{LM\omega q}A_{\sigma}^{LM\omega qq'}B_{\sigma\tau\rho}^{LM\omega qq'}$$
$$\times\langle 0|\left[(\eta_j)^{\omega-l+q'-q-\tau+\rho}(\bar{\eta}_j)^{q-q'+l+\tau-\rho}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q-\tau+\rho}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{q'-q+\tau-\rho}\right]^{\dagger}(\eta_j)^{\omega-l'}(\bar{\eta}_j)^{l'}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q'}|0\rangle, \tag{B9}$$

where again we have utilized the fact that the scalar product is real.

By a formula similar to (B4) but with $\bar{\eta}_m$ replacing to η_m, the scalar product on the right-hand side of (B9) is seen to be equal to

$$D_{\tau\rho}^{LMqq'}\langle 0|\left[(\eta_j)^{\omega-l+q'-q-\tau+\rho}(\bar{\eta}_j)^{q-q'+l+\tau-\rho}(\boldsymbol{\eta}\cdot\bar{\boldsymbol{\eta}})^{q'-q+\tau-\rho}\right]^{\dagger}(\eta_j)^{\omega-l'}(\bar{\eta}_j)^{l'}(\bar{\boldsymbol{\eta}}\cdot\bar{\boldsymbol{\eta}})^{q'-q+\tau-\rho}|0\rangle \tag{B10}$$

with

$$D_{\tau\rho}^{LMqq'} = \frac{q'!(2q'+2l'+n-2)!!2^{q-\tau+\rho}}{(q'-q+\tau-\rho)!(2q'+2l'+n-2-2q+2\tau-2\rho)!!}, \tag{B11}$$

and the scalar product in (B10) can easily be found to have the value

$$E_{\tau\rho}^{LM\omega qq'} = (\omega-l')!(q-q'+l+\tau-\rho)!(q'-q+\tau-\rho)!2^{q'-q+\tau-\rho}. \tag{B12}$$

Thus our final result is

$$\langle LM\omega q'|LM\omega q\rangle = C_{00}^{LM\omega q'}\sum_{\rho\sigma\tau} C_{\sigma\tau}^{LM\omega q}A_{\sigma}^{LM\omega qq'}B_{\sigma\tau\rho}^{LM\omega qq'}D_{\tau\rho}^{LMqq'}E_{\tau\rho}^{LM\omega qq'}. \tag{B13}$$

Inserting the values of A,B,D,E, we obtain the expression given in (5.5).

[1] M. Moshinsky, J. Math. Phys. **25**, 1555 (1984).
[2] W. Zickendraht, J. Math. Phys. **10**, 30 (1969); **12**, 1663 (1971).
[3] A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[4] A. J. McConnell, *Applications of the Absolute Differential Calculus* (Blackie, London and Glasgow, 1942), Chap. XI.
[5] M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968).
[6] E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981).
[7] L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics* (McGraw-Hill, New York, 1935), pp. 275–8.
[8] E. M. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957), pp. 51–3.
[9] M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
[10] I. S. Gradsteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), 4th ed., pp. 844, 1037–9.
[11] O Castaños, E. Chacón, A. Frank, P. O. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[12] E. Chacón and M. Moshinsky, Kinam **3**, 3 (1981).
[13] I. S. Gradsteyn and I. M. Ryzhik, Ref. 10, pp. 1035–6.
[14] M. A. Lohe, "The Development of the Boson Calculus for the Orthogonal and Symplectic Groups," thesis, University of Adelaide, 1974.
[15] E. Chacón, M. Moshinsky, and R. T. Sharp, J. Math. Phys. **17**, 668 (1976).
[16] M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[17] A. J. Dragt, J. Math. Phys. **6**, 533 (1965).
[18] O. Castaños, E. Chacón, and A. Frank, J. Math. Phys. (to be published).
[19] E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).
[20] O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).

On the analytic structure of the wave function for a hydrogen atom in an analytic potential

Robert Nyden Hill
Physics Department, University of Delaware, Newark, Delaware 19711

(Received 2 August 1983; accepted for publication 7 October 1983)

The rate of convergence of an approximate method for solving Schrödinger's equation depends on the ability of the approximating sequence to mimic the analytic structure of the unknown exact wave function. Thus a knowledge of the analytic structure of the wave function can be of great value when approximation schemes are designed. Consider the Schrödinger equation $[-\frac{1}{2}\nabla^2 - r^{-1} + V(\mathbf{r})]\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$ for a hydrogen atom in a potential $V(\mathbf{r})$. The general theory of elliptic partial differential equations implies that Ψ is analytic at regular points, but no general theory is available at singular points. The present paper investigates the Coulomb singular point at $r = 0$ and shows that, if $V(\mathbf{r}) = V_1(x,y,z) + rV_2(x,y,z)$ where V_1 and V_2 are analytic functions of x, y, z at $x = y = z = 0$, then the wave function has the form $\Psi(\mathbf{r}) = \Psi_1(x,y,z) + r\Psi_2(x,y,z)$ where Ψ_1 and Ψ_2 are analytic functions of x, y, z at $x = y = z = 0$.

PACS numbers: 31.15. + q, 02.30. + g, 03.65.Ge

I. INTRODUCTION

In most cases of physical interest, the wave function Ψ which satisfies the Schrödinger equation $H\Psi = E\Psi$ cannot be determined exactly. When only approximations are available, rapid convergence of those approximations is clearly desirable. Because the rate of convergence of an approximation method usually depends on the ability of the approximating sequence to mimic the analytic structure of an unknown exact Ψ, information about this analytic structure can be of great value when approximation methods are designed. Thus the desire for better methods of practical numerical computation leads in a natural way to the study of the analytic structure of Ψ.

The mathematical foundation for the study of N-particle quantum-mechanical systems was laid in Kato's famous 1951 paper,[1] which gave the first rigorous definition of the Hamiltonian, proved that it is essentially self-adjoint (Hermitian), proved that eigenfunctions belonging to whatever discrete eigenvalues the Hamiltonian has will satisfy the Schrödinger equation as a differential equation except at singular points of the potential, and proved that the discrete eigenvalues (below the continuum) and their eigenfunctions could be characterized by the familiar Rayleigh–Ritz variational principle. Kato's subsequent 1957 paper,[2] which was devoted to properties of the eigenfunctions, proved that the eigenfunctions are continuous throughout the configuration space, proved that they have partial derivatives of first order (except at Coulomb singular points) which are bounded, and established his famous cusp condition. A number of subsequent papers by other authors have extended and generalized his cusp condition.[3]

A general theory of the analytic structure of solutions to partial differential equations at singular points, analagous to the theory for ordinary differential equations at singular points, does not exist. The few results which are available in the literature[4] apply only to special cases. The present paper treats a special case of physical interest, namely, the Schrödinger equation $H\psi = E\psi$, where

$$H = -\tfrac{1}{2}\nabla^2 - r^{-1} + V(\mathbf{r}). \tag{1.1}$$

The Hamiltonian H describes a hydrogenic electron in a potential $V(\mathbf{r})$. Potentials of the form

$$V(\mathbf{r}) = V_1(x,y,z) + rV_2(x,y,z) \tag{1.2}$$

with V_1 and V_2 analytic functions of x, y, z at $x = y = z = 0$ are treated. The main result is that the wave function ψ has the form

$$\Psi(\mathbf{r}) = \Psi_1(x,y,z) + r\Psi_2(x,y,z) \tag{1.3}$$

in a neighborhood of $r = 0$, where ψ_1 and ψ_2 are analytic functions of x, y, z at $x = y = z = 0$.

This paper is the first in a projected series of papers which will discuss the analytic structure of the wave function at Coulomb singular points for arbitrary atoms and molecules. Because of certain additional technical difficulties which arise when discussing Coulomb singular points for many-particle systems, it was decided to treat only the simplest case in this first paper of the series. Nevertheless the results obtained apply to both the Zeeman and Stark effects for one-electron atoms, which are problems of continuing interest.

The analysis, which is quite elementary, begins by observing that the analytic structure of ψ is a local property. It is assumed that global questions have already been answered by invoking functional analysis to prove the existence of a "weak" or "generalized" global ψ in an appropriate Sobolev space. If P is a regular point of the potential $-r^{-1} + V(\mathbf{r})$, the Hamiltonian operator H is elliptic and ψ is an analytic function of the Cartesian coordinates at P. If P is a singular point of $-r^{-1} + V(\mathbf{r})$, choose a neighborhood N of P, and consider the interior Dirichlet problem $H\tilde{\psi} = E\tilde{\psi}$ in N, $\tilde{\psi} = \psi$ on ∂N. For N sufficiently small, this Dirichlet problem can be shown to have the unique solution $\tilde{\psi} = \psi$ in N. The analytic structure of ψ at P is then established by explicit construction of a series solution for $\tilde{\psi}$ which is general enough to satisfy $\tilde{\psi} = \psi$ on ∂N; by uniqueness this series coincides with ψ throughout N. The approach outlined

 0022-2488/84/051577-07$02.50

above works because the differential operator $-\nabla^2$ dominates, with the potential $-r^{-1}+V(\mathbf{r})$ acting as a small perturbation, in the neighborhood N. This approach should be useful for showing that formal series solutions of partial differential equations are sufficiently general to include the solutions of physical interest in other problems as well.

Section II of this paper gives precise statements of the results obtained for solutions of $H\psi = E\psi$. The proofs are in Sec. III.

II. STATEMENT OF RESULTS

First some notation. The setting is the (local) Sobolev space $H^1(\Omega)$ of functions for which the norm

$$\|\Psi\|_1 = \left[\int_\Omega (|\Psi|^2 + |\nabla\Psi|^2) d^3\mathbf{r}\right]^{1/2} \tag{2.1}$$

is finite. Ω_a is the subspace $0 \leqslant r < a$ of $\mathbb{R}^3$. The spherical harmonics $Y_{l,m}$ are defined in the usual way by

$$Y_{l,m}(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}\, P_l^m(\cos\theta) e^{im\phi}, \tag{2.2}$$

where

$$P_l^m(\xi) = \frac{(-1)^m}{2^l l!}(1-\xi^2)^{m/2}\left[\frac{d^{l+m}}{d\xi^{l+m}}(\xi^2-1)^l\right] \tag{2.3}$$

is an associated Legendre function as defined in Magnus, Oberhettinger, and Soni[5] and in the Bateman project.[6]

Theorem 1 and Corollary 1 give some properties of spherical harmonic expansions of analytic functions. These properties have been stated as a separate theorem because of their value for the interpretation of numerical data on rates of convergence. Theorem 2 and Corollary 2 spell out the analytic structure of solutions for the Hamiltonian (1.1) in a neighborhood of $r = 0$.

Theorem 1: Let r_1, r_2, δ, and δ_2 be real numbers satisfying $0 \leqslant r_1 < r_2 < \infty$, $\delta > 0$, $\delta_2 > 0$. Define the domains D, D_1, D_2, and D_3 by

$$\begin{aligned} D(r_1, r_2, \delta) := \{x, y, z \,|\, x\in\mathbb{C}, y\in\mathbb{C}, z\in\mathbb{C}, r_1^2 \leqslant [\mathrm{Re}(x)]^2 \\ + [\mathrm{Re}(y)]^2 + [\mathrm{Re}(z)]^2 \leqslant r_2^2, [\mathrm{Im}(x)]^2 \\ + [\mathrm{Im}(y)]^2 + [\mathrm{Im}(z)]^2 \leqslant \delta^2\}, \end{aligned} \tag{2.4}$$

$$\begin{aligned} D_1(r_1, r_2, \delta) := \{r, \theta, \phi \,|\, r\in\mathbb{C}, \theta\in\mathbb{C}, \phi\in\mathbb{C}, \\ r_1^2 \leqslant [\mathrm{Re}(r)]^2 + |r|^2 I(\theta,\phi) \leqslant r_2^2, [\mathrm{Im}(r)]^2 + |r|^2 I(\theta,\phi) \leqslant \delta^2, I(\theta,\phi) \\ = \sinh^2[\mathrm{Im}(\theta)] + \{\sin^2[\mathrm{Re}(\theta)] + \sinh^2[\mathrm{Im}(\theta)]\}\sinh^2[\mathrm{Im}(\phi)]\}, \end{aligned} \tag{2.5}$$

$$\begin{aligned} D_2(r_1, r_2, \delta, \delta_2) := \{r, \theta, \phi \,|\, r\in\mathbb{C}, \theta\in\mathbb{C}, \phi\in\mathbb{C}, \\ r_1^2 \leqslant [\mathrm{Re}(r)]^2, [\mathrm{Re}(r)]^2 + |r|^2 I_2(\theta,\phi) \leqslant r_2^2, \\ [\mathrm{Im}(r)]^2 + |r|^2 I_2(\theta,\phi) \leqslant \delta^2, I_2(\theta,\phi) \\ = \sinh^2[|\mathrm{Im}(\theta)| + |\mathrm{Im}(\phi)| + \delta_2]\}, \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} D_3(r_1, r_2, \delta, \delta_3) := \{r \,|\, r\in\mathbb{C}, r_1 \leqslant |\mathrm{Re}(r)|, \\ [\mathrm{Re}(r)]^2 + |r|^2 \sinh^2(\delta_3) \leqslant r_2^2, \\ [\mathrm{Im}(r)]^2 + |r|^2 \sinh^2(\delta_3) \leqslant \delta^2\}. \end{aligned} \tag{2.7}$$

D_1 is the image of D under the transformation from Cartesian to spherical coordinates.

Let $f(x,y,z)$ be a function of the three complex variables x, y, z, analytic and bounded in absolute value by a real constant M, for x, y, z in the tubular domain D. Define F by

$$F(r,\theta,\phi) := f(r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta). \tag{2.8}$$

Then F is an analytic function of r, θ, ϕ in the domain D_1, and F has an expansion

$$F(r,\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} r^l G_{l,m}(r) Y_{l,m}(\theta,\phi) \tag{2.9}$$

which converges uniformly to F in any domain $D_2(r_1, r_2, \delta, \delta_2)$ for which $\delta_2 > 0$ has been chosen small enough to make D_2 nonempty. The coefficient $r^l G_{l,m}(r)$ is given by

$$r^l G_{l,m}(r) = \int_0^\pi \left[\int_0^{2\pi} F(r,\theta,\phi) Y_{l,m}(\theta,\phi) d\phi\right] \sin\theta\, d\theta. \tag{2.10}$$

$G_{l,m}(r)$ is an analytic function of r for $r_1 \leqslant |\mathrm{Re}(r)| \leqslant r_2$, $|\mathrm{Im}(r)| \leqslant \delta$. The terms of the expansion (2.9) have the bound

$$|r^l G_{l,m}(r) Y_{l,m}(\theta,\phi)| \leqslant \frac{M\Gamma(l+\frac{3}{2})}{2\pi^{3/2} l!} \exp[-(l+1)\delta_2] \tag{2.11}$$

for (r,θ,ϕ) in $D_2(r_1, r_2, \delta, \delta_2)$. The expansion coefficients have the bound

$$|r^l G_{l,m}(r)| \leqslant \frac{M\Gamma(l+\frac{3}{2})}{2\pi^{3/2} l!} \exp[-(l+1)\delta_3] \tag{2.12}$$

for r in D_3 and δ_3 chosen small enough to make D_3 nonempty.

Corollary 1: Let R and δ_2 be real numbers satisfying $R > 0$, $\delta_2 > 0$. Define the domains D', D_1', D_2', and D_3' by

$$\begin{aligned} D'(R) := \{x, y, z \,|\, x\in\mathbb{C}, y\in\mathbb{C}, z\in\mathbb{C}, \\ |x|^2 + |y|^2 + |z|^2 \leqslant R^2\}, \end{aligned} \tag{2.13}$$

$$\begin{aligned} D_1'(R) := \{r, \theta, \phi \,|\, r\in\mathbb{C}, \theta\in\mathbb{C}, \phi\in\mathbb{C}, \\ |r|^2[1 + \sinh^2[\mathrm{Im}(\theta)] + \{\sin^2[\mathrm{Re}(\theta)] \\ + \sinh^2[\mathrm{Im}(\theta)]\}\sinh^2[\mathrm{Im}(\phi)]]\}, \end{aligned} \tag{2.14}$$

$$\begin{aligned} D_2'(R, \delta_2) := \{r, \theta, \phi \,|\, r\in\mathbb{C}, \theta\in\mathbb{C}, \phi\in\mathbb{C}, \\ |r|^2[1 + 2\sinh^2(|\mathrm{Im}(\theta)| + |\mathrm{Im}(\phi)| + \delta_2)] \leqslant R^2, \end{aligned} \tag{2.15}$$

and

$$D_3'(R, \delta_3) := \{r \,|\, r\in\mathbb{C}, |r|^2[1 + 2\sinh^2(\delta_3)] \leqslant R^2\}. \tag{2.16}$$

D_1' is the image of D' under the transformation from Cartesian to spherical coordinates. Then if the tube D of Theorem

1 is replaced by the ball D', the conclusions of Theorem 1 continue to hold with the domains D_1, D_2, and D_3 replaced by D_1', D_2', and D_3', respectively, and with the domain of analyticity of $G_{l,m}$ replaced by $0 \leqslant |r| \leqslant R$.

Theorem 2: Let Ψ be any solution to $H\Psi = E\Psi$, with H given by (1.1), whose restriction to Ω_R lies in $H^1(\Omega_R)$. Let V_1 and V_2 in (1.2) be analytic functions of the Cartesian coordinates x, y, z for complex x, y, z in the tube $0 \leqslant [\mathrm{Re}(x)]^2 + [\mathrm{Re}(y)]^2 + [\mathrm{Re}(z)]^2 \leqslant R^2$, $[\mathrm{Im}(x)]^2 + [\mathrm{Im}(y)]^2 + [\mathrm{Im}(z)]^2 \leqslant \delta^2$. Then Ψ has an expansion of the form

$$\Psi(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} r^l g_{l,m}(r) Y_{l,m}(\theta, \phi). \tag{2.17}$$

There exist real numbers $\epsilon > 0, \epsilon_2 > 0$ such that the expansion (2.17) converges uniformly to Ψ for (r, θ, ϕ) in the tubular domain $D_2(0, R, \epsilon, \epsilon_2)$. The domain D_2, defined in Theorem 1, includes the Coulomb singularity at $r = 0$. There exists a real constant M, independent of r, θ, ϕ, l, and m, such that the terms of (2.17) are bounded by

$$|r^l g_{l,m}(r) Y_{l,m}(\theta, \phi)| \leqslant M(l + \tfrac{1}{2})\exp(-\epsilon_2 l) \tag{2.18}$$

for (r, θ, ϕ) in $D_2(0, R, \epsilon, \epsilon_2)$. There exists a real constant M', independent of r, l, and m, such that the bound

$$|r^l g_{l,m}(r)| \leqslant M'(l + \tfrac{1}{2})^{1/2}\exp(-\epsilon_2 l) \tag{2.19}$$

holds for r in $D_3(0, R, \epsilon, \epsilon_2)$. If V_1 and V_2 are polynomials in x, y, and z, then the $g_{l,m}(r)$ are entire functions of r.

Corollary 2: Under the hypotheses of Theorem 2, there exists a real number $a > 0$ such that, for $0 \leqslant |x|^2 + |y|^2 + |z|^2 \leqslant a$, the solution Ψ can be written in the form

$$\Psi(\mathbf{r}) = \Psi_1(x, y, z) + r\Psi_2(x, y, z), \tag{2.20}$$

where Ψ_1 and Ψ_2 are analytic functions of x, y, and z for $0 \leqslant |x|^2 + |y|^2 + |z|^2 \leqslant a$.

III. PROOFS

Proof of Theorem 1: The analyticity of F in the domain $D_1(r_1, r_2, \delta)$ follows from the fact that an analytic function of an analytic function is analytic; $D_1(r_1, r_2, \delta)$ is just the image of $D(r_1, r_2, \delta)$ under the change of variables from x, y, z to r, θ, ϕ. The coefficient $r^l G_{l,m}(r)$ is analytic in r because the integrand in (2.10) is a continuous function of the three variables r, θ, ϕ, analytic in r for fixed θ, ϕ, for all $r \in \mathbb{C}$, $\theta \in \mathbb{R}$, $\phi \in \mathbb{R}$ such that $r_1 \leqslant |\mathrm{Re}(r)| \leqslant r_2$, $|\mathrm{Im}(r)| \leqslant \delta$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. The detailed argument, based on uniform convergence of the Riemann sum to the integral, parallels the argument given in Copson[7] for an integral over one variable. It follows that $G_{l,m}(r)$ is analytic in r except possibly at $r = 0$. If $|x|^2 + |y|^2 + |z|^2 = 0$ is included in the domain of analyticity of $f(x, y, z)$, then there is a real $\epsilon > 0$ such that $f(x, y, z)$ has a power series in x, y, z, convergent for $|x|^2 + |y|^2 + |z|^2 \leqslant \epsilon^2$, which can be written in the form

$$f(x, y, z) = f_1(x, y, z) + f_2(x, y, z), \tag{3.1}$$

where

$$f_1(x, y, z) = \sum_{\substack{i,j,k \\ i+j+k<l}} a_{i,j,k} x^i y^j z^k \tag{3.2}$$

and

$$f_2(x, y, z) = \sum_{\substack{i,j,k \\ i+j+k>l}} a_{i,j,k} x^i y^j z^k. \tag{3.3}$$

The polynomial f_1 does not contribute to $r^l G_{l,m}(r)$ because $Y_{l,m}$ is orthogonal to all spherical harmonics of order less than l. The series f_2, which does contribute to $r^l G_{l,m}(r)$, is of order r^l for $r \to 0$. Thus $G_{l,m}(r)$ is analytic at $r = 0$ if $r = 0$ is contained in D_1.

The uniform convergence of the expansion (2.9) can be established by using the addition theorem for the spherical harmonics to write it in the form

$$F(r, \theta, \phi) = \sum_{l=0}^{\infty} H_l(r, \theta, \phi), \tag{3.4}$$

where

$$H_l(r, \theta, \phi) = \frac{2l+1}{4\pi} \int P_l(\cos\omega) F(r, \theta'', \phi'') \sin\theta'' \, d\theta'' \, d\phi'' \tag{3.5}$$

with $\cos\omega = \cos\theta\cos\theta'' + \sin\theta\sin\theta''\cos(\phi - \phi'')$. Define rotation matrices $R_y(\alpha)$, $R_z(\alpha)$ via

$$R_y(\alpha) := \begin{pmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{pmatrix}, \tag{3.6}$$

$$R_z(\alpha) := \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.7}$$

The coordinate rotation

$$\begin{pmatrix} \sin\theta''\cos\phi'' \\ \sin\theta''\sin\phi'' \\ \cos\theta'' \end{pmatrix} = R_z(\phi) R_y(\theta) R_z(\phi') R_y(\theta') \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.8}$$

on the dummy variables of integration in (3.4) carries ω into θ' and brings (3.4) to the form

$$H_l(r, \theta, \phi) = \int_0^{\pi} P_l(\cos\theta') h_l(r, \theta, \phi; \theta') \sin\theta' \, d\theta' \tag{3.9}$$

with

$$h_l(r, \theta, \phi; \theta') = \frac{2l+1}{4\pi} \int_0^{2\pi} F(r, \theta'', \phi'') d\phi', \tag{3.10}$$

where $F(r, \theta'', \phi'')$ is to be obtained from (2.8) via (3.8). Let $\mathbf{v} = (v_1, v_2, v_3)$ be a vector with real components, and introduce the usual norm

$$\|\mathbf{v}\|_2 = (v_1^2 + v_2^2 + v_3^2)^{1/2}. \tag{3.11}$$

It is then straightforward to show that, for α complex and β equal to either y or z,

$$\|\mathrm{Re}[R_\beta(\alpha)\mathbf{v}]\|_2 \leqslant \cosh[\mathrm{Im}(\alpha)] \|\mathbf{v}\|_2 \tag{3.12}$$

and

$$\|\mathrm{Im}[R_\beta(\alpha)\mathbf{v}]\|_2 \leqslant \sinh[|\mathrm{Im}(\alpha)|] \|\mathbf{v}\|_2. \tag{3.13}$$

Repeated application of the inequalities (3.12) and (3.13) on (3.8) shows that

$$[\mathrm{Re}(\sin\theta''\cos\phi'')]^2+[\mathrm{Re}(\sin\theta''\sin\phi'')]^2+[\mathrm{Re}(\cos\theta'')]^2 \leqslant \cosh^2[|\mathrm{Im}(\theta')|+|\mathrm{Im}(\phi')|+|\mathrm{Im}(\theta)|+|\mathrm{Im}(\phi)|] \tag{3.14}$$

and that

$$[\mathrm{Im}(\sin\theta''\cos\phi'')]^2+[\mathrm{Im}(\sin\theta''\sin\phi'')]^2+[\mathrm{Im}(\cos\theta'')]^2 \leqslant \sinh^2[|\mathrm{Im}(\theta')|+|\mathrm{Im}(\phi')|+|\mathrm{Im}(\theta)|+|\mathrm{Im}(\phi)|]. \tag{3.15}$$

It follows that (r, θ'', ϕ'') will remain in the domain D_1 of analyticity of F if (r, θ, ϕ) is in D_2, $|\mathrm{Im}(\theta')|\leqslant\delta_2$, and $\phi'\in\mathbb{R}$. Hence[7] $h_l(r, \theta, \phi; \theta')$ is an analytic function of r, θ, ϕ, θ' for r, θ, ϕ in $D_2(r_1, r_2, \delta, \delta_2)$ and $-\infty<\mathrm{Re}(\theta')<\infty$, $|\mathrm{Im}(\theta')|\leqslant\delta_2$. Let $\zeta=e^{i\theta'}$, and define k_l by

$$k_l(r, \theta, \phi; \zeta)=h_l(r, \theta, \phi; \theta'). \tag{3.16}$$

Then k_l is an analytic function of r, θ, ϕ, ζ for r, θ, ϕ in $D_2(r_1, r_2, \delta, \delta_2)$ and ζ in the annular region $e^{-\delta_2}\leqslant|\zeta|\leqslant e^{\delta_2}$. H_l is given by

$$H_l(r, \theta, \phi)=\frac{1}{2\pi i}\oint_C Q_l((\zeta+\zeta^{-1})/2)k_l(r, \theta, \phi; \zeta)(\zeta^{-2}-1)d\zeta, \tag{3.17}$$

where the contour C is the circle $|\zeta|=e^{\delta_2}$. Q_l is the second solution of Legendre's equation, which can be defined by the integral representation[8]

$$Q_l(w)=\frac{1}{2}\int_{-\infty}^{\infty}[w+(w^2-1)^{1/2}\cosh(t)]^{-l-1}dt. \tag{3.18}$$

Here $|\arg[w^{-1}(w^2-1)]|\leqslant\pi/2$, the w plane is cut along the real axis from -1 to 1, and neighborhoods of ± 1 are excluded. The previous expression (3.9) for H_l can be recovered from (3.17) by shrinking the contour C to the circle $|\zeta|=1+\epsilon$ where $\epsilon>0$ is arbitrarily small and using the formula[9]

$$Q_l(w)=\frac{1}{2}P_l(w)\ln\left(\frac{w+1}{w-1}\right)+R(w), \tag{3.19}$$

where $R(w)$ is a polynomial of degree $l-1$ in w, to compute the jump across the branch cut which runs from $w=-1$ to $w=+1$ along the real w axis. The uniform convergence of (3.3), and therefore of (2.9), now follows from the discussion given by Szego[10] of the convergence of Jacobi series of analytic functions. The bound (2.11) follows from (3.17) with the aid of the bound

$$|Q_l(w)|\leqslant\frac{\Gamma(l+\frac{1}{2})\sqrt{\pi}}{l!\sqrt{2}|w^2-1|^{1/4}|w+(w^2-1)^{1/2}|^{l+\frac{1}{2}}}, \quad |\arg[w^{-1}(w^2-1)^{1/2}]|\leqslant\pi/2. \tag{3.20}$$

The bound (3.20) can be obtained from the integral representation (3.18) by changing the dummy variable of integration from t to s via $\cosh(t)-1=s^2/2$ and choosing the path of integration so that $\arg[1+w(w^2-1)^{-1/2}]=\arg(s^2)$. On this path, $|\arg(s^2)|\leqslant\pi/2$, $|w+(w^2-1)^{-1/2}\cosh(t)|$ $=|(w^2-1)^{1/2}/2|[|2+2w(w^2-1)^{-1/2}|+|s^2|]$, and $|dt/ds|=|(1+s^2/4)^{-1/2}|\leqslant 1$. These imply the bound

$$|Q_l(w)|\leqslant\tfrac{1}{2}|2(w^2-1)^{-1/2}|^{l+1}\times\int_{-\infty}^{\infty}[|2+2w(w^2-1)^{-1/2}|+|s|^2]^{-l-1}d|s|, \tag{3.21}$$

from which (3.20) follows by carrying out the integration. A saddle point evaluation of (3.18) (the saddle point is at $t=0$) shows that the inequality in (3.20) approaches an equality as $l\to\infty$. The bound (2.12) is obtained by squaring (2.11), multiplying by $\sin\theta$, integrating $0\leqslant\theta\leqslant\pi$, $0\leqslant\phi\leqslant 2\pi$, and taking the (positive) square root of the result.

Corollary 1 can be proven in the same way as Theorem 1.

In order to make the basic ideas of the proof of Theorem 2 easier to follow, the messier details are relegated to Lemmas 1, 2, and 3. These lemmas are stated first. The proof of Theorem 2 is then given, followed by proofs of the lemmas and Corollary 2.

Lemma 1: Let $V(\mathbf{r})$ satisfy the hypothesis of Theorem 2. Then the lowest eigenvalue of the eigenvalue problem

$$H\tilde{\Psi}=E\tilde{\Psi}, \quad \tilde{\Psi}\in H^1(\Omega_{\underline{a}}), \tag{3.22}$$

$$\tilde{\Psi}(\mathbf{r})|_{r=\underline{a}}=0 \tag{3.23}$$

can be made arbitrarily large by choosing $\underline{a}$ sufficiently small.

Lemma 2: The spherical harmonics are bounded by

$$|Y_{l,m}(\theta, \phi)|\leqslant[(2l+1)/(4\pi)]^{1/2}\times\exp[l|\mathrm{Im}(\theta)|-m\,\mathrm{Im}(\phi)]. \tag{3.24}$$

Lemma 3: Let V_1 and V_2 in (1.2) be functions of the three Cartesian coordinates x, y, z, analytic and bounded in absolute value by a positive real constant M_V, for complex x, y, z in and on the ball $0\leqslant|x|^2+|y|^2+|z|^2\leqslant\delta^2, 0<\delta$. Let the value of $\Psi(\mathbf{r})$ for $r=\underline{a}$ be given by

$$\Psi(\mathbf{r})|_{r=\underline{a}}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}b_{l,m}Y_{l,m}(\theta, \phi), \tag{3.25}$$

where $b_{l,m}$ is bounded by

$$|b_{l,m}|\leqslant M_\Psi\exp[-l\delta_\Psi(\underline{a})] \tag{3.26}$$

for some positive real numbers M_Ψ independent of $\underline{a}$ and $\delta_\Psi(\underline{a})$, which may depend on $\underline{a}$. Then for sufficiently small $\underline{a}>0$, the partial differential equation $H\Psi=\bar{E}\Psi$ with the boundary condition (3.25) has a solution with the series representation (2.17), with the series (2.17) uniformly and absolutely convergent throughout the domain

$$|r|\leqslant\underline{a}, \quad |r|\exp[|\mathrm{Im}(\theta)|+|\mathrm{Im}(\phi)|]\leqslant 2\underline{a}. \tag{3.27}$$

There is a constant M'_Ψ such that the bound

$$|a^l g_{l,m}(r)|\leqslant M'_\Psi\exp[-l\delta_\Psi(\underline{a})] \tag{3.28}$$

holds for $|r|\leqslant\underline{a}$. If V_1 and V_2 are polynomials, the functions $g_{l,m}(r)$ in (2.17) are entire functions.

Proof of Theorem 2: For any Ψ satisfying the hypothesis of Theorem 2 and any real number $\underline{a}$ in $0<\underline{a}<R$, the general theory of elliptic partial differential equations[11] implies the existence of a real $\epsilon'(\underline{a})>0$ such that Ψ is an analytic function of x, y, z in the tubular domain $D(\underline{a}, R, \epsilon'(\underline{a}))$ where D is defined in Theorem 1. The expansion (2.17) then follows

from Theorem 1 for $\underline{a} \leqslant r \leqslant R$. The bound (2.12) implies that, for any real $\epsilon_3'(\underline{a}) > 0$,

$$|r^l g_{l,m}(r)| \leqslant \frac{M_\Psi \Gamma(l+\frac{3}{2})}{2\pi^{3/2} l!} \exp[-(l+1)\epsilon_3'(\underline{a})] \tag{3.29}$$

for r real in $\underline{a} \leqslant r \leqslant \underline{b}$ where $\underline{b}$ is the lesser of $R/\cosh[\epsilon_3'(\underline{a})]$ and $\epsilon(\underline{a})/\sinh[\epsilon_3'(\underline{a})]$. The constant M_Ψ in (3.29) can be chosen independent of $\underline{a}$ because $\Psi \in H^1(\Omega_R)$ implies that Ψ is continuous in Ω_R. Now consider the Dirichlet problem

$$H\tilde{\Psi} = E\tilde{\Psi}, \quad \tilde{\Psi} \in H^1(\Omega_a), \tag{3.30}$$

$$\tilde{\Psi}(\mathbf{r})|_{r=a} = \Psi(\mathbf{r})|_{r=a}. \tag{3.31}$$

The problem (3.30), (3.31) has the obvious solution $\tilde{\Psi}$ = the restriction of Ψ to Ω_a; thus existence is trivial. Uniqueness can be established for sufficiently small $\underline{a}$ by assuming that there are two solutions $\tilde{\Psi}_1$ and $\tilde{\Psi}_2$. The difference $\tilde{\Psi}_3 = \tilde{\Psi}_2 - \tilde{\Psi}_1$ then satisfies (3.30) with the homogeneous boundary condition

$$\tilde{\Psi}_3(\mathbf{r})|_{r=a} = 0. \tag{3.32}$$

The homogeneous problem (3.30), (3.32) is just the Dirichlet eigenvalue problem for Ω_a; Lemma 1 implies that all eigenvalues can be made larger than E by choosing $\underline{a}$ sufficiently small. For such $\underline{a}$, (3.30), (3.32) has only the trivial solution $\tilde{\Psi}_3 = 0$ and uniqueness follows.

Now choose $\underline{a}$ sufficiently small both to guarantee uniqueness and to satisfy the requirements of Lemma 3. Lemma 3 then implies that the Dirichlet problem (3.30), (3.31) has a solution $\tilde{\Psi}$ given by a series of the form (2.17), uniformly and absolutely convergent for $|r| \leqslant \underline{a}$, $|r|\exp[|\mathrm{Im}(\theta)| + |\mathrm{Im}(\phi)|] \leqslant 2\underline{a}$. Uniqueness implies that $\tilde{\Psi} = \Psi$. Now choose positive real numbers ρ_1 and ρ_2 such that

$$\rho_1^2 + \rho_2^2 \leqslant 1, \tag{3.33}$$

$$2\rho_2 + (\rho_1^2 + \rho_2^2)^{1/2} \leqslant 2, \tag{3.34}$$

and

$$[2\rho_2 + (\rho_1^2 + \rho_2^2)^{1/2}]\exp[-\delta_\Psi(\underline{a})] < 1. \tag{3.35}$$

Choose positive real numbers ϵ and ϵ_2 such that $\epsilon \leqslant \epsilon'(\rho_1 \underline{a})$, $\epsilon \leqslant \rho_2 \underline{a}$, $D_2(\rho_1 \underline{a}, R, \epsilon, \epsilon_2)$ is nonempty, and $\epsilon_2 \leqslant \delta_\Psi(\underline{a}) - \ln[2\rho_2 + (\rho_1^2 + \rho_2^2)^{1/2}]$. The uniform convergence of the expansion (2.17) and the bounds (2.18), (2.19) for $|\mathrm{Re}(r)| \geqslant \rho_1 \underline{a}$ now follow from Theorem 1 and the analyticity of Ψ at regular points. The uniform convergence of (2.17) and the bounds (2.18), (2.19) for $|\mathrm{Re}(r)| \leqslant \rho_1 \underline{a}$ follow from Lemmas 2 and 3. The fact that the $g_{l,m}$ are entire when V_1 and V_2 are polynomials follows from Lemma 3.

Proof of Lemma 1: The following simple argument was suggested by John D. Morgan III. H can be written as the sum $H = H_a + H_b + H_c$ where $H_a = -\frac{1}{4}\nabla^2$, $H_b = -\frac{1}{4}\nabla^2 - r^{-1}$, and $H_c = V(\mathbf{r})$. Then $E \geqslant E_a + E_b + E_c$ where E_a, E_b, E_c are the lowest points in the spectrum of H_a, H_b, and H_c, respectively. A simple calculation shows that $E_a = \pi^2/(4\underline{a}^2)$. E_b is bounded below by -1, the eigenvalue for the infinite domain.[12] $E_c = \min_{r \in \Omega_a} V(\mathbf{r})$, which is bounded below by the minimum of $V(\mathbf{r})$ over $r \in \mathbb{R}$, $0 \leqslant |r| \leqslant R$, which minimum exists because $V(\mathbf{r})$ is continuous in that domain. The conclusion now follows.

Remark: It can be shown that the lowest eigenvalue of (3.22), (3.23) is actually $\pi^2/(2\underline{a}^2) + O(\underline{a}^{-1})$ for $\underline{a} \to 0$.

Proof of Lemma 2: The estimate

$$|\cos\theta + i\sin\theta\cos t| \leqslant \exp[|\mathrm{Im}(\theta)|] \tag{3.36}$$

applied to the integral representation

$$P_l(\cos\theta) = \frac{1}{2\pi}\int_0^{2\pi} (\cos\theta + i\sin\theta\cos t)^l dt \tag{3.37}$$

yields the bound

$$|P_l(\cos\theta)| \leqslant \exp[l\,|\mathrm{Im}(\theta)|]. \tag{3.38}$$

The addition theorem for the spherical harmonics, when written in the form

$$\sum_{m=-l}^{l} Y_{l,m}(\theta', \phi') Y_{l,m}(\theta, \phi) = \frac{(2l+1)}{4\pi} \exp(2im\phi) P_l(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi')), \tag{3.39}$$

is valid for arbitrary complex θ, ϕ, θ', and ϕ'. It follows from (3.39) that

$$Y_{l,m}(\theta, \phi) Y_{l,m}(\theta', \phi') = \frac{(2l+1)}{8\pi^2} \exp(2im\phi) \times \int_0^{2\pi} P_l(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos t) e^{imt}\, dt. \tag{3.40}$$

Setting $\theta = \theta'$, $\phi = \phi'$, and taking absolute values in (3.40) yields

$$|Y_{l,m}(\theta, \phi)|^2 \leqslant \frac{2l+1}{4\pi} \exp[-2m\,\mathrm{Im}(\phi)] \times \operatorname*{Max}_{0 \leqslant t \leqslant 2\pi} |P_l(\cos^2\theta + \sin^2\theta\cos t)|. \tag{3.41}$$

If $\cos\omega = \cos^2\theta + \sin^2\theta\cos t$ with t real, it can be shown that $\mathrm{Im}|(\omega)|$ is largest when $\cos t = -1$ and $\mathrm{Im}|(\omega)| = 2\,\mathrm{Im}|(\theta)|$. The bound (3.24) then follows from (3.41) with the aid of (3.38).

Proof of Lemma 3: Corollary 1 implies that $V(\mathbf{r})$ has an expansion

$$V(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} r^l V_{l,m}(r) Y_{l,m}(\theta, \phi), \tag{3.42}$$

uniformly convergent for r, θ, ϕ in $D_2'(\delta, \delta_2')$ where $\delta_2' > 0$ is arbitrary, and that $V_{l,m}$ is analytic for $0 \leqslant |r| \leqslant \delta$. This analyticity implies that $r^l V_{l,m}(r)$ has an expansion

$$r^l V_{l,m}(r) = \sum_{n=0}^{\infty} v_{l,m,n} r^{l+n}, \tag{3.43}$$

uniformly convergent for $0 \leqslant |r| \leqslant \delta$. The series

$$\Psi(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{n=0}^{\infty} g_{l,m,n} r^{l+n} Y_{l,m}(\theta, \phi) \tag{3.44}$$

will satisfy $H\Psi = E\Psi$ formally if the coefficient $g_{l,m,n}$ satisfy the recursion relations

$$g_{l,m,1} = -(l+1)^{-1} g_{l,m,0}, \tag{3.45}$$

$$g_{l,m,n+1} = 2(n+1)^{-1}(n+2l+2)^{-1}[\,-g_{l,m,n} - Eg_{l,m,n-1} + \underline{a}_{l,m,n-1}\,], \quad n \geqslant 1, \tag{3.46}$$

where the $\underline{a}_{l,m,n}$ are the coefficients in the expansion

$$V(\mathbf{r})\Psi(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{n=0}^{\infty} \underline{a}_{l,m,n} r^{l+n} Y_{l,m}(\theta, \phi). \tag{3.47}$$

The Wigner 3 j-symbol can be used to write the product of two spherical harmonics in the form

$$\begin{aligned} Y_{l_1,m_1}(\theta,\phi) Y_{l_2,m_2}(\theta,\phi) &= \sum_{l=|l_1-l_2|}^{l_1+l_2} \left[\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}\right]^{1/2} \\ &\times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & -m \end{pmatrix} \\ &\times \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} (-1)^m Y_{l,m}(\theta,\phi), \end{aligned} \tag{3.48}$$

where $m = m_1 + m_2$. It follows from (3.42)–(3.44), (3.47), and (3.48) that

$$\begin{aligned} \underline{a}_{l,m,n} = \sum_{n_1=0}^{n} \sum_{s=0}^{l} \sum_{t=0}^{[(n-n_1)/2]} \sum_{m_1=m_{1\min}}^{m_{1\max}} & v_{l_1,m_1,n_1} g_{l_2,m_2,n_2} (-1)^m \\ &\times \left[\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}\right]^{1/2} \\ &\times \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & -m \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{3.49}$$

with $l_1 = s + t$, $l_2 = l - s + t$, $m_2 = m - m_1$, $m_{1\min} = -\min(l_1, l_2 - m)$, $m_{1\max} = \min(l_1, l_2 + m)$, and $n_2 = n - n_1 - 2t$. The summation limits come from the fact that either or both of the 3 j-symbols in (3.49) are zero unless $l_1 + l_2 + l$ is even, l_1, l_2, and l satisfy the triangle condition, and $m = m_1 + m_2$. The coefficients $g_{l,m,n}$ can be calculated recursively from (3.45), (3.46), and (3.49) if the $g_{l,m,0}$ are prescribed. Bounds on the coefficients $g_{l,m,n}$ will now be calculated. Corollary 1 implies the bound

$$|r^l V_{l,m}(r)| \leqslant \frac{M_v \Gamma(l+\frac{3}{2})}{2\pi^{3/2} l!} \exp[\,-(l+1)\delta_3'\,] \tag{3.50}$$

for $|r| \leqslant \delta\,[1 + 2\sinh^2(\delta_3')]^{-1/2}$. The inequality

$$\Gamma(l+\tfrac{3}{2}) \leqslant [(l+\tfrac{1}{2})\pi]^{1/2} l! \tag{3.51}$$

can be proven by using $\Gamma(l+\frac{3}{2}) = (l+\frac{1}{2})(l-\frac{1}{2})\cdots(\frac{3}{2})(\frac{1}{2})\pi^{1/2}$ and $l! = l(l-1)\cdots(3)(2)(1)$ to write

$$\begin{aligned} \frac{\Gamma(l+\frac{3}{2})}{l!} &= [(l+\tfrac{1}{2})\pi]^{1/2}\left[1 - \frac{l}{4l^2}\right]^{1/2}\left[1 - \frac{1}{4(l-1)^2}\right]^{1/2} \\ &\cdots [\,1 - \tfrac{1}{16}\,]^{1/2}[\,1 - \tfrac{1}{4}\,]^{1/2} \leqslant [(l+\tfrac{1}{2})\pi]^{1/2}. \end{aligned}$$

The bound

$$|v_{l,m,n}| \leqslant (2l+1)^{1/2} C_v a_v^l b_v^n, \tag{3.52}$$

where

$$C_v = 2^{-3/2}\pi^{-1} M_v \exp(-\delta_3'), \tag{3.53}$$

$$b_v = \delta^{-1}[\,1 + 2\sinh^2(\delta_3')\,]^{1/2}, \tag{3.54}$$

and

$$a_v = b_v \exp(-\delta_3'), \tag{3.55}$$

follows from (3.50) and (3.51) with the aid of a Cauchy integral. The orthogonality conditions for the 3 j-symbols imply the (unitarity) bound

$$\left|\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}\right| \leqslant (2j_{\max}+1)^{-1/2}, \tag{3.56}$$

where $j_{\max}$ is the largest of j_1, j_2, j_3. The bounds (3.52) and (3.56) can be used in (3.49) to obtain

$$\begin{aligned} |\underline{a}_{l,m,n}| \leqslant C_v \sum_{n_1=0}^{n} \sum_{s=0}^{l} \sum_{t=0}^{[(n-n_1)/2]} & (2l_1+1)^2 \\ &\times a_v^l b_v^{n_1} \max_{m_2} |g_{l_2,m_2,n_2}|. \end{aligned} \tag{3.57}$$

Make the definition

$$C(\xi,\eta,\zeta) := 9C_v(1-\xi)^{-1}(1-\eta)^{-3}(1-\zeta)^{-3}. \tag{3.58}$$

Let a_g, b_g, C_g be positive real numbers satisfying $a_v a_g^{-1} < 1$, $b_v b_g^{-1} < 1$, $a_v a_g b_g^{-2} < 1$, and

$$b_g \geqslant \tfrac{1}{2} + [\,|E| + \tfrac{1}{4} + C(b_v b_g^{-1}, a_v a_g^{-1}, a_v a_g b_g^{-2})\,]^{1/2}. \tag{3.59}$$

Then

$$|g_{l,m,0}| \leqslant C_g a_g^l \tag{3.60}$$

implies

$$|g_{l,m,n}| \leqslant C_g a_g^l b_g^n, \tag{3.61}$$

as will now be shown via mathematical induction. Assume first that (3.61) holds for $n \leqslant N$. Then (3.57) implies that

$$|\underline{a}_{l,m,n}| \leqslant C_v C_g a_g^l b_g^n \sigma\left(\frac{b_v}{b_g}, \frac{a_v}{a_g}, \frac{a_v a_g}{b_g^2}\right) \tag{3.62}$$

for $n \leqslant N$, where

$$\sigma(\xi,\eta,\zeta) := \sum_{n_1=0}^{n} \sum_{s=0}^{l} \sum_{t=0}^{[(n-n_1)/2]} (2s+2t+1)^2 \xi^n \eta^s \zeta^t. \tag{3.63}$$

Replacing the upper limits by ∞ in (3.63) yields the bound

$$\sigma(\xi,\eta,\zeta) \leqslant (1-\xi)^{-1}(1-\eta)^{-3}(1-\zeta)^{-3}\tau(\eta,\zeta),$$

valid for $0 \leqslant \xi < 1$, $0 \leqslant \eta < 1$, $0 \leqslant \zeta < 1$, where

$$\begin{aligned} \tau(\eta,\zeta) := 9 &- \eta(1-\eta)(\tfrac{9}{2}\zeta^2 + 6) \\ &- \zeta(1-\zeta)(\tfrac{9}{2}\eta^2 + 6) \\ &- \tfrac{3}{2}\eta\zeta(\eta+\zeta) - \eta^2 - \zeta^2 - 4\eta\zeta. \end{aligned} \tag{3.64}$$

Clearly $\tau(\eta,\zeta) \leqslant 9$ for $0 \leqslant \eta \leqslant 1$, $0 \leqslant \zeta \leqslant 1$. The inequality (3.61) for $n = N+1$ now follows from (3.46), (3.62), (3.64), and $\tau \leqslant 9$, thus completing the induction proof. The bounds (3.24) and (3.61) then imply that the series (3.44) converges uniformly and absolutely in the domain defined by the inequalities

$$|r| \leqslant \rho b_g^{-1}, \quad |r|\exp[\,|\mathrm{Im}(\theta)| + |\mathrm{Im}(\phi)|\,] \leqslant \rho a_g^{-1}, \quad 0 < \rho < 1. \tag{3.65}$$

It follows that (3.44) converges to a solution of the partial differential equation $H\Psi = E\Psi$ in the restriction of this domain to the real values for which $H\psi = E\psi$ was defined.

It will now be shown that, for sufficiently small $\underline{a}$, there is a choice of the constants $g_{l,m,0}$ satisfying (3.60) for which the boundary condition (3.25) is satisfied. The series

$$g_{l,m,n} = \sum_{p=0}^{\infty} \gamma_{l,m,n;p} a^{p-l} \tag{3.66}$$

will satisfy the boundary conditions formally if

$$\gamma_{l,m,0;0} = b_{l,m} \tag{3.67}$$

and

$$\gamma_{l,m,0;p} = -\sum_{n=1}^{p} \gamma_{l,m,n;p-n}, \quad p \geqslant 1. \tag{3.68}$$

The $\gamma_{l,m,n;p}$ are to be calculated recursively from the $\gamma_{l,m,0;p}$ via (3.45), (3.46), and (3.49) [just replace $g_{l,m,n}$ by $\gamma_{l,m,n;p}$ in (3.45), (3.46), and (3.49)]. The condition (3.67), the bound (3.26) on $b_{l,m}$, and the fact that (3.60) implies (3.61) can be used to show that

$$|\gamma_{l,m,n;0}| \leqslant M_{\Psi} \exp(-l\delta_{\Psi}) \underline{b}_g^{\,n}. \tag{3.69}$$

Mathematical induction on p can then be used to show that

$$|\gamma_{l,m,n;p}| \leqslant \tfrac{1}{2} M_{\Psi} \exp(-l\delta_{\Psi}) \underline{b}_g^{\,n} (2\underline{b}_g)^p. \tag{3.70}$$

It follows that the series (3.66) converges uniformly and absolutely for

$$\underline{a} \leqslant \tfrac{1}{2} \rho \underline{b}_g^{-1}, \quad 0 < \rho < 1. \tag{3.71}$$

It will now be shown that a suitable choice of the constants $\underline{a}_g$, $\underline{b}_g$, and $\underline{a}$ exists. The constants M_v and δ are determined by the potential $V(\mathbf{r})$; δ'_3 can be chosen freely. With δ'_3 chosen, $\underline{a}_v$, $\underline{b}_v$, and $\underline{c}_v$ are determined by (3.53)–(3.55). Choose $\underline{a}_g = 2\underline{a}_v$, and anticipate that $\underline{b}_g \geqslant \mathrm{Max}(2\underline{a}_v, 2\underline{b}_v)$. Then $C(\underline{b}_v \underline{b}_g^{-1}, \underline{a}_v \underline{a}_g^{-1}, \underline{a}_v \underline{a}_g \underline{b}_g^{-2}) \leqslant C(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) = 1152\, C_v$. Choose $\underline{b}_g = \mathrm{Max}[2\underline{a}_v, 2\underline{b}_v, \frac{1}{2} + (|E| + \frac{1}{4} + 1152\, C_v)]$. Choose ρ in $0 < \rho < 1$. Then for any $\underline{a}$ in $0 < \underline{a} \leqslant \frac{1}{2} \rho \underline{b}_g^{-1}$, the series (3.44), (3.66) converge to a solution of $H\Psi = E\Psi$ with the boundary value (3.42) throughout the restriction of the domain (3.27) [which is a subset of (3.65)] to the reals. The function $g_{l,m}(r)$ of (2.17) is given by

$$g_{l,m}(r) = \sum_{n=0}^{\infty} g_{l,m,n} r^n. \tag{3.72}$$

The bound (3.28) follows from (3.66) and (3.68)–(3.70) with $M_{\Psi'} = (2-\rho)(1-\rho)^{-1} M_{\Psi}$.

It remains to show that $g_{l,m}$ is an entire function when V_1 and V_2 are polynomials. For polynomials V_1 and V_2, there exist $l_{\max}$, $n_{\max}$ such that $v_{l,m,n} = 0$ for $l > l_{\max}$ or $n > n_{\max}$ or both. The terms of the sum (3.49) are then zero for $n_1 > n_{\max}$ or $l_1 > l_{\max}$. It follows that the nonzero terms of (3.49) have $n_2 \geqslant n - n_{\max} - 2l_{\max}$. Define K by

$$K = (n_{\max} + 2l_{\max} + 2)/2, \tag{3.73}$$

where the trivial case $n_{\max} = l_{\max} = 0$ is excluded. Define $B(n)$ by

$$B(n) = \begin{cases} 1, & n \leqslant 2K-2, \\ \{\Gamma[(n+2-K)/K]\}^{-1}, & n > 2K-2. \end{cases} \tag{3.74}$$

The induction proof that (3.60) implies (3.61) can be sharpened to show that (3.60) implies

$$|g_{l,m,n}| \leqslant C_g a_g^l b_g^n / \Gamma[(n+K)/K]. \tag{3.75}$$

The key observation is that, because now $n_2 \geqslant n - n_{\max} - 2l_{\max}$, the validity of (3.75) for $n \leqslant N$ implies that

$$|\underline{a}_{l,m,n}| \leqslant C_v C_g a_g^l b_g^n C(b_v b_g^{-1}, a_v a_g^{-1}, a_v a_g b_g^{-2}) B(n) \tag{3.76}$$

for $n \leqslant N$. The bound (3.75) implies that the $g_{l,m}(r)$ are entire functions.

Proof of Corollary 2: Make the definitions

$$\Psi_1 := \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} g_{l,m,2n} r^{l+2n} Y_{l,m}(\theta, \phi), \tag{3.77}$$

$$\Psi_2 := \sum_{n=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} g_{l,m,2n+1} r^{l+2n} Y_{l,m}(\theta, \phi), \tag{3.78}$$

where the $g_{l,m,n}$ are the expansion coefficients from the expansion (3.44) introduced in the proof of Lemma 3. The bound (3.61) shows that the series (3.77) and (3.78) converge uniformly and absolutely in the domain (3.65). Comparison with (3.44) shows that $\Psi = \Psi_1 + r\Psi_2$. Since $r^l Y_{l,m}(\theta, \phi)$ is a polynomial in x, y, z while $r^2 = x^2 + y^2 + z^2$, the series (3.77) and (3.78) define analytic functions of x, y, and z in the domain (3.65).

Remark: A proof of (2.20) for larger r could not be given because nothing is known about the growth of $g_{l,m}(r)$ with l for $|r| > a$ and $r < 0$.

ACKNOWLEDGMENTS

The author would like to thank Professors Richard J. Weinacht and John D. Morgan III for a number of helpful conversations. This work was supported by the National Science Foundation under Grant No. Phy-8106771.

[1] T. Kato, Trans. Amer. Math. Soc. **70**, 195–211 (1951).
[2] T. Kato, Commun. Pure Appl. Math. **10**, 151–177 (1957).
[3] R. T. Pack and W. Byers Brown, J. Chem. Phys. **45**, 556 (1966); V. Fock, K. Nor. Vidensk. Selsk. Forh. **31**(23), 145 (1958); R. J. White and F. H. Stillinger, J. Chem. Phys. **52**, 5800 (1970); D. L. Knirk, J. Chem. Phys. **60**, 66, 760 (1974); M. Hoffmann-Ostenhof and R. Seiler, Phys. Rev. A **23**, 21 (1981); Bruce R. Johnson, Phys. Rev. A **24**, 2339–2343 (1981); A. E. Carlsson and N. W. Ashcroft, Phys. Rev. B **25**, 3474–3481 (1982).
[4] See, for example, A. Weinstein, *Singular Partial Differential Equations*, Symposium of Institute for Fluid Dynamics and Applied Mathematics, University of Maryland (Gordon and Breach, New York, 1963); R. P. Gilbert, *Function Theoretic Methods in Partial Differential Equations* (Academic, New York, 1969); N. S. Hall, D. W. Quinn, and R. J. Weinacht, SIAM J. Math. Anal. **5**, 111–118 (1974); R. J. Weinacht, SIAM J. Math. Anal. **5**, 147–152 (1974).
[5] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 3rd ed. (Springer, New York, 1966), p. 198.
[6] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. C. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2, p. 182.
[7] E. T. Copson, *An Introduction to the Theory of Functions of a Complex Variable* (Oxford U.P., New York, 1935), Sec. 5.5, pp. 107–109.
[8] Gabor Szego, *Orthogonal Polynomials*, 4th ed. American Mathematical Society Colloquim Publications, Vol. 23 (AMS, Providence, RI, 1975), p. 88, Eq. 4.81.2.
[9] Reference 8, p. 78, Eq. 4.62.7.
[10] Reference 8, pp. 251–253. Szego's α, β, x, y are 0, 0, 1, $\frac{1}{2}(\zeta + \zeta^{-1})$, respectively.
[11] The fact that weak solutions of linear elliptic partial differential equations with C^{∞}-coefficients are (C^{∞}) classical solutions (Weyl's lemma) follows from Sobolev embedding theorems. See Avner Friedman, *Partial Differential Equations* (Holt, Rinehart and Winston, New York, 1969), pp. 46–60. The fact that classical solutions of linear elliptic partial differential equations with analytic coefficients are analytic was proven by Fritz John, Commun. Pure Appl. Math. **3**, 273–304 (1950).
[12] See Hans F. Weinberger, *Variational Methods for Eigenvalue Approximation*, SIAM Regional Conference Series in Applied Mathematics (SIAM, Philadelphia, 1974), pp. 58–60, especially Example. 7.2.

Radiation reaction dynamics in an electromagnetic wave and constant electric field

E. Atlee Jackson
Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green Street, Urbana, Illinois 61801

(Received 6 September 1983; accepted for publication 9 November 1983)

The relativistic motion of a charged particle is studied when it is acted on simultaneously by a constant electric field and a plane electromagnetic wave, propagating in the direction of the electric field (x axis). The dynamics includes the radiation reaction (self-force) on the particle through a standard approximation of the Lorentz–Dirac equation. The interest is to determine the result of the competition between the average acceleration due to the electromagnetic wave ("radiation pressure") and the acceleration due to the constant force of the static field. Each of these actions alone of course produce an unbounded particle energy asymptotically in time. However, it is proved first that, when the "forces" are in opposite directions, the particle can never accelerate (on the average) indefinitely in the x direction, regardless how weak the electric field (E_0) is compared to the amplitude of the wave (A). It is then proved that all solutions converge to a region of zero area in a suitable velocity phase space and, if there exists a periodic solution [in the phase $\xi = \omega(t - x/c)$] in a specified region of this phase space, then all solutions must converge to this solution asymptotically ($\xi \to +\infty$). In the case when (E_0A^2/ω^2) has a specified bound (ω: wave frequency), an iterative method is developed which explicitly yields such a periodic solution, showing that the energy remains bounded. The direction of the average drift is determined in terms of (A, E_0, ω). When the parameter (E_0A^2/ω^2) is above this bound, a combination of numerical and analytic results are obtained which indicate that this periodic solution persists. These results indicate that *all motions tend to states with bounded energy*, regardless of the field strengths.

PACS numbers: 41.70. + t

I. INTRODUCTION

The dynamics of a charged particle, when it reacts to its own radiation, has been of interest for a long time. A number of special solutions have been obtained illustrating the effects of the energy loss by this radiation.[1–5] In the case of a particle in an electromagnetic wave, this energy loss brings about an ever increasing energy of the particle, which is commonly described as a "radiation pressure." Without this radiation loss the particle's energy would remain bounded for all time, but the radiation reaction causes the particle to establish a phase relation with the wave such that it continues to gain energy (on the average) for all time. The gain in the energy, as shown by Gunn and Ostriker,[3] only increases as $t^{1/3}$, indicating that the "radiation pressure" does not act like a constant force, even asymptotically (for γ would then become proportional to t).

The present study is concerned with the simultaneous influence of such "radiation pressure" and a constant (in space and time) electric field which opposes this "pressure" and, in particular, how this radiation reaction is modified by this constant force. Since, as noted above, electromagnetic radiation acting alone is not equivalent to a constant force, it might seem reasonable to expect that a constant force (due to an electric field) would always dominate the "radiation pressure" effect. In this case, a particle would always have an unbounded energy in time, due to the influence of the constant force. Rather surprisingly, this is not the case. It will be shown that when the electric force and "radiation force" are in opposite directions all initial states of the particle tend asymptotically to a state of motion which has a bounded energy and that this feature is essentially independent of the strength of the constant electric field E_0, the amplitude A, or the frequency ω, of the plane electromagentic wave (linearly polarized). What this result means is that the radiation reaction readjusts the dynamics, through the average phase relation which the particle establishes with the wave, in such a manner that it causes the particles to always emit or absorb precisely the same average energy which it acquires by its drift motion along the electric field. This very pretty response and balance between the two actions on the particle was very unexpected to the author.

In Sec. II, the standard approximation of the Lorentz–Dirac equation, which is applicable to relativistic dynamics, will be reviewed. These equations will then be applied to the case of a particle in a plane electromagnetic wave plus a constant electric field along its direction of propagation, and this system will be reduced to the basic form (2.26). While the dynamics are characterized by the three dimensionless parameters

$$\epsilon = E_0/A, \quad \omega_a = qA/mc\omega, \quad \tau_1 = \omega\tau_0 = \omega(2q^2/3mc^3), \tag{1.1}$$

this reduction shows that the dominant parameter is $\epsilon\tau_1\omega_a^3$. In Sec. III, it will be shown that all solutions are captured in a bounded region of a suitable velocity phase space, and from this it is proved that the particle cannot be indefinitely accelerated in the direction of the wave propagation, regardless of the amplitude of the wave. In Sec. IV, it is proved that all solutions tend to a region of zero area in this phase space and that, if there is a solution which remains in a specified region, all solutions will tend asymptotically ($t \to +\infty$) to this solution. In Sec. V, a convergent iterative method is used to de-

 0022-2488/84/051584-08$02.50

termine a solution which is periodic in the phase $\xi = \omega(t - x/c)$ and which is in this specified region, provided that (E_0A^2/ω^2) satisfies a given upper bound. Numerical examples illustrate these figure eight limit cycles, and the direction of the asymptotic drift motion of the particle is determined analytically in terms of (A,E_0,ω). For large values of (E_0A^2/ω^2), a combination of analytic and numerical results are presented which indicate that this globally attracting limit cycle persists for all values of the parameters. The conclusion is that all motions of a particle in these fields maintain a bounded energy for all time, regardless of the field strengths and the frequency (the bound obviously depends on these values).

II. REDUCTION OF THE LORENTZ–DIRAC EQUATION

The classical description of the dynamics of a charged (q) particle when it reacts to its own radiation (self-force) is given by the Lorentz–Dirac[6,7] equation

$$\dot{u}^\mu = \omega^{\mu\nu}u_\nu + \tau_0(\ddot{u}^\mu - \dot{u}^\nu\dot{u}_\nu u^\mu/c^2), \tag{2.1}$$

where $\tau_0 = 2q^2/3mc^3$, $u_\mu = \dot{x}_\mu$ are the components of the four-velocity, the dot represents differentiation with respect to the proper time,

$$\omega^{\mu\nu} = \frac{q}{mc}\begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix} \tag{2.2}$$

is the electromagnetic field tensor, and summation notation is used. In the case of relativistic motion, the $\ddot{u}^\mu$ term ("Schott term") is negligible, and the remainder of the radiation reaction terms can be treated in an iterative approximation,[8] which is valid both at intermediate and extreme relativistic limits.[5] The iteration involves substituting the radiationless dynamics $\dot{u}^\mu = \omega^{\mu\nu}u_\nu$ into the last terms of (2.1). This leads to the following system of equations, which forms the basis of the present study

$$\dot{u}^\mu = \omega^{\mu\nu}u_\nu - Pu^\mu. \tag{2.3}$$

This equation can also be written in the form

$$\frac{d}{dt}(\gamma\boldsymbol{\beta}) = \frac{q}{mc}(\mathbf{E} + \boldsymbol{\beta}\times B) - P\boldsymbol{\beta}. \tag{2.4}$$

The function

$$p = \tau_0\,\omega_{\mu\nu}u^\nu\omega^{\mu\lambda}\frac{u_\lambda}{c^2} = \frac{2q^4\gamma^2}{3m^3c^5}\{(\mathbf{E} + \boldsymbol{\beta}\times B)^2 - (\boldsymbol{\beta}\cdot\mathbf{E})^2\} \tag{2.5}$$

is the instantaneous power radiated by the particle, in units of mc^2.

We now consider a plane linearly polarized electromagnetic wave propagating in the positive x direction, and a constant electric field E_0 along this axis, so that

$$\mathbf{E} = (E_0, A\sin\xi, 0), \quad \mathbf{B} = (0, 0, A\sin\xi), \tag{2.6}$$

$$\xi \equiv \omega t - kx = \omega(t - x/c). \tag{2.7}$$

We introduce the dimensionless proper time $\omega\,dt = \gamma\,d\tau$ and use the dimensionless parameters in (1.1), to write (2.3) in the form

$$\frac{d}{d\tau}u_x = \epsilon\omega_a\gamma + \omega_a u_y\sin\xi - \tau_1 u_x R, \tag{2.8}$$

$$\frac{d}{d\tau}u_y = \omega_a(\gamma - u_x)\sin\xi - \tau_1 u_y R, \tag{2.9}$$

where

$$R = \omega_a^2\{(\epsilon\gamma + u_y\sin\xi)^2 + \sin^2\xi(\gamma - u_x)^2 - [\epsilon u_x + u_y\sin\xi]^2\}. \tag{2.10}$$

This can also be written in the form

$$R = \omega_a^2\{\epsilon^2(\gamma^2 - u_x^2) + 2\epsilon u_y\sin\xi(\gamma - u_x) + \sin^2\xi(\gamma - u_x)^2\}$$

and since $\gamma^2 = 1 + u_x^2 + u_y^2$, the factor $\gamma^2 - u_x^2 = 1 + u_y^2$. This suggests that a more natural function is

$$w = \gamma - u_x. \tag{2.11}$$

Provided that we again use the approximation $\gamma^2 - 1 \simeq \gamma^2$, then

$$\frac{d\gamma}{d\tau} \simeq \epsilon\omega_a u_x + \omega_a u_y\sin\xi - \tau_1 R\gamma \tag{2.12}$$

and using (2.8) one obtains the equation of motion for w

$$\frac{dw}{d\tau} = -\epsilon\omega_a w - \tau_1 R w.$$

We next introduce the phase ξ as the independent variable, and noting that

$$\frac{d\xi}{d\tau} = w, \tag{2.13}$$

the above equation for w then becomes simply

$$\frac{dw}{d\xi} = -\epsilon\omega_a - \tau_1 R. \tag{2.14}$$

From the last expression for R, we obtain

$$R = \omega_a^2\{\epsilon^2 + (\epsilon u_y + w\sin\xi)^2\} \tag{2.15}$$

the equation (2.9) for u_y likewise becomes

$$w\frac{du_y}{d\xi} = \omega_a w\sin\xi - \tau_1 u_y R. \tag{2.16}$$

We note that the term containing R can be eliminated between (2.14) and (2.16)

$$w\frac{du_y}{d\xi} - u_y\frac{dw}{d\xi} = \omega_a(\epsilon u_y + w\sin\xi)$$

and the resulting function on the right side is what occurs in R, (2.15). We can write this in the form

$$\frac{d}{d\xi}\left(\frac{u_y}{w}\right) = \frac{\omega_a}{w}\left(\epsilon\frac{u_y}{w} + \sin\xi\right).$$

From this, it appears that the more natural functions are (note that $w > 0$)

$$u = \frac{-\epsilon\omega_a}{w}, \quad v = \frac{\epsilon u_y}{w}, \tag{2.17}$$

where a scaling has also been introduced, to simplify the coefficients in the following equations. One then obtains

$$\frac{du}{d\xi} = -\alpha u^2 + \sigma(v + \sin\xi)^2, \tag{2.18}$$

$$\frac{dv}{d\xi} = -u(v + \sin\xi), \tag{2.19}$$

where

$$\alpha = 1 + \epsilon\omega_a\tau_1, \quad \sigma = -\epsilon\tau_1\omega_a^3. \tag{2.20}$$

The Eq. (2.18) suggests two final transformations. One is simply to introduce the function

$$p \equiv v + \sin\xi \tag{2.21}$$

so that

$$\frac{dp}{d\xi} = -up + \cos\xi, \tag{2.22}$$

and the other to scale out the coefficient σ in (2.18). In order not to change the simple coefficients of (2.22) the only possible scaling is

$$u' = \delta^{-1}u, \quad p' = \delta p, \quad \xi' = \delta\xi. \tag{2.23}$$

One then readily finds that, if

$$\delta^4 = \sigma \tag{2.24}$$

the equations become

$$\frac{du'}{d\xi'} = -\alpha(u')^2 + (p')^2; \quad \frac{dp'}{d\xi'} = u'p' + \cos\left(\frac{\xi'}{\delta}\right).$$

Now, dropping the primes, setting

$$\eta = \sigma^{-1/4} = [-\epsilon\tau_1\omega_a^3]^{-1/4}, \tag{2.25}$$

and letting the dot now represent differentiation with respect to the (final) ξ variable, the final form of the equations are

$$\dot{u} = -\alpha u^2 + p^2; \dot{p} = -up + \cos(\eta\xi). \tag{2.26}$$

The physical case of interest is when the force due to the constant electric field opposes the "radiation pressure," which here is the positive x-direction [the direction of the Poynting vector, $(\mathbf{E}\times\mathbf{B})c/4\pi$]. This means that $q_0E_0 < 0$, or $\epsilon\omega_a = qE_0/mc\omega < 0$. This is the reason for introducing the negative sign in (2.20) [so that η is real, (2.25)] and in (2.17) [so that $u > 0$]. This means that the case of interest is when $1 > \alpha$. Moreover all electric fields E_0 will be assumed to satisfy

$$1 > -(qE_0/mc)(2q^2/3mc^3). \tag{2.27}$$

For example, for electrons, (2.27) implies that E_0 is less than 3×10^{20} volts/m. For field strengths greater than (2.27), the Eqs. (2.3) are invalid because the quadratic effect of E_0 in (2.5) can dominate the Lorentz force (linear in E_0), thereby violating the assumption basic to the iterative approximation, (2.3) (see also Shen[5]). To be quite certain that this limit is not approached, we shall put a lower bound on α, and also specify that the initial condition of $u(\xi)$ is positive

$$1 \geqslant \alpha \geqslant \tfrac{1}{2}, \quad u(0) > 0. \tag{2.28}$$

III. BOUNDED ATTRACTIVE REGION

We will first prove that all solutions of (2.26) and (2.28) enter and remain in a bounded region of the $(p,u>0)$ phase plane. More specifically it will be proven that

Theorem 1: All solutions of (2.26) and (2.28) tend asymptotically ($\xi \to +\infty$) to the region

$$(p^2 + u^2) < (1.8)^2 + (1.8/\alpha)^{2/3}.$$

Once they *enter* this region they remain there. The physical consequence of this theorem is that all motions asymptotically (in time, $t \to +\infty$) satisfy

$$\sigma^{1/2}\left(\frac{\epsilon\beta_y}{1-\beta_x} + \sin\xi\right)^2 + \sigma^{-1/2}(\omega_e/\gamma(1-\beta_x))^2 < 3.25 + 1.5\alpha^{-2/3}, \tag{3.1}$$

where $\omega_e = \epsilon\omega_a = qE_0/mc\omega$. Note that ξ is a monotonic function of time, by (2.13), so that $\xi \to +\infty$ implies that $\tau \to +\infty$, to yield (3.1). This inequality yields a number of useful physical conclusions, which will be discussed following the proof of Theorem 1.

The proof consists of two parts. First it is shown that all solutions must pass through a specified region (R_4, below) in the phase plane $(p,u>0)$. One then shows that any solution in this region remains bounded for all $\xi \to +\infty$ (in a slightly larger region).

We first define several regions R_i in the phase plane $(p,u>0)$, together with their complements, $\bar{R}_i$ (see Fig. 1). First let

$$R_1 = \{|p| < \alpha u^3\}; \quad \bar{R}_1 = \{|p| > \alpha u^3\}, \tag{3.2}$$

and note that

$$\frac{1}{2}\frac{d}{d\xi}(p^2+u^2) = -p\cos(\eta\xi) - \alpha u^3 \leqslant |p| - \alpha u^3.$$

Therefore, the "radius"

$$\rho = [p^2 + u^2]^{1/2} \tag{3.3}$$

decreases with increasing ξ in R_1,

$$\frac{d\rho}{d\xi} < 0 \ (\text{in } R_1). \tag{3.4}$$

Next, let

$$R_2 = \{|p|u > 1\}, \bar{R}_2 = \{|p|u < 1\},$$
$$R_3 = \{|p| > \sqrt{\alpha}u\}, \ \bar{R}_3 = \{|p| < \sqrt{\alpha}u\}. \tag{3.5}$$

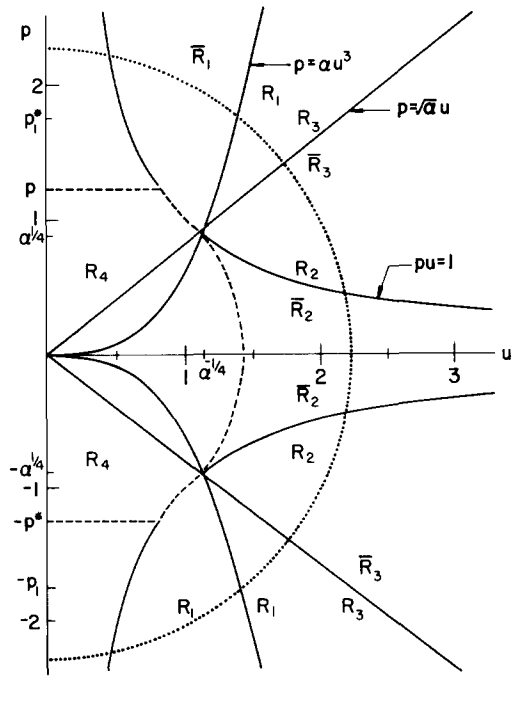


FIG. 1. The boundaries of the regions in the phase space $(p,u>0)$ defined in the text. The region R_4 is enclosed by the dashed curve.

Solutions in these regions, according to (2.26) have the following properties (for all ξ):

$$\operatorname{sgn}(p)\dot{p} < 0 \quad (\text{in } R_2)$$
$$\dot{u} > 0 \quad (\text{in } R_3); \quad \dot{u} < 0 \quad (\text{in } \bar{R}_3). \tag{3.6}$$

Finally we define a rather awkward region

$$R_4 = \bar{R}_2 \cap \bar{R}_3 \cap \{\rho < [\alpha^{1/2} + \alpha^{-1/2}]^{1/2}\} + \bar{R}_2 \cap R_3 \cap \{|p| < p^*\}, \tag{3.7}$$

where p^* will be defined below. These regions are illustrated in Fig. 1, where R_4 is enclosed by a dashed curve.

Because of the properties (3.4) and (3.6), it is clear that all solutions in the following regions proceed (as ξ increases) according to the following sequence:

$$R_1 \cap R_3 \to R_2 \cap \bar{R}_3 \to (R_1 \cap \bar{R}_2) \setminus R_4 \to R_4. \tag{3.8}$$

Next consider $R_2 \cap \bar{R}_1$, in which $\operatorname{sgn}(p)\,\dot{p} < 0$ and $\dot{u} > 0$. It follows that solutions in this region either enter $R_1 \cap R_3$ [thereafter following the route (3.8)] or else they enter $\bar{R}_2 \cap R_3$. What we want to do is to show that, if p^* is selected correctly, they can in fact only go into R_4 (and not into $\bar{R}_2 \cap R_3 \cap \{|p| > p^*\}$). In other words, solutions in $R_2 \cap \bar{R}_1$ can enter $\bar{R}_2 \cap R_3$ only if $p < p^*$. The slope of this boundary is $\operatorname{sgn}(p)\,dp/du = -u^2$ whereas, in $R_2 \cap \bar{R}_1$

$$0 > \operatorname{sgn}(p)\frac{dp}{du} > \frac{-1-pu}{p^2 - \alpha u^2}\operatorname{sgn}(p). \tag{3.9}$$

Therefore, a solution can cross from $R_2 \cap \bar{R}_1$ to $\bar{R}_2 \cap R_3$ only if $\operatorname{sgn}(p)\,(dp/du)_{|p|u=1} < -u^{-2}$, or $(2+\alpha)\,u^4 > 1$. In other words if $p < (2+\alpha)^{1/4}$. Thus, we define

$$p^* \equiv (2+\alpha)^{1/4} \tag{3.10}$$

in which case, all points in $R_2 \cap \bar{R}_1$ now either enter $R_1 \cap R_3$ or R_4 (directly).

Thus we have shown that all solutions are either in R_4 or enter R_4, with the possible exception of the solutions in $\bar{R}_2 \cap \{|p| > p^*\}$. Since all these solutions satisfy $\dot{u} > 0$, they either enter R_4 directly, or $\bar{R}_1 \cap R_2$ (hence to R_4, as just shown), or else $|p| \to \infty$. We now show that the latter possibility does not happen. In $\bar{R}_2$

$$\frac{dp}{du}\operatorname{sgn}(p) = \frac{-up - \cos(\eta\xi)}{p^2 - \alpha u^2}\operatorname{sgn}(p) < \frac{1}{p^2 - \alpha u^2}.$$

Consider an arbitrary initial point

$$|p| = p_0, \quad p_0 u_0 < 1 \quad (p_0 \geqslant p^*), \tag{3.11}$$

then

$$\frac{dp}{du}\operatorname{sgn}(p) < \frac{1}{p_0^2 - (\alpha/p_0^2)}$$

both initially and also for all ξ such that $|p| \geqslant p_0$. Hence, we conclude that, for the initial state (3.11)

$$|p| < p_0 + \frac{u}{p_0^2 - (\alpha/p_0^2)},$$

and hence, since $|p|u \leqslant 1$, there is a maximum value of $|p|$ given by

$$|p| \leqslant p_1 \equiv p_0 + p_1^{-1}[p_0^2 - (\alpha/p_0^2)]^{-1} \quad (\text{in } \bar{R}_2).$$

This yields

$$p_1 = \frac{p_0}{2}\left[1 + \left[1 + \frac{4}{p_0^4 - \alpha}\right]^{1/2}\right]. \tag{3.12}$$

This shows that all solutions in $\bar{R}_2 \cap \{|p| > p^*\}$ leave this region at a finite $|p| < p_1(p_0)$, given by (3.12). This completes the proof that all solutions not in R_4 must enter this region.

They do not, however, necessarily stay in R_4, but we can now easily obtain a bounded region in which they remain. Solutions which leave R_4 have a maximum $|p|$ given by

$$|p| = p^* \quad |p|u < 1,$$

after which their maximum value of $|p|$ is given by (3.12) (with $p_0 = p^*$)

$$p_1^* = \tfrac{1}{2}(2+\alpha)^{1/4}[1 + 3^{1/2}] < 1.8, \tag{3.13}$$

where the inequality is nearly satisfied if $\alpha = 1$ ($p_1^* = 1.797\,79$). If these solutions enter $\bar{R}_1\,R_2$ they cannot obtain a larger value of $|p|$ [since $\operatorname{sgn}(p)\,\dot{p} < 0$], so the largest possible value of ρ when they enter R_1 is

$$\rho^2 = (p_1^*)^2 + (p_1^*/\alpha)^{2/3}. \tag{3.14}$$

This bound on ρ can be reduced by obtaining an upper negative bound on $(dp/du)\operatorname{sgn}(p)$ in $\bar{R}_1\,R_2$, but this refinement will not be pursued. Using the upper value $p^* = 1.8$ of (3.13) in (3.14) gives the bound quoted in the theorem.

The bound, as expressed in the form (3.1), shows that in all cases $\beta_x \not\to +1$ as $\tau \to +\infty$, because $\lim_{\beta_x \to +1}\gamma(1-\beta_x) = 0$. In other words, *regardless how small the static electric field, E_0, the radiation reaction can not cause an indefinite acceleration in the direction of the Poynting vector.* This result is of considerable interest, possibly surprising, but also possibly not (due to the temporal argument in the introduction). In any case it can not be concluded from the inequality (3.1) that β_x does not go to -1, nor even that γ is bounded. Indeed, on the basis of the aforementioned temporal argument, it might seem likely that the particle is always dominated by the constant electric field and hence would expect (since $q\,E_0 < 0$) that $\beta_x \to -1$ as $t \to +\infty$. Certainly this seems reasonable if E_0 is large [but within the bound (2.27)] relative to the amplitude of the electromagnetic wave (say $\epsilon = E_0/A > 1$). Most surprisingly, the Eqs. (2.26) and (2.28) do not yield this result.

IV. ASYMPTOTIC RESULTS

In this section we will establish several more results concerning the asymptotic solutions of (2.26) and (2.28). We know from the proof of Theorem I that all these solutions pass through the region R_4 (Fig. 1), and thereafter remain in the bounded region indicated in that theorem. Unfortunately all points in this region do not necessarily remain in it (only those orbits which filter through R_4), so one cannot employ Brouwer's fixed point theorem to conclude that this region has a periodic solution. However, we can establish several facts about the asymptotic solutions, and then use these to draw conclusions about an explicit periodic solution which will be obtained in the next section.

The first result is:

Theorem II. All solutions of (2.26), (2.28) are asymptotic ($t \to +\infty$) to a region of zero area within the bounded region of Theorem I.

This follows simply from the fact that $\partial\dot{u}/\partial u + \partial\dot{p}/\partial p = -(2\alpha+1)u$, which is negative for all $u > 0$. Thus the area of all regions in the relevant phase space contracts as t increases.

Although the motion is in a plane, it is governed by nonautonomous equations and hence trajectories can have crossings in the plane. Therefore, a convergent flow does not necessarily tend to a closed or nonintersecting curve. One possibility might be for solutions to approach some type of strange attractor, associated frequently with forced nonlinear oscillators. In the present case, the autonomous system is not an oscillator, but simply has a fixed point (global attractor) at the origin. Thus there is no reason to expect this exotic behavior, since there are no competing oscillatory motions involved. At least in the case of large η [small σ, (2.20)] we will be able to show that the region of zero area is in fact a single closed curve (a *unique* periodic solution). To arrive at this property of uniqueness, we next prove:

Theorem III: If any solution of (2.26) and (2.28) remains in the region $4\alpha u^2 > p^2$, then all solutions converge to it.

It of course follows that, if the solution in question is a periodic solution, it is the unique global attractor (limit cycle) of all solutions.

To prove this result assume that (u_1, p_1) is a solution satisfying

$$4\alpha u_1^2 > p_1^2, \tag{4.1}$$

and that (u_2, p_2) is any other solution. Consider the change in the distance between these solutions

$$\begin{aligned}\frac{1}{2}\frac{d}{d\xi}&[(u_1-u_2)^2+(p_1-p_2)^2]\\ &=(u_1-u_2)(-\alpha u_1^2+p_1^2+\alpha u_2^2-p_2^2)\\ &\quad+(p_1-p_2)(-u_1p_1+u_2p_2)\\ &=-\alpha(u_1-u_2)^2(u_1+u_2)+(p_1-p_2)\\ &\quad\times(p_2u_1-p_1u_2).\end{aligned} \tag{4.2}$$

The last factor, considered as a function of (p_1, p_2), has a maximum which is positive. For a fixed p_1 the maximum occurs at $2u_1p_2 = p_1(u_1+u_2)$ at which point the last factor in (4.2) equals $(p_1^2/4u_1)(u_1-u_2)^2$. Therefore,

$$\begin{aligned}\frac{1}{2}\frac{d}{d\xi}&[(u_1-u_2)^2+(p_1-p_2)^2]\\ &\leqslant\left\{-\alpha(u_1+u_2)+\frac{p_1^2}{4u_1}\right\}(u_1-u_2)^2,\end{aligned}$$

which is negative if $(u_1-u_2)\neq 0$ and $4\alpha u_1(u_1+u_2) > p_1^2$. But since $u_2 > 0$, this inequality holds, due to (4.1). Moreover, if $u_1 = u_2$, then the derivative is still negative according to (4.2). It therefore follows that the distance between any distinct solution (p_2, u_2) and the solution satisfying (4.1) decreases as ξ (or t) increases.

This result means that if such a periodic solution (satisfying $4\alpha u^2 > p^2$) can be found, it is the only periodic solution of the system (2.26), (2.28), and it is moreover the global attractor of all solutions. We will not proceed to obtain such a periodic solution in the case of large η (small σ). The case of small η (large σ) will be examined, in Sec. VI.

V. GLOBAL LIMIT CYCLE FOR "SMALL" E_0A^2/ω^2

In this section we will obtain an iterative solution for Eqs. (2.26) and (2.28) which is most useful in the limit

$$\eta = \left[-\left(\frac{mc}{qE_0}\right)\left(\frac{me\omega}{qA}\right)^2\Big/\tau_0\right]^{1/4} \gg 1. \tag{5.1}$$

This condition is usually well satisfied unless the frequency of the radiation is very low. In such a case, the particle experiences a near-constant equal and orthogonal electric and magnetic field (due to the wave), which can produce long intervals of acceleration, so that any ultimate cyclic motion requires a different treatment (see the next section).

The solution of the equations of motion which we seek is the periodic function of the phase which, if it satisfies the condition of Theorem III, will be the global attractor (limit cycle) of all initial states. We begin by setting

$$u = \sum_{k=0}^{\infty} u_k, \quad p = \sum_{k=0}^{\infty} p_k, \tag{5.2}$$

where these functions are required to satisfy the equations

$$\begin{aligned}\dot{u}_0 = 0; \quad \dot{u}_k = &-\alpha\delta_{k_0}u_0^2\\ &-\alpha\sum_{l=0}^{k}u_{k-l}u_l\\ &+\sum_{l=0}^{k-1}p_{k-1-l}p_e \quad (k\geqslant 1),\end{aligned} \tag{5.3}$$

$$\dot{p}_k = \delta_{k0}\cos(\eta\xi) - \sum_{l=0}^{k}p_{k-l}u_l. \tag{5.4}$$

If these equations are satisfied, then the functions (5.2) satisfy the basic Eqs. (2.26). We moreover, are only interested in the periodic solution of (5.3) and (5.4), which is readily found to be the asymptotic solution of these equations. We therefore require that the average of all right-hand sides of (5.3) and (5.4) vanish. More specifically, we require

$$u_k\left(\xi+\frac{2\pi}{\eta}\right)-u_k(\xi)=\int_{\xi}^{\xi+2\pi/\eta}\dot{u}_k(\xi')d\xi'=0, \tag{5.5}$$

$$p_k\left(\xi+\frac{2\pi}{\eta}\right)-p_k(\xi)=\int_{\xi}^{\xi+2\pi/\eta}\dot{p}_k(\xi')d\xi'=0.$$

The lowest-order equations

$$\dot{u}_0 = 0; \quad p_0 = -\cos(\eta\xi) - p_0u_0$$

yield a constant u_0 (unspecified) and the periodic solution

$$p_0 = -\sin(\eta\xi+\phi_0)/[\eta^2+u_0^2]^{1/2}, \quad \tan\phi_0 = u_0/\eta. \tag{5.6}$$

In the next order

$$\dot{u}_1 = -2\alpha u_0u_1 - \alpha u_0^2 + p_0^2; \quad \dot{p}_1 = -u_0p_1 - u_1p_0, \tag{5.7}$$

the condition (5.5) applied to the equation for u_1, and taking u_1 to have no constant portion, determines u_0 in terms of η,

$$u_0^2 = -\tfrac{1}{4}D_{11}^2D_{1\alpha}\sin(2\eta\xi+2\phi_0+\psi_0), \quad \tan\psi_0 = \alpha u_0/\eta, \tag{5.8}$$

where we have introduced the notation

$$D_{jk} = ((j\eta)^2+(ku_0)^2)^{-1/2}. \tag{5.9}$$

Using (5.8) in the above equation for p_1, one obtains

$$\begin{aligned}p_1 = &-\tfrac{1}{8}D_{11}^4D_{1\alpha}\sin(\eta\xi+2\phi_0+\psi_0)\\ &+\tfrac{1}{8}D_{11}^3D_{31}D_{1\alpha}\sin(3\eta\xi+3\phi_0+\psi_0+\phi_0),\\ &\tan\phi_1 = u_0/3\eta.\end{aligned} \tag{5.10}$$

The equation determining u_2 is then found to be

$$\dot{u}_2 = -2\alpha u_0 u_2 - (\alpha/32) D_{11}^4 D_{1\alpha}^2 + \tfrac{1}{8} D_{11}^5 D_{1\alpha} \cos(\phi_0 + \psi_0) - \tfrac{1}{8} D_{11}^4 D_{1\alpha} [D_{11} \cos(2\eta\xi + 3\phi_0 + \psi_0) + D_{31} \cos(2\eta\xi + 2\phi_0 + \psi_0 + \phi_1)] + \tfrac{1}{8} D_{11}^4 D_{1\alpha} [(\alpha/4) D_{1\alpha} \cos(4\eta\xi + 4\phi_0 + 2\psi_0) + D_{31} \cos(4\eta\xi + 4\phi_0 + \psi_0 + \phi_1)], \quad (5.11)$$

and the application of (5.5) determines the constant portion of u_2

$$\bar{u}_2 = -\frac{1}{64u_0} D_{11}^4 D_{1\alpha}^2 + \frac{1}{16\alpha u_0} D_{11}^5 D_{1\alpha} \cos(\phi_0 + \psi_1), \quad (5.12)$$

which is the dominant portion of u_2 in the limit (5.1).

To estimate the order of the functions in (5.2) in the limit of large η, note that according to (5.7) and (5.9)

$$u_0 \sim 1/\eta, \quad D \sim 1/\eta,$$

so that the oscillatory portion of u_2, according to (5.11), goes as $D^6 \eta^{-1} = \eta^{-7}$, whereas the constant factor (5.12) decreases as η^{-4}, and hence is dominant. Proceeding with this analysis on the higher-order equations in (4.3) and (4.4), it is found that the order of the functions for large k is

$$u_k \sim \eta^{-2k}, \quad p_k \sim \eta^{-2(k+1)}.$$

Therefore, the series (5.2) are convergent for $\eta > 1$, but they are of course most useful if η is large. It might be noted that even when η tends to zero, $u_0 \sim 2^{-1/2}$ and $D \sim 2^{1/4}$. Since this is only marginally larger than unity, the series (5.2) may still be convergent due to the other factors (indicated in the above equations). However, in that limit we will proceed by another route in the next section.

If the equations of motion (2.26) are numerically integrated, one readily finds that the solutions tend to this limit cycle. Two examples are illustrated in Figs. 2 and 3, for moderate values of η. Note the small variation in u from its mean value, and how this rapidly decreases even for only a modest increase in η. The dashed curve in these figures indicate the boundary described in Theorem III, which shows that these limit cycles are indeed global attractors.

We can compare the lowest-order prediction of the mean value of u, (5.7), with these results. For $\eta = 3.162$ and $\alpha = 1$, this predicts $u_0 = 2.236$, whereas for $\eta = 4.729$ this yields $u_0 = 0.1495$, both of which are accurate to three significant figures.

While these results clearly indicate that the particle's energy always remains bounded if η is greater than unity, it does not yet give the interesting information about the average drift of the particle along the x axis. An interesting question is "What is the average value of β_x as a function of (A, E_0, ω)?" Unfortunately, this is not very easy to determine, but one might at least be able to answer the question "what is the sign of the average value of β_x?" Knowledge of the sign at least indicates whether the radiation "pressure" or the constant electric field dominates the average drift velocity.

To examine these questions, we note that for the original four velocities $\gamma^2 = 1 + u_x^2 + u_y^2$ and using (2.11) one can readily obtain the expressions

$$u_k = (1 + u_y^2 - w^2)/2w, \quad \gamma = (1 + u^2 + w^2)/2w.$$

Since $\gamma \beta_k = u_k$, this can be expressed in terms of the functions (2.17). Finally the scaling (2.23) can be introduced for u and p in order to conform with the analysis of the last section. In terms of these functions

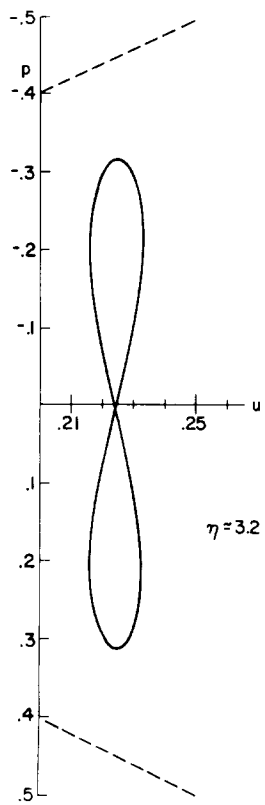


FIG. 2. The limit cycle for the case $\eta = 3.162$. The dashed lines represent the condition of Theorem III, proving that this limit cycle is a global attractor of all solutions.

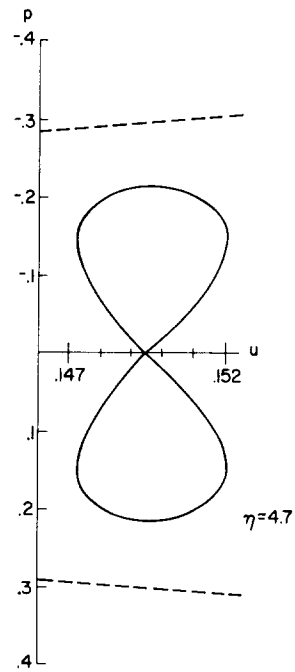


FIG. 3. The limit cycle (global attractor) for the case $\eta = 4.729$. Note the expanded scale on the u axis.

$$\beta_x = [(u/\eta\epsilon\omega_a)^2 + \epsilon^{-2}(\eta p - \sin\eta\xi)^2 - 1]$$
$$\times [(u/\eta\epsilon\omega_a)^2 + \epsilon^{-2}(\eta p - \sin\eta\xi)^2 + 1]^{-1}. \quad (5.13)$$

Unfortunately, the average of this expression is not easily obtained, and even its sign is not obvious. The question of the sign may, however, also be approached by examining the average $dx/d\xi = u_k/w$. From the above relationships, one finds that this transcribes to the expression

$$\frac{dx}{d\xi} = \frac{1}{2}\left[\left(\frac{u}{\eta\epsilon\omega_a}\right)^2 + \epsilon^{-2}(\eta p - \sin\eta\xi)^2 - 1\right], \quad (5.14)$$

which is essentially the numerator of (5.13). While it is clear that the parameter dependence in (5.14) is fairly involved, one can obtain the result for the direction of the drift in the limit of large η without much difficulty. One first substitutes for u and p in (5.14) the results u_0 and $p_0 + p_1$ of (5.6), (5.7), and (5.10). The function p_1 must be considered (in contrast with u_1) because the average of $[\sin(\eta\xi + \phi_0) - \sin(\eta\xi)]^2$ is $1 - \cos(\phi_0) = 1 - \eta(\eta^2 + u_0^2)^{-1/2}$, which goes to $1/(4\alpha\eta^4)$ in the limit of large η and the first factor of p_1, (5.10), can make a similar contribution. There are in fact two such terms which contribute

$$\tfrac{1}{8}\eta D_{11}^4 D_{1\alpha}[\eta D_{11}\cos(\phi_0 - \psi_0) - \cos(2\phi_0 + \psi_0)],$$

and their net contribution decreases as η^{-6}, and hence can be neglected. The conclusion is that the sign of the drift velocity $\bar{\beta}_x$, is given by the sign of

$$1 + (2/\omega_a^2) - 4\alpha\epsilon^2\eta^4 = 1 + (2/\omega_a^2) + 4\alpha(\epsilon/\omega_a^3\tau_1). \quad (5.15)$$

Recall that the wave velocity is positive and the case of interest is when $qE_0 < 0$, or $\epsilon/\omega_a^3 < 0$. The parameters in (5.15) are defined in (1.1) and (2.20). If $\omega_a = qA/mc\omega$ is small, then the sign of the drift velocity is given by the sign of

$$1 + 2\alpha(E_0/A)(mc/qA\tau_0),$$

which is now independent of the frequency. Since $\tau_0 = 6.2\times 10^{-24}$ s (for an electron), it is clear that the drift velocity is negative unless E_0 is very small compared to the wave amplitude.

We next turn to the case of very large values of E_0A^2/ω^2.

VI. VERY STRONG FIELDS, LOW FREQUENCIES

In the last section it was shown that if

$$\eta = -(A/E_0)(mc/qA)^3(\omega^2/\tau_0) \quad (6.1)$$

is large, then the asymptotic dynamics is a limit cycle in the velocity phase space ($p,u > 0$) which has a figure eight structure (Figs. 2 and 3). As η is decreased, it is found numerically that this limit cycle persists, with the only change being that the figure eight is distorted, as illustrated in Fig. 4. An essential point to note is that this limit cycle remains in the region defined in Theorem III, and hence, is still a global attractor of all solutions of (2.26). The analytic method of the last section may in fact be convergent even in this case, but it does not really shed light on the limiting form of this limit cycle, so a new approach will now be taken.

To see the nature of this structure, consider the equations

$$\dot{p} = -up - C, \quad \dot{u} = -\alpha u^2 + p^2, \quad (6.2)$$

which are the same as the basic Eqs. (2.26) except that the function $\cos(\eta\xi)$ has been replaced by a constant C. Shortly we will consider what happens if C varies adiabatically with ξ [corresponding to a small η in $\cos(\eta\xi)$]. For a fixed C, the solutions of (6.2) tend to the fixed point

$$u_0 = \alpha^{-1/4}|C|^{1/2}, \quad p_0 = -C/u_0. \quad (6.3)$$

Note that C may be positive or negative. That is, the characteristic exponents associated with the fixed point always has a negative real part, $-(\alpha + 1/2)u_0$. Indeed a simple phase plane analysis shows that the fixed point is a global attractor.

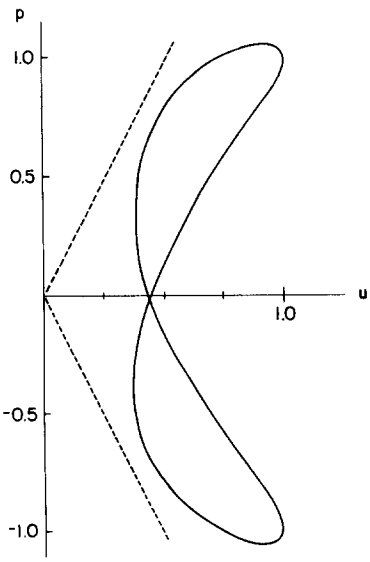


FIG. 4. The limit cycle (global attractor) for a smaller value of η (= 0.5318), indicating both the larger range of u and the distorted figure eight shape.

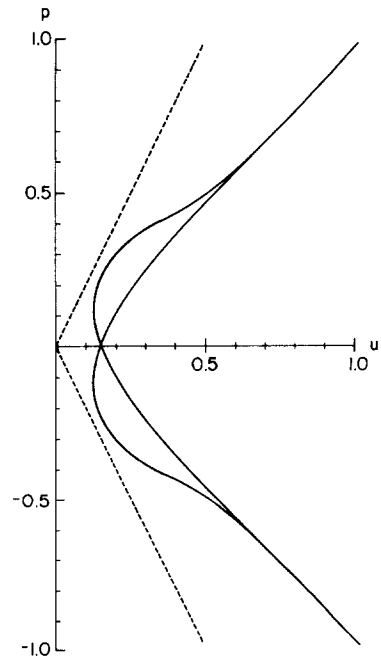


FIG. 5. The structure of the limit cycle for $\eta = 0.0178$. The degenerate straight line portion corresponds to the adiabatically varying "fixed point" (6.3).

If the fractional rate of change in the fixed point is small compared with the decay rate, $(\alpha + 1/2)u_0$, then the solution of (6.2) with variable $C(\xi)$ will closely follow the variable fixed point. Thus the condition for $u_0(\xi)$ and $p_0(\xi)$ to be the solution is, e.g.,

$$\frac{1}{u_0}\frac{du_0}{d\xi} \ll \left(\alpha + \frac{1}{2}\right)u_0,$$

and reintroducing $C(\xi) = \cos(\eta\xi)$, this yields the condition

$$(\cos(\eta\xi))^{3/2} > \frac{\alpha^{1/4}}{(2\alpha + 1)}\eta. \tag{6.4}$$

Therefore, the limit cycle reduces to a V-shaped curve, except where the inequality (6.4) is not satisfied. In the region near the origin, where the inequality (6.4) fails, numerical solutions show that the "crossover" from positive to negative values of p has the character shown in Fig. 5. Unfortunately, no analytic expression for this crossover region has yet been obtained. The limit cycle, however, clearly remains in the region of Theorem III.

VII. CONCLUSION

The very unexpected and beautiful effect found in this study is the radiation reaction which occurs, in the presence of the combined influence of an electromagentic wave and a constant force (any force will do) opposing the "radiation pressure," such that it always establishes an equality between the average energy gained (from both fields) and that lost to radiation. The surprise is not so much that the constant force controls the energy input due to the wave (Sec. III), for the wave alone only causes γ to increase as $t^{1/3}$. What is quite surprising is that this constant force cannot increase the particle's energy indefinitely, apparently because the wave produces just enough deflections to cause the particle to "shed" its excess energy through radiation.

It is believed[9,10] that one important application of this effect is in the structure of pulsar atmospheres (rotating, highly magnetized neutron stars), where the dipole radiation "pressure" is opposed by the charging of the inner region caused precisely by this pressure acting only on charges with one sign (because they are the only particles locally available). In the course of our research concerning radiative dynamics in spatially inhomogeneous fields, Richard Martin[11] found numerical evidence that the parameter $\tau_1 \omega_a^3$ plays an important role in determining the asymptotic $(t \to \infty)$ behavior of the particle. In contrast to the present governing parameter $\eta = -\epsilon\tau_1\omega_a^3$, the influence of ϵ in the inhomogeneous case has not yet been established. We hope in the near future to be able to find a relationship between the spatial limit cycles found numerically in certain spatially inhomogeneous cases by Martin, and the present velocity limit cycles, which generally represent spatial drift motion. This research is presently in progress.

ACKNOWLEDGMENTS

The author acknowledges the helpful support of the University of Illinois at Urbana-Champaign, Department of Physics.

[1]G. N. Plass, Rev. Mod. Phys. **33**, 37 (1961).
[2]T. Erber, Fortschr. Phys. **9**, 343 (1961).
[3]J. E. Gunn and J. P. Ostriker, Astrophys. J. **165**, 523 (1971).
[4]M. Grewing and H. Heintzmann, Phys. Rev. Lett. **28**, 381 (1972).
[5]C. S. Shen, Phys. Rev. D **17**, 434 (1978).
[6]P. A. M. Dirac, Proc. R. Soc. London Ser. A **167**, 148 (1938).
[7]F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, MA, 1965).
[8]L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975).
[9]E. A. Jackson, Astrophys. J. **222**, 675 (1978).
[10]E. A. Jackson, On the Physical Bases of Pulsar Atmospheres [preprint, ILL-(AST)-82-15, May, 1982].
[11]R. F. Martin, Jr., *Theoretical Studies of the Pulsar Magnetosphere*, thesis, Department of Physics, University of Illinois at Urbana-Champaign, 1983.

On the representation of two-dimensional scalar wave fields in the complex plane

M. Nieto-Vesperinas
Instituto de Optica, C.S.I.C., Serrano 121, Madrid 6, Spain

(Received 16 March 1983; accepted for publication 5 August 1983)

Scalar wave fields satisfying the Helmholtz equation in two dimensions are represented by means of a complex variable associated with the two-dimensional physical plane. This characterizes the wave functions as generalizations of analytic functions, which allows the existence of a generalized Cauchy integral formula constituting the nucleus of well-known theorems of optics such as the theorem of Helmholtz and Kirchhoff and the Ewald–Oseen extinction theorem. It also seems useful in the interpretation of inverse diffraction and scattering problems.

PACS numbers: 42.10.Hc, 02.30.Dk

1. INTRODUCTION

The representation of scalar wave fields by means of functions of a complex variable has an interpretative value in optics.[1,2] Also, in the interaction of electromagnetic fields with cylindrical scatterers, there is a useful representation of the wave function in the complex plane which permits the location of singularities and analytic continuation of the fields.[3,4]

As in the case of some elliptic equations[5,6] and, in particular, electrostatic problems, solutions of the two-dimensional Helmholtz equation may be studied in the context of complex variable theory. It will be shown that functions of a complex variable, associated with two-dimensional scalar wave fields, are generalizations of analytic functions. They have a similarity to the generalized analytic functions (in Vekua's terminology[7]), or pseudoanalytic functions (as introduced by Bers[8]). In fact, there exists a *generalized Cauchy integral formula* for those functions, which constitutes the central equation for dealing with boundary-value problems (in the generalized sense of Hilbert).[7,9] As will be seen, this formula yields a Poisson representation theorem and constitutes the complex variable formulation from which the theorem of Helmholtz and Kirchhoff[10,11] for scalar wave fields is obtained. It also provides a framework for the interpretation of inverse diffraction problems and other subjects of recent active research such as the extinction theorem of Ewald and Oseen. This theorem, which was initially a result of molecular optics, describing the extinction of an incident wave inside a medium, and its replacement by another wave with different wave number, has recently received a great deal of attention (see, e.g., Refs. 12–14 and references therein). It has been generalized to different media and to quantum potential scattering and has been interpreted by Wolf and Pattanayak as a nonlocal boundary condition for determining fields in the interior of the medium. It will be seen in this paper that the mathematical aspect of the *extinction theorem* is given by the form adopted by the *generalized Cauchy integral formula* evaluated at a point exterior to a doubly connected domain.

In the next section we shall establish properties characterizing scalar wave fields in the complex plane associated to the $\mathbb{R}^2$ physical plane. This will allow the derivation of the Cauchy formula (Sec. 3) and its consequences for diffraction and scattering problems (Secs. 4 and 5). The two-dimensional treatment presented here is enough to show the essentials of the ideas that we try to put forward. We believe that a full 3-D treatment, although important from a strict formal point of view when dealing with wave fields in three dimensions, would be much more complicated as it should require the use of quaternionic analysis,[15] but would not add many more qualitative results, at least up to the scope pursued here.

2. MATHEMATICAL PRELIMINARIES

Let us consider a two-dimensional monochromatic scalar field $u(x, y)$ satisfying the Helmholtz equation in a domain $\mathscr{D}$, free of sources, of the $\mathbb{R}^2$ plane:

$$\nabla^2 u(x, y) + k^2 u(x, y) = 0\,, \tag{1}$$

k being the field propagation constant.

Let us associate with the $\mathbb{R}^2$ plane the complex E plane by introducing the new variables

$$z = x + iy\,, \quad \bar{z} = x - iy \tag{2}$$

with the operators

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right). \tag{3}$$

The partial derivatives (3) in a domain $\mathscr{D} \subset E$ must be taken in the generalized sense of Sobolev[7] or, equivalently, in the areal sense of Pompeiu.[9] If a continuous function $u(z)$ has these derivatives in $\mathscr{D}$, then that function belongs, respectively, to the manifolds $D_{\bar{z}}\{\mathscr{D}\}$ and $D_z\{\mathscr{D}\}$. Also such a function satisfies[7]

$$\frac{\partial^{p+q} u}{\partial z^p \partial \bar{z}^q} = \frac{\partial^{p+q} u}{\partial \bar{z}^q \partial z^p}\,. \tag{4}$$

The function $u(x, y)$ satisfying Eq. (1) in $\mathscr{D} \subset \mathbb{R}^2$ will satisfy, as a function $u(z) = u_1(x, y) + iu_2(x, y)$ in $\mathscr{D} \subset E$ the following equation:

$$\frac{\partial^2 u(z)}{\partial z\, \partial \bar{z}} + \frac{k^2}{4} u(z) = 0\,. \tag{5}$$

That is, $u(z)$ will belong to the class $D_{\bar{z}}\{\mathscr{D}\}$ and $D_z\{\mathscr{D}\}$. Let us consider $u(z)$ as a function of the manifold $D_{\bar{z}}\{\mathscr{D}\}$; then, evi-

 0022-2488/84/051592-07$02.50

dently, Eq. (5) is equivalent to the pair of equations

$$\frac{\partial u(z)}{\partial \bar{z}} = \pm \frac{k}{2} v(z), \tag{6a}$$

$$\frac{\partial v(z)}{\partial z} = \mp \frac{k}{2} u(z), \tag{6b}$$

which are the complex form of the generalized Cauchy–Riemann systems:

$$\frac{\partial u_1}{\partial x} - \frac{\partial u_2}{\partial y} = \pm k v_1, \quad \frac{\partial u_2}{\partial x} + \frac{\partial u_1}{\partial y} = \pm k v_2, \tag{7a}$$

$$\frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} = \mp k u_1, \quad \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y} = \mp k u_2. \tag{7b}$$

When $k = 0$, then $\partial u/\partial \bar{z} = 0$, i.e., $u(z)$ becomes analytic in $\mathscr{D}$ and $u(x, y)$ is harmonic.

We shall denote by $\mathscr{U}(\mathscr{D})$ the class of functions u satisfying Eqs. (6) in $\mathscr{D} \subset E$. [We could develop a parallel formalism for the function $v(z)$, but we shall be only interested in the function $u(z)$ as the descriptor of the scalar wave field.] Functions of the class $\mathscr{U}(\mathscr{D})$ have a close relationship with the class of *generalized analytic functions* in the sense of Vekua[7] or *pseudoanalytic functions* as defined by Bers.[8] In fact, as will be seen, functions of the manifold $\mathscr{U}(\mathscr{D})$ satisfy some theorems that are similar to those established by Vekua and Bers.

Theorem 1: The function $u \in \mathscr{U}(\mathscr{D})$ may be expressed as

$$u(z) = \Phi(z) \mp \frac{k}{2\pi} \int\int_{\mathscr{D}} \frac{v(\zeta)}{\zeta - z} d\xi\, d\eta, \tag{8}$$

$$\zeta = \xi + i\eta,$$

Φ being an analytic function.

Proof: Since $u(z) \in D_z(\mathscr{D})$, $D_{\bar{z}}(\mathscr{D})$ it satisfies the well-known Green formulas[7]

$$\frac{1}{2i} \int_\Gamma u(z)\, dz = \int\int_{\mathscr{D}} \frac{\partial u}{\partial \bar{z}} dx\, dy, \tag{9a}$$

$$-\frac{1}{2i} \int_\Gamma u(z)\, d\bar{z} = \int\int_{\mathscr{D}} \frac{\partial u}{\partial z} dx\, dy, \tag{9b}$$

Γ being the boundary of $\mathscr{D}$.

Let z be a fixed point of $\mathscr{D}$. Applying Eq. (9a) to the domain $\mathscr{D}_\epsilon$, which is the intersection of $\mathscr{D}$ and the domain $|z - \zeta| > \epsilon$, where ϵ is a sufficiently small positive number, one easily obtains

$$u(z) = -\frac{1}{2\pi i} \int_\Gamma \frac{u(\zeta)\, d\zeta}{\zeta - z} - \frac{1}{\pi} \int\int_{\mathscr{D}} \frac{\partial u(\zeta)}{\partial \bar{\zeta}} \frac{d\xi\, d\eta}{\zeta - z}. \tag{10}$$

The first term of the right-hand side of Eq. (10) does not depend on $\bar{z}$; hence it satisfies the ordinary Cauchy–Riemann equations, and thus it is an analytic function. Substituting the value of $\partial u/\partial \bar{\zeta}$ given by Eq. (6a) into the second term of Eq. (10), one obtains Eq. (8).

The following theorem is analogous to the reciprocal theorem of Vekua[7] or the similarity principle of Bers[8]:

Theorem 2: The function $u(z) \in \mathscr{U}(\mathscr{D})$ may be written as

$$u(z) = \Phi(z) e^{\omega(z)}, \tag{11}$$

where $\Phi(z)$ is an analytic function of $\omega(z)$ is a continuous function in $\mathscr{D}$ of the class $D_{\bar{z}}(\mathscr{D})$. Moreover, $\omega(z)$ is given by the expression

$$\omega(z) = \mp \frac{k}{2\pi} \int\int_{\mathscr{D}} \frac{v(\zeta)/u(\zeta)}{\zeta - z} d\xi\, d\eta. \tag{12}$$

Proof: From Eq. (11) one has

$$\frac{\partial u}{\partial \bar{z}} = u(z) \frac{\partial \omega}{\partial \bar{z}}. \tag{13}$$

Comparing Eqs. (6a) and (13)

$$\frac{\partial \omega}{\partial \bar{z}} = \pm \frac{k}{2} \frac{v(z)}{u(z)}. \tag{14}$$

If $z = z_0$ is a zero of order k of $u(z)$, then $u(z)$ may be written

$$u(z) = A(z)(z - z_0)^k,$$

$$A(z_0) \neq 0. \tag{15}$$

But, according to Eq. (6a), from Eq. (15) is obtained

$$v(z) = \pm \frac{2}{k} \frac{\partial A}{\partial \bar{z}} (z - z_0)^k. \tag{16}$$

And from Eq. (16) one has by virtue of Eq. (6a) and (6b)

$$k \frac{\partial A}{\partial \bar{z}} = -\left[\frac{\partial^2 A(z)}{\partial z\, \partial \bar{z}} + \frac{k^2}{4} A(z)\right](z - z_0). \tag{17}$$

The right-hand side of Eq. (17) is a function that vanishes at $z = z_0$, therefore the quotient

$$\frac{v(z)}{u(z)} = \pm \frac{2}{k} \frac{1}{A(z)} \frac{\partial A}{\partial \bar{z}} \tag{18}$$

will vanish at $z = z_0$ [$A(z) \neq 0$ at $z = z_0$].

We conclude, thus, that the right-hand side of Eq. (14) is zero at those points where $u(z)$ is zero. Moreover, we infer from Eq. (14) and Theorem 1 that $\omega(z)$ satisfies Eq. (12).

The representation (11) permits associating to every analytic function $\Phi(z)$ of $\mathscr{D} \subset E$ a function $u(z) \in \mathscr{U}(\mathscr{D})$. Since $e^{\omega(z)}$ is regular and has no zeros, the singularities and zeros of $u(z)$ coincide with the singularities and zeros of $\Phi(z)$. Hence, it follows that if $u(z)$ does not vanish identically, its poles and zeros are isolated and the multiplicity of a zero and the order of a pole are positive integers.

However, unlike analytic and generalized analytic functions, $u(z)$ of the class $\mathscr{U}(\mathscr{D})$ does not satisfy the maximum modulus principle (see also Ref. 16) nor is every uniformly bounded function of the class $u(E)$ a constant, although it satisfies the following theorem:

Theorem 3: Every function $u(z) \in \mathscr{U}(E)$ continuous and uniformly bounded on the whole complex plane E has the form

$$u(z) = c e^{\omega(z)}, \tag{19}$$

c being a constant.

Proof: By virtue of Liouville's theorem for analytic functions under the hypotheses of this theorem the function $\Phi(z)$ of Theorem 2 will become a constant c.

An example of a function $u(z)$ satisfying Theorem 3 is the plane wave given by Eq. (19) with

$$\omega(z) = \tfrac{1}{2} k\, [(q + ip)z - (q - ip)\bar{z}],$$

$$p^2 + q^2 = 1, \tag{20}$$

q and p being, respectively, the cosine directors of the direc-

tion of propagation. The function $\omega(z)$ given by (20) satisfies the equations of Theorem 2 and the corresponding $u(z)\in\mathscr{U}(E)$.

Another example of wave functions $u(z)$ satisfying Theorem 3 is given by the class of *source-free fields* studied by Sherman[17], which by virtue of Theorem II of Ref. 17 satisfy Eqs. (6) on the whole E.

Analogously to Theorem 3, it is straight-forward to prove:

Theorem 4: A function $u(z)\in\mathscr{U}(\mathscr{D})$ which is zero at a set of points of $\mathscr{D}$ with a limit point will vanish identically in $\mathscr{D}$.

As a consequence of Theorem 4 one obtains the following corollary:

Corollary 1: Two functions of the class $\mathscr{U}(\mathscr{D})$ that coincide at a set of points with a limit point in $\mathscr{D}$ are identical in $\mathscr{D}$.

This corollary establishes, in analogy with the theory of analytic functions, that two functions of the class $\mathscr{U}(\mathscr{D})$ that are identical on an arc of $\mathscr{D}$ are identical in the whole $\mathscr{D}$. In fact, as we shall see, there exists a generalized Cauchy integral formula that connects the values of a function $u\in\mathscr{U}(\mathscr{D})$ inside a domain with the values of u on its contour. This leads to the Poisson representation for the Helmholtz equation and constitutes the formulation of the Huygens description of wave field propagation.

3. THE FUNDAMENTAL KERNELS AND THE GENERALIZED CAUCHY FORMULA

Let $\Omega_1(z,t)$ and $\overline{\Omega}_2(z,t)$ be solutions of Eqs. (6). i.e., $\Omega_1(z,t)$ and $\Omega_2(z,t)$ [complex conjugate of $\overline{\Omega}_2(z,t)$], satisfy

$$\frac{\partial\Omega_1}{\partial\bar{z}} = \pm\frac{k}{2}\overline{\Omega}_2\,, \tag{20a}$$

$$\frac{\partial\Omega_2}{\partial\bar{z}} = \mp\frac{k}{2}\overline{\Omega}_1 \tag{20b}$$

and correspond, respectively, to the analytic functions $\Phi_1(z) = 1/t - z$ and $\Phi_2(z) = 0$, t being a fixed point of E. By virtue of Theorem 1, Ω_1 and Ω_2 satisfy the representations

$$\Omega_1(z,t) = \frac{1}{t-z} \mp \frac{k}{2\pi}\int\int_{\mathscr{D}} \frac{\overline{\Omega}_2(\zeta,t)}{\zeta - z}\,d\xi\,d\eta\,, \tag{21a}$$

$$\Omega_2(z,t) = \pm\frac{k}{2\pi}\int\int_{\mathscr{D}} \frac{\overline{\Omega}_1(z,t)}{\zeta - z}\,d\xi\,d\eta\,. \tag{21b}$$

Evidently, Ω_1 and Ω_2 hold the following conditions:

$$\lim_{z\to t}(t-z)\Omega_1(z,t) = 1\,, \tag{22a}$$

$$\lim_{z\to t}(t-z)\Omega_2(z,t) = 0\,. \tag{22b}$$

The functions $\Omega_1(z,t)$ and $\Omega_2(z,t)$ are called the *fundamental kernels* of the class $\mathscr{U}(\mathscr{D})$.

Now, let us establish the following theorem:

Theorem 5: The function $u(z)\in(\mathscr{D})$ is given at points $z\in E$ from its values at a contour Γ enclosing the domain $\mathscr{D}$ by means of the *generalized Cauchy integral*

$$\frac{1}{2\pi i}\int_\Gamma [v(\zeta)\overline{\Omega}_2(\zeta,z)d\bar{\zeta} - u(\zeta)\Omega_1(\zeta,z)d\zeta\,] = \begin{cases} u(z) & \text{when } z\in\mathscr{D}\,,\\ \frac{1}{2}u(z) & \text{when } z\in\Gamma\,,\\ 0 & \text{when } z\notin\mathscr{D}+\Gamma\,. \end{cases} \tag{23}$$

Proof: Let Γ be a contour of a domain $\mathscr{D}$. It is the union of a finite number of simply smooth closed Jordan curves in which the function $u(z)$ and $\Omega_1(z,t)$, $\Omega_2(z,t)$ satisfy, respectively, Eqs. (6) and (20) and $\mathscr{D}+\Gamma$ does not contain the point $z=t$; both $u(z)\Omega_1(z,t)$ and $v(z)\Omega_2(z,t)$ are continuous and belong, respectively, to the manifolds $D_{\bar{z}}(\mathscr{D})$ and $D_z(\mathscr{D})$. Hence, the Green formulas, Eqs. (9), may be applied:

$$\frac{1}{2i}\int_\Gamma u(z)\Omega_1(z,t)dz = \int\int_{\mathscr{D}}\frac{\partial}{\partial\bar{z}}[u(z)\Omega_1(z,t)]dx\,dy$$

$$= \int\int_{\mathscr{D}}\left[\Omega_1(z,t)\frac{\partial u}{\partial\bar{z}} + u(z)\frac{\partial\Omega_1}{\partial\bar{z}}\right]dx\,dy\,.$$

And by means of Eqs. (6a) and (20a) one obtains

$$\frac{1}{2i}\int_\Gamma u(z)\Omega_1(z,t)dz = \pm\frac{k}{2}\int\int_{\mathscr{D}}[u(z)\overline{\Omega}_2(z,t) + \Omega_1(z,t)v(z)]dx\,dy\,. \tag{24}$$

In a similar fashion, one obtains from Eq. (9b), and by using Eqs. (6b) and (20b)

$$-\frac{1}{2i}\int_\Gamma v(z)\overline{\Omega}_2(z,t)d\bar{z} = \mp\frac{k}{2}\int\int_{\mathscr{D}}[v(z)\Omega_1(z,t) + \overline{\Omega}_2(z,t)u(z)]dx\,dy\,. \tag{25}$$

Then, summing (24) and (25) we obtain

$$\int_\Gamma [u(z)\Omega_1(z,t)dz - v(z)\overline{\Omega}_2(z,t)d\bar{z}] = 0\,. \tag{26}$$

If $t\in\mathscr{D}$, applying Eq. (26) to the domain $\mathscr{D}_\epsilon$ bounded by Γ and the circle $\Gamma_\epsilon \equiv |z-t| = \epsilon$, we get

$$\int_\Gamma [u(z)\Omega_1(z,t)dz - v(z)\overline{\Omega}_2(z,t)d\bar{z}] - \int_{\Gamma_\epsilon}[u(z)\Omega_1(z,t) - v(z)\overline{\Omega}_2(z,t)d\bar{z}] = 0\,.$$

Therefore, when $\epsilon\to 0$ we obtain by virtue of the conditions (22)

$$\int_\Gamma [u(z)\Omega_1(z,t)dz - v(z)\overline{\Omega}_2(z,t)d\bar{z}] = -2\pi i u(t)\,. \tag{27}$$

When $t\in\Gamma$ we obtain an analogous expression in the right-hand side of which appear $\pi i u(t)$. In other words, we have derived the *generalized Cauchy formula* of Eq. (23).

Evidently, when $k=0$ then $\Omega_1(\zeta,z) = 1/(z-\zeta)$ and $\Omega_2(z,\zeta) = 0$, so that Eq. (23) becomes the classical Cauchy formula for analytic functions.

As can be guessed and will be seen later, the fundamental kernels Ω_1 and Ω_2 are closely related with the Green's function of the Helmholtz equation. As such, we shall look for solutions to Eqs. (20) that satisfy conditions (22), and besides for $z\neq 0$, $t\to\infty$,

$$\Omega_1(z,t) = O(|t|^{-1/2})\,, \tag{28a}$$

$$\Omega_2(z,t) = O(|t|^{-1/2}), \tag{28b}$$

which coincide with the behavior at infinity of cylindrical waves. The solutions to Eqs. (20) subject to the conditions (22) and (28) are

$$\Omega_1(z,t) = \pm \pi i \frac{k}{2} \frac{|t-z|}{t-z} H_1^{(2)}(k|t-z|), \tag{29a}$$

$$\Omega_2(z,t) = \pi i \frac{k}{2} H_0^{(2)}(k|t-z|), \tag{29b}$$

$$|t-z| = [(t-z)(\bar{t}-\bar{z})]^{1/2}.$$

$H_\nu^{(1)}$ and $H_\nu^{(2)}$ are respectively, the first and second Hänkel functions of order ν, whose limiting forms are[18]

$kr \to 0$:

$$-iH_0^{(1)}(kr) \sim iH_0^{(2)}(kr) \sim (2/\pi)\ln(kr),$$
$$-iH_\nu^{(1)}(kr) \sim iH_\nu^{(2)}(kr) \sim -(1/\pi)\Gamma(\nu)(\tfrac{1}{2}kr)^{-\nu} \quad (\mathrm{Re}\,\nu > 0), \tag{30}$$

$kr \to \infty$:

$$H_\nu^{(2)}(kr) \sim \sqrt{\frac{2}{\pi kr}} e^{\pm i(kr - \nu\pi/2 - \pi/4)}. \tag{31}$$

The two signs in front of the right-hand side of expression (29a) as well as in Eqs. (20) and (6) must be understood in the following way: the upper sign is associated to $H_1^{(1)}$ and $H_0^{(2)}$ whereas the lower sign is chosen with $H_1^{(2)}$ and $H_0^{(1)}$. As will be seen, $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are, respectively, associated to outgoing and incoming waves.

It can easily be verified that Eq. (23) can be also applied to an unbounded domain if we impose upon $u(z)$ the additional condition

$$\frac{|t-z|}{t-z}\frac{\partial u}{\partial \bar{z}} \pm i\frac{k}{2}u = 0(|z|^{-1/2}) \tag{32}$$

for t fixed and $z \to \infty$.

This condition implies the following asymptotic law for $u(z)$:

$$u(z) \sim \frac{\pi i k}{2}\sqrt{\frac{2}{\pi k|t-z|}}\, e^{\pm i[k|t-z| - \pi/4]}. \tag{33}$$

That is, $u(z)$ vanishes at infinity as fast as a diverging (upper sign) or converging (lower sign) cylindrical wave. Equation (32) is the counterpart in the E-complex plane of Sommerfeld radiation condition.

With condition (32), and providing that $u(z)$ satisfies Eqs. (6) in the upper half plane (uhp), $y > 0$, Eq. (23) yields a representation for $u(z)$ for $y > 0$, in terms of the generalized Cauchy integral on the real axis:

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} [v(\xi)\bar{\Omega}_2(\xi,z) - u(\xi)\Omega_1(\xi,z)]d\xi = \begin{cases} u(z), & y>0, \\ \tfrac{1}{2}u(z), & y=0, \\ 0, & y<0, \end{cases} \tag{34}$$

$$\zeta = \xi + i\eta,$$

where the contour Γ of Eq. (23) has been chosen to be the real axis and a semicircle in $y > 0$ centered at the origin and whose radius R is taken in the limit $R \to \infty$.

Equation (34) constitutes the Poisson representation for $u(z) \in \mathscr{U}(\mathscr{D})$, $\mathscr{D}$ being now the upper half of the plane E. As such, it can be regarded as an equivalent representation to that of Heins[19] (see also Ref. 20).

When only $u(\xi)$ is specified on $\eta = 0$, subtracting from Eq. (34) the same integral but now evaluated at $\bar{z}$ instead of at z yields

$$-\frac{1}{\pi i}\int_{-\infty}^{\infty} u(\xi)[\Omega_1(\xi,z) - \Omega_1(\xi,\bar{z})]d\xi = u(z), \quad y>0, \tag{35}$$

which is another Poisson representation equivalent to the Rayleigh–Sommerfeld diffraction integral in the $\mathbb{R}^2$-plane since evidently one may obtain from Eq. (29a)

$$\Omega_1(\xi,z) - \Omega_1(\xi,\bar{z}) = \frac{i}{4}\frac{\partial G}{\partial \mathbf{n}}, \tag{36}$$

$\mathbf{n}$ being the unit normal to Ox along the positive Oy-axis, and G being Sommerfeld's Green function[21]

$$G(k|z-\zeta|) = \pm \pi i\left[H_0^{(2)}(k|z-\zeta|) - H_0^{(2)}(k|\bar{z}-\zeta|)\right]. \tag{37}$$

Let us denote the left-hand side of Eq. (23) by $\Sigma_\Gamma^{(2)}(z)$:

$$\Sigma_\Gamma^{(2)}(z) = \frac{1}{2\pi i}\int_\Gamma [v(\zeta)\bar{\Omega}_2(\zeta,z)d\bar{\zeta} - u(\zeta)\Omega_1(\zeta,z)d\zeta]. \tag{38}$$

The first of Eqs. (23), corresponding to $z \in \mathscr{D}$, is analogous to the theorem of Helmholtz and Kirchhoff in $\mathbb{R}^2$ for scalar wave fields.[10,11] Hence, the generalized Cauchy integral represents for points inside $\mathscr{D}$ the formulation in the E-plane of the Huygens–Fresnel description. In fact, $\Sigma_\Gamma^{(2)}(z)$ given by Eq. (38) may be written as

$$\Sigma_\Gamma^{(2)}(z) = \mp\frac{1}{2}\int_\Gamma \left[H_0^{(2)}(k|\zeta - z|)\frac{\partial u}{\partial \bar{\zeta}}d\bar{\zeta} + u(\zeta)\frac{\partial}{\partial \zeta}H_0^{(2)}(k|\zeta - z|)d\zeta\right].$$

By making use of Eqs. (2) and (3), the above takes the form

$$\Sigma_\Gamma^{(2)}(z) = \mp\frac{1}{4}\int_\Gamma \left[u(\nabla H_0^{(2)}\cdot d\mathbf{s} + i|\nabla H_0^{(2)}\times d\mathbf{s}|) + H_0^{(2)}(\nabla u \cdot d\mathbf{s} - i|\nabla u \times d\mathbf{s}|)\right],$$

$d\mathbf{s}$ being the arc element vector of Γ taken counterclockwise. Then, $\nabla f \cdot d\mathbf{s}$ and $|\nabla f \times d\mathbf{s}|$ represent, respectively, the components of ∇f along the tangent $\mathbf{t}$ and the normal $\mathbf{n}$ to Γ[22]. Therefore we have

$$\Sigma_\Gamma^{(2)}(z) = \mp\frac{1}{4}\int_\Gamma \nabla(H_0^{(2)}u)\cdot d\mathbf{s} \mp \frac{i}{4}\int_\Gamma (u\nabla H_0^{(2)}\cdot \mathbf{n}\, ds - H_0^{(2)}\nabla u \cdot \mathbf{n}\, ds). \tag{39}$$

The first integral of Eq. (39) is evidently zero since Γ is closed. Thus we are left with

$$\sum_{\Gamma}^{(1)(2)}(z)=\frac{1}{4\pi}\int_{\Gamma}\left(G^{(2)}\frac{\partial u}{\partial n}-u\frac{\partial G^{(2)}}{\partial n}\right)ds\,, \tag{40}$$

where $G^{(2)}$ denotes the Green's function, respectively, for outgoing and incoming waves,

$$G^{(2)}(k\,|\zeta-z|)=\pm\,\pi i H_0^{(2)}(k\,|\zeta-z|)\,. \tag{41}$$

Equation (40) is the Helmholtz–Kirchhoff integral in $\mathbb{R}^2$.

4. BOUNDARY VALUES. THE EXTINCTION THEOREM

Let us consider a domain $\mathscr{D}^{(-)}$ of the E-plane containing a wave-field source distribution, and bounded by a contour Γ. $\mathscr{D}^{(-)}$ is included in a larger domain $\mathscr{D}$ of boundary Γ'. We shall denote by $z_<$ a point $z\in\mathscr{D}^{(-)}$ and by $z_>$ a point $z\in\mathscr{D}^{(+)}=\mathscr{D}-\mathscr{D}^{(-)}$ (Fig. 1). Evidently $u(z)$ satisfies Eqs. (6) outside $\mathscr{D}^{(-)}$, i.e., $u(z)\in\mathscr{U}(\mathscr{D}^{(+)})$.

For $z\in\mathscr{D}^{(-)}$, the generalized Cauchy formulas, Eqs. (23) applied to the doubly connected domain $\mathscr{D}^{(+)}$ yield

$$0=\sum_{\Gamma'}^{(1)(2)}(z_<)-\sum_{\Gamma}^{(1)(2)}(z_<)\,, \tag{42}$$

where the notation of Eq. (38) has been used.

Analogously, for $z\in\mathscr{D}^{(+)}$, Eqs. (23) applied to the domain $\mathscr{D}^{(+)}$ give us

$$u(z_>)=\sum_{\Gamma'}^{(1)(2)}(z_>)-\sum_{\Gamma}^{(1)(2)}(z_>)\,. \tag{43}$$

Note that in (42) and (43) the integrals $\Sigma_{\Gamma'}^{(1)(2)}$ (and $\Sigma_{\Gamma}^{(1)(2)}$) involve the values of $u(\zeta)$ and $\partial u/\partial\bar{\zeta}$ on Γ' (and Γ).

In a direct diffraction or *scattering problem* the values $u(z_>)$ are to be determined from $\Sigma_{\Gamma}^{(1)}$ by means of Eq. (43). In fact, by writing the total field as

$$u(z)=u^{(i)}(z)+u^{(s)}(z)\,,$$

where $u^{(i)}(z)$ is a field wave function of the class $\mathscr{U}(E)$, i.e., a function that satisfies Eqs. (6) everywhere in E. It is usual to identify $u^{(i)}(z)$ either with an incident field or zero according to whether there are in $\mathscr{D}^{(-)}$ secondary or primary sources. On the other hand, $u^{(s)}(z)$ represents the field generated by the sources of $\mathscr{D}^{(-)}$ and is usually referred to as the *scattered field* or the field *radiated* by the sources at $\mathscr{D}^{(-)}$. $u^{(s)}(z)$ satisfies the radiation condition for outgoing waves [Eq. (32) with the upper sign]. Then, it is evident that $\Sigma_{\Gamma}^{(1)}(z_>)$ is equal to

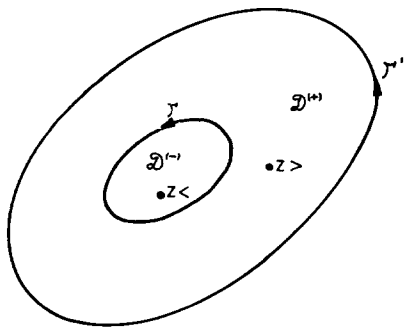


FIG. 1. Contours and domains in the E-plane.

the value of $\Sigma_{c^{(\infty)}}^{(1)}$, i.e., to the integral (38) with the upper sign over a circle of radius that is taken in the limit tending to infinity. Then we have

$$\sum_{c^{(\infty)}}^{(1)}=u^{(i)}(z)\,, \tag{44}$$

since $u^{(i)}$ satisfies the first of Eqs. (23) everywhere in the E-plane and the component of $\Sigma_{c^{(\infty)}}^{(1)}$ due to $u^{(s)}(z)$ is zero by virtue of Eq. (32) with the upper sign. Equation (44) is independent of whether $z\in\mathscr{D}^{(-)}$ or $z\in\mathscr{D}^{(+)}$.

Thus, Eqs. (42) and (43) yield for the direct problem

$$0=u^{(i)}(z_<)-\sum_{\Gamma}^{(1)}(z_<)\,, \tag{45}$$

$$u(z_>)=u^{(i)}(z_>)-\sum_{\Gamma}^{(1)}(z_>)\,. \tag{46}$$

Eq. (45) may be recognized as the *extinction theorem* of optics, which constitutes a boundary condition (nonlocal) on Γ for $u(z_<)$.[12–14] On the other hand, Eq. (46) represents the direct exterior solution. We thus see how both expressions, and in particular the extinction theorem of molecular optics, ultimately are consequences of the generalized Cauchy integral formulas for functions of the class $\mathscr{U}(\mathscr{D}^{(+)})$. More specifically, the extinction theorem constitutes the result obtained from the generalized Cauchy formula applied to a point exterior to a doubly connected domain $\mathscr{D}^{(+)}$.

In an *inverse diffraction* or *scattering problem* either the values $u(z_>)$ or even the values $u(\zeta)$ and $\partial u/\partial\bar{\zeta}$ on Γ are to be determined from $u(\zeta)$ and $\partial u/\partial\bar{\zeta}$ on Γ', i.e., from $\Sigma_{\Gamma'}^{(2)}$. In this case the choice of the lower superindex of Ω_1, and Ω_2 in Eqs. (29) indicates that we are considering convergent cylindrical waves and reconstructing the initial values of $u(z)$ from those on a surface Γ' towards which the wave field $u(z)$ has propagated. In this case Eqs. (42) and (43) yield

$$0=\sum_{\Gamma'}^{(2)}(z_<)-\sum_{\Gamma}^{(2)}(z_<)\,, \tag{47}$$

$$u(z_>)=\sum_{\Gamma'}^{(2)}(z_>)-\sum_{\Gamma}^{(2)}(z_>)\,. \tag{48}$$

Note that $u(z)$ may be decomposed according to Eq. (44), it satisfies the same radiation condition at infinity and, thus, now $\Sigma_{\Gamma'}^{(2)}$, *does not* become $u^{(i)}$. This is a fundamental difference between direct and inverse problems.

Analogously to the extinction theorem, Eq. (45), the new equation (47) constitutes a nonlocal boundary condition for determining the interior field $u(z_<)$ and in particular its value $u(\zeta)$ and $\partial u/\partial\bar{\zeta}$ on the boundary Γ of $\mathscr{D}^{(-)}$. The *inverse exterior solution* $u(z_<)$ is then determined from these values and $\Sigma_{\Gamma'}^{(2)}$ by means of Eq. (48).

In the case in which $u(z_>)$ represents a *source free field*[17] the contribution of $\Sigma_{\Gamma}^{(2)}(z_>)$ in Eq. (48) is negligible. This is in agreement, on the other hand, with the fact that if $u(z_>)$ is source-free it may be extended into the domain $\mathscr{D}^{(-)}$ (Ref. 17) to obtain a bounded solution of Eqs. (6) for all the E-plane and, thus, the Cauchy formula for $u(z)$ will apply to any simply connected domain $\mathscr{D}^{(+)}\subset E$, without excluding a domain of sources as in Eq. (48). Hence, the Cauchy integral formula yields for a source-free field the following *inverse*

equation instead of Eq. (48):

$$u(z_>) = \sum_{\Gamma'}^{(2)} (z_>) . \tag{49}$$

In particular, when Γ' is chosen as the real axis and a semicircle in $y > 0$ centered at the origin and whose radius R is taken in the limit $R \to \infty$, Eq. (49) becomes the 2-D analogous to the inverse diffraction equation of Shewell and Wolf [Eq. (4.9) of Ref. 23].

In fact the boundary values $\partial u/\partial \zeta$ and $u(\zeta)$ may be considered as given on the real axis and any parallel to the real axis in the uhp, $y > 0$, by means of a conformal mapping that transforms the Γ-contour into the line $y = 0$. The domain $\mathscr{D}^{(-)}$ thus becoming the lhp which contains the sources, and the domain exterior to $\mathscr{D}^{(-)}$ being mapped into the uhp free of sources.

5. THE GENERALIZED CAUCHY FORMULA IN A DOMAIN CONTAINING SOURCES

We should add to the previous analysis, that it is even possible to obtain a generalized Cauchy formula for a function $u(z)$ in a domain $\mathscr{D}$ containing sources, such as the domains $\mathscr{D}^{(-)}$ or $\mathscr{D}$ of Fig. 1.

This is the case for a wave field $u(z)$ satisfying the complex form of the inhomogeneous Helmholtz equation,

$$\frac{\partial^2 u(z)}{\partial z\, \partial \bar{z}} + \frac{k^2}{4} u(z) = \frac{1}{4} \rho(z) , \tag{50}$$

$\rho(z)$ denoting the source distribution.

Equation (50) is equivalent to the pair of equations

$$\frac{\partial u}{\partial \bar{z}} = \pm \frac{k}{2} v(z) , \tag{51a}$$

$$\frac{\partial v}{\partial z} = \mp \frac{k}{2} u(z) \pm \frac{2}{k} \rho(z) . \tag{51b}$$

Of course, the function $u(z)$ satisfying Eqs. (51) is no longer a function of the class $\mathscr{U}(\mathscr{D})$. However, in an analogous way as Theorem 5 was proven, it is easy to derive the following generalized Cauchy formula for $u(z)$:

$$\frac{1}{2\pi i} \int_\Gamma [v(\zeta)\bar{\Omega}_2(\zeta,z)d\bar{\zeta} - u(\zeta)\Omega_1(\zeta,z)d\zeta] \pm \frac{1}{2\pi k} \int\int_{\mathscr{D}} dx\, dy\, \rho(z)\bar{\Omega}_2(\zeta,z) = \begin{cases} u(z) , & z \in \mathscr{D} , \\ \frac{1}{2} u(z) , & z \in \Gamma , \\ 0 , & z \notin \mathscr{D} + \Gamma , \end{cases} \tag{52}$$

Γ being the boundary of $\mathscr{D}$.

Equations (52), when used together with Eqs. (42) and (43), may transform contour integrals into domain integrals and, in particular, may be used for solving the inverse interior problem, namely, that of finding the source distributions $\rho(z)$ from the knowledge of $\Sigma^{(2)}_{\Gamma,(1)}(z)$.[24,25]

6. SUMMARY AND CONCLUSIONS

We have established a complex variable characterization of scalar wave fields that satisfy the Helmholtz equation in two dimensions. This permits us to obtain fundamental mathematical properties for the wave functions in the complex plane associated to the physical $\mathbb{R}^2$-plane. Specifically, these functions have been seen to be generalizations of analytic functions with many similar properties; in particular, the possibility of establishing a generalized Cauchy integral formula which is fundamental for solving boundary-value problems. This formula is revealed as the nucleus of classical theorems of optics such as the Helmholtz–Kirchhoff theorem and the Ewald and Oseen extinction theorem, and may also be useful in the formulation of inverse scattering and diffraction problems. The analytic nature of the wave functions, in the general sense used here, seems therefore to be a fundamental characteristic of scattered fields: specifically, it is at the root of the conservation of information in wave propagation (Huygens' principle) as stems from the Cauchy integral.

ACKNOWLEDGMENTS

The author acknowledges several useful discussions with Dr. A. Méndez.

[1] H. M. Nussenzveig, "Causality and analyticity in optics," in *AIP Conference Proceedings* **65** (American Institute of Physics, New York, 1981), pp. 9–30.

[2] G. Ross, M. A. Fiddy, and M. Nieto-Vesperinas, "The inverse scattering problem in structural determinations," in *Inverse Scattering Problems in Optics*, edited by H. P. Baltes, Topics in Current Physics **20** (Springer-Verlag, Berlin, 1980), pp. 15–71; M. A. Fiddy, G. Ross, M. Nieto-Vesperinas, and A. M. J. Huiser, "Encoding of information in inverse optical problems," Optica Acta **29**, 23–40 (1982).

[3] R. F. Millar: "Singularities of two-dimensional exterior solutions of the Helmholtz equation," Proc. Camb. Phil. Soc. **69**, 175–188 (1971).

[4] M. Neviere, M. Cadilhac, and R. Petit, "Applications of conformal mappings to the diffraction of E.M. waves by a grating," IEEE Trans. Antennas Propag. **AP-21**, 37–46 (1973).

[5] I. N. Vekua, *New Methods for Solving Elliptic Equations* (North-Holland, Amsterdam, 1968).

[6] C. Miranda, *Partial Differential Equations of Elliptic Type* (Springer-Verlag, Berlin, 1970), §53.

[7] I. N. Vekua, *Generalized Analytic Functions* (Pergamon-Addison Wesley, Reading, MA, 1962).

[8] L. Bers, "Function theoretic aspects of the theory of elliptic partial differential equations," in *Methods of Mathematical Physics*, by R. Courant and D. Hilbert (Interscience, New York, 1962), Vol. 2, pp. 374–406; L. Bers, "An outline of the theory of pseudoanalytic functions," Bull. Am. Math. Soc. **62**, 291–331 (1956).

[9] F. D. Gakhov, *Boundary Value Problems* (Pergamon-Addison Wesley, Reading, MA, 1966), Chap. 5.

[10] M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1975), 5th ed., §8.3.

[11] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, 1968), Chap. 3.

[12] D. N. Pattanayak and E. Wolf, "General form and a new interpretation of the Ewald-Oseen extinction theorem," Opt. Commun. **6**, 217–220 (1972).

[13] D. N. Pattanayak and E. Wolf, "Scattering states and bound states as solutions of the Schrödinger equation with nonlocal boundary conditions", Phys. Rev. D **13**, 913–923 (1976).

[14] E. Wolf, "A generalized extinction theorem and its role in scattering theory," in *Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf (Plenum, New York, 1973), pp. 339–357.

[15] A. Sudbery, "Quaternionic analysis," Proc. Camb. Phil. Soc. **85**, 199–225 (1979).

[16] C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer-Verlag, Berlin, 1969).

[17] G. C. Sherman, "Diffracted wavefields expressible by plane-wave expansions containing only homogeneous waves," Phys. Rev. Lett. **21**, 761–764 (1968); **21**, 1220 (E) (1968); J. Opt. Soc. Am. **59**, 697–711 (1969).

[18] F. W. J. Olver, in *Handbook of Mathematical Functions*, edited by M.

Abramowitz and I. Stegun (Dover, New York, 1972), Chap. 9.
[19]A. E. Heins, "On an integral equation in diffraction theory," Compos. Math. **18**, 49–54 (1967).
[20]R. F. Millar, "The analytic continuation of solutions to elliptic boundary value problems in two independent variables," J. Math. Anal. Appl. **76**, 498–515 (1980).
[21]A. Sommerfeld, *Optics* (Academic, New York, 1964), §34.
[22]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1, §4.1.
[23]J. R. Shewell and E. Wolf, "Inverse diffraction and a new reciprocity theorem," J. Opt. Soc. Am. **58**, 1596–1603 (1968).
[24]N. Bleistein and J. K. Cohen, "Nonuniqueness in the inverse source problem in acoustics and electromagnetics," J. Math. Phys. **18**, 194–201 (1977).
[25]N. N. Bojarski, "Inverse scattering inverse source theory," J. Math. Phys. **22**, 1647–1650 (1981).

A mechanical system with a "wild" horseshoe

Jair Koiller
Instituto de Matemática, Universidade Federal do Rio de Janeiro, Caixa Postal 68530, Rio de Janeiro-RJ, Brazil 21944

(Received 28 April 1983; accepted for publication 18 July 1983)

We consider a gyrostat in its simplest form, namely, a rigid body with a flywheel attached. It is well known that, although the motions are coupled, when the rotor is symmetric, the full system is completely integrable in terms of elliptic functions. The Euler equations for the angular momentum components with respect to the rigid body now depend on the parameter I = angular momentum of the flywheel (a constant of motion) and the phase portrait of the system undergoes several bifurcations. Analytical formulas, in terms of elementary functions, can be given for the separatrices. A small imperfection in the flywheel breaks the S^1 symmetry; as a result, the separatrices split with transversal intersection, producing horseshoes. We applied the techniques of Holmes and Marsden to write down the Melnikov function of the system. The integral is computed by the method of residues, in the limit case $I \to 0^-$. Somewhat surprisingly, the amplitude of the Melnikov function diverges. We propose an explanation for this "paradox."

PACS numbers: 46.10. + z

1. INTRODUCTION

Considerable effort has been devoted in recent years to the study of "chaotic behavior in deterministic systems."[1] A relatively simple mechanism for the presence of erratic trajectories was discovered by Poincaré and Birkhoff[2] and set into an abstract framework by Smale,[3] who called it the horseshoe.

Topologically, the horseshoe is a map $f: \mathbb{R}^2 \to \mathbb{R}^2$ looking like Fig. 1(a). Although the dynamics of f is complicated, it can be well understood on a certain subset.

Horseshoes appear frequently in Poincaré maps associated with differential equations. Consider an autonomous system in the plane (x,y) containing a separatrix connecting two saddles, as in Fig. 1(b). If this system is subjected to a T-periodic time-dependent perturbation, then generically the Poincaré map between sections $t = t_0$ and $t = t_0 + T$ will have horseshoes. This phenomenon is due to the transversal splitting of the separatrix [Fig. 1(c)].

The quantitative aspects of the issue were treated by Melnikov and Arnold.[4] The Melnikov function $M = M(t_0)$, measures, to first approximation on a perturbation parameter, the gap between the curves C^u and C^s. Here t_0 is the time along the unperturbed separatrix $(\bar{x}(t), \bar{y}(t))$. When $M(t_0)$ has simple zeros, then C^u and C^s intersect transversally; so $M(t_0) = 0$ means that a transversal intersection occurs near $(\bar{x}(t_0), \bar{y}(t_0))$.

In typical examples, the Melnikov function is an oscillating trigonometric polynomial, whose amplitude can be interpreted as a quantitative measurement of the "observability" of the horseshoe. See Holmes and Marsden.[5] In a series of papers, Holmes and Marsden further developed the method of Melnikov.[6,7]

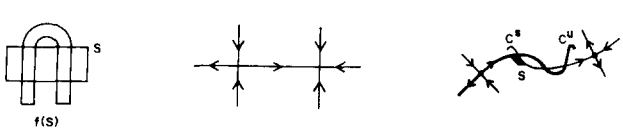


FIG. 1. (a) The horseshoe; (b) the separatrix $(\bar{x}(t), \bar{y}(t))$; (c) transversal splitting: C^u = unstable manifold, C^s = stable manifold.

In this work we apply their techniques to the example consisting of a rigid body with an attachment. This example was also proposed by Holmes and Marsden.[8] This paper is organized as follows: In Sec. 2 we treat the unperturbed system and sketch its phase portrait; the results are probably classical; in Sec. 3 we review the Holmes–Marsden theorem and suggest a slight modification in order to simplify computations; in Sec. 4 we write down the Melnikov function and establish the existence of horseshoes; finally, conclusions and related questions are presented in Sec. 5.

2. THE UNPERTURBED SYSTEM

A system of rigid bodies is usually meant as any finite number of rigid bodies coupled in arbitrary fashion. In recent years the dynamics of multibody systems became a field of increasing interest, motivated by such various applications as spacecraft engineering, biodynamics, and molecular chemistry. It was realized that even the problem of writing down the equations of motion is nontrivial, and several authors have devised ingenious ways to do it.[9] Here we deal with a very simple multibody system.

The system studied in this paper is depicted in Fig. 2. Here J_1, J_2, J_3 are the the inertia moments of the main body, $K_1 = K_2 = K$, K_3 the inertia moments of the attachment; moreover, m is the mass of the rotor and $L = \|0P\|$. This is the unperturbed system; in the next section we will consider small imperfections of mass $\mu/2$ added to the points P_1, P_2 at distance r of P. Then μ will be the perturbation parameter introducing the horseshoes. To simplify the analysis, we neglect the gravity ($g = 0$) which can also be achieved when $L = 0$.

Without loss of generality, we may assume that the attachment point $P = CM_2$ belongs to the line $0Z_1$. Notice, however, that the direction cosines (a,b,c) of the axis of rota-

 0022-2488/84/051599-06$02.50

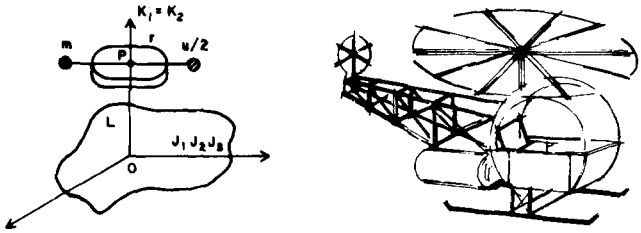


FIG. 2. (a) The example to be analyzed: configuration space $SO(3)\times S^1$; (b) rigid body with two flywheels: a mechanical system on $SO(3)\times T^2$ ($T^2 = S^1\times S^1$) whose Lagrangian is a left-invariant metric.

tion, with respect to the main body frame, can be arbitrary. For the time being we do not require that $0X_1Y_1Z_1$ be the principal axis of the rigid body. So let A_1 be its inertia matrix; then the Lagrangian of the unperturbed system is given by

$$L = T_0 = \frac{1}{2}\langle A_1\Omega,\Omega\rangle + \frac{1}{2}(K+mL^2)(\Omega_1^2+\Omega_2^2) + \frac{1}{2}K_3\Omega_3^2 + \tfrac{1}{3}K_3\dot\theta^2 + K_3(a\Omega_1+b\Omega_2+c\Omega_3)\dot\theta, \tag{2.1}$$

where $\Omega = (\Omega_1,\Omega_2,\Omega_3)$ is the angular velocity vector of the main body and $\dot\theta$ the angular velocity of the flywheel.

Notice the coupling term between these velocities. To save space, we omit the derivation of this formula. It can be done, in a straightforward way, using the elementary techniques of textbooks such as Arnold's.[10] We remark that T_0 is a left-invariant metric on the direct product $SO(3)\times S^1$.

It is convenient to pass to the Hamiltonian formulation. For simplicity we assume now that the attachment axis PZ_2 is coincident with a principal axis (say $0Z_1$) of the rigid body. After some manipulations we obtain

$$H^0 = \frac{1}{2}\left[\frac{M_1^2}{J_1+K+mL^2} + \frac{M_2^2}{J_2+K+mL^2} + \frac{1}{J_3}(M_3-I)^2 + \frac{I^2}{K_3}\right], \tag{2.2}$$

where the Legendre transformation is given by

$$\dot\theta = \left(\frac{1}{J_3}+\frac{1}{K_3}\right)I - \frac{1}{J_3}M_3, \tag{2.3}$$

$$\Omega_1 = \frac{M_1}{J_1+K+mL^2},\quad \Omega_2 = \frac{M_2}{J_2+K+mL^2},$$

$$\Omega_3 = -\frac{I}{J_3}+\frac{M_3}{J_3}.$$

It is well known that I is a constant of motion,[11] but notice that $\dot\theta = I/K_3 - \Omega_3$, showing that the flywheel reacts to the movement of the rigid body. It is easy to produce a modern derivation of several related classical results, using the notations of Marsden and Weinstein.[12] It follows that the solution curves $M(t)$ lie on the intersection of spheres

$$M_1^2+M_2^2+M_3^2 = l^2 \tag{2.4}$$

with the ellipsoids

$$\frac{M_1^2}{\alpha}+\frac{M_2^2}{\beta}+\frac{(M_3-I)^2}{\gamma} = 2h - \frac{I^2}{\delta}, \tag{2.5a}$$

where

$$\alpha = J_1+K+mL^2,\ \beta = J_2+K+mL^2,\ \gamma = J_3,\ \delta = K_3. \tag{2.5b}$$

The solution curves satisfy the differential equations

$$\dot M_1 = M_2M_3\left(\frac{1}{\gamma}-\frac{1}{\beta}\right) - \frac{IM_2}{\gamma},$$

$$\dot M_2 = M_1M_3\left(\frac{1}{\alpha}-\frac{1}{\gamma}\right) + \frac{IM_1}{\gamma}, \tag{2.6}$$

$$\dot M_3 = M_1M_2\left(\frac{1}{\beta}-\frac{1}{\alpha}\right).$$

For $I = 0$ these are the traditional Euler equations, except by the fact that α,β,γ are no longer geometrically constrained.[13] We wonder if this remark could be of interest for higher dimensions [$SO(n)$, $n\geqslant 4$], where the integrability of the Euler equations depend on these relations.[14]

In Fig. 3 we sketch the phase portraits of the system on the sphere $\|M\| = l$, using I as bifurcation parameter. There are two different cases:

(a) $\alpha>\beta>\gamma, I\geqslant 0$ ($I<0$ and/or $\gamma>\beta>\alpha$ are analogous);

(b) $\alpha>\gamma>\beta, I\geqslant 0$ ($I<0$ and/or $\alpha<\gamma<\beta$ are analogous).

A remarkable fact is that the separatrices can be quantitatively described in terms of elementary functions!

3. THE MELNIKOV FUNCTION

In this section we make some observations on the technique of Holmes and Marsden[8] to show the existence of chaotic behavior in perturbations of integrable Hamiltonian systems. Our aim is to apply their Theorem 6.4 to the rigid body with a flywheel.

Following their notation, the four-dimensional symplectic manifold P is here the product $S^2(l)\times T^*S^1$ of the sphere $M_1^2+M_2^2+M_3^2 = l^2$ with the cylinder $T^*S^1 = \{(\theta,I)\}$. The notations correspond as follows:

Holmes/Marsden	ours
u	M
ψ	θ
J	I
ϵ	μ

The Poisson bracket of g,h: $P\to R$ is given by

$$\{g,h\} = \frac{\partial g}{\partial\theta}\frac{\partial h}{\partial I} - \frac{\partial g}{\partial I}\frac{\partial h}{\partial\theta} - M\cdot\underset{M}{\text{grad}}\, g\times\underset{M}{\text{grad}}\, h, \tag{3.1}$$

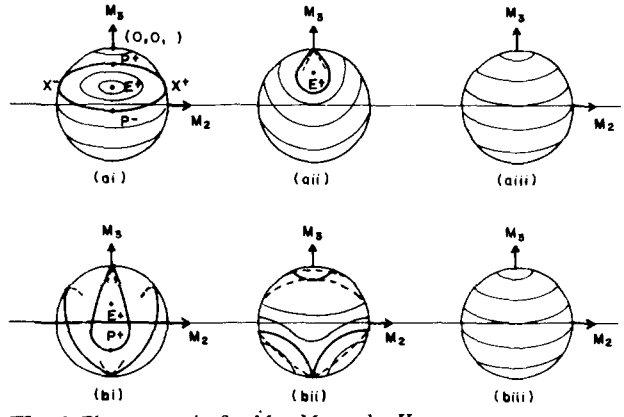


FIG. 3. Phase portraits for $\dot M = Mx\ \text{grad}_M H$.

where $\text{grad}_M f = (\partial f/\partial M_1, \partial f/\partial M_2, \partial f/\partial M_3)$.

The Hamiltonian of the system described by Fig. 2(a), when the imperfection is present, is of the form

$$H^\mu = H^0 + \mu H^1 + O(\mu^2), \tag{3.2}$$

where $H^0 = H^0(M,I)$ is as in (2.2) and $H^1 = H^1(M,\theta,I)$ is found, after some manipulations, to be given by

$$H^1 = \frac{1}{2}\left[\frac{r^2\cos^2\theta + L^2}{\alpha^2}M_1^2 + \frac{r^2\sin^2\theta + L^2}{\beta^2}M_2^2 + \frac{2r^2\sin\theta\cos\theta}{\alpha\beta} + \frac{r^2 I^2}{\gamma^2}\right]. \tag{3.3}$$

Since we are going to use a slightly different Melnikov function, it is perhaps worth reviewing Holmes–Marsden Theorem 6.4. In their notation, it is as follows:

(1) Choose and fix a value of J such that the unperturbed system has a separatrix $\bar{u}(t)$. Compute

$$\Omega(t) = \frac{\partial H^0}{\partial J}(\bar{u}(t),J),$$
$$\psi(t) = \int_0^t \Omega(t')\,dt'. \tag{3.4}$$

Assume $\Omega(t) > 0$ and $\lim_{t\to\pm\infty}\psi(t) = \pm\infty$ (this is implicitly assumed by Holmes and Marsden, although not stated explicitly).

(2) Compute the Melnikov function

$$M(\psi_0) = \int_{-\infty}^{\infty}\left\{H^0, \frac{H^1}{\Omega}\right\}dt, \tag{3.5}$$

where $\{H^0, H^1/\Omega\}$ is the u-Poisson bracket (J and ψ kept frozen) and the integral evaluated along

$$u = \bar{u}(t),\ J,\ \psi = \psi(t) - \psi_0. \tag{3.6}$$

Suppose that $M(\psi_0)$ changes sign and all zeros are simple. Then, for $\epsilon > 0$ sufficiently small, the perturbed system $H^\epsilon = H^0 + \epsilon H^1 + 0(\epsilon^2)$ has transversal splitting of the separatrix on the energy surface $H^\epsilon = h = H^0(\bar{u},J)$.

The key idea of the reasoning of Holmes and Marsden is to consider the restriction of the system to the energy manifold $H(\bar{u}(t), J) = h$ and to eliminate the time variable, replacing it by $\psi \in (-\infty, \infty)$. The reduced system is still Hamiltonian, and it is given by

$$L^\epsilon(u,\psi) = L^0(u) + \epsilon L^1(u,\psi) + O(\epsilon^2),$$

where

$$L^0(u) = \text{the value of } J \text{ given implicitly by } H^0(u,J) = h,$$
$$L^1(u,\psi) = -H(u,\psi,L^0(u))/\Omega(u,L^0(u)). \tag{3.7}$$

Then Holmes and Marsden apply the results of a previous work,[5] by which the Melnikov function for the family L^ϵ of 2-degrees-of-freedom systems is given by

$$M(\psi_0) = \int_{-\infty}^{\infty}\{L^0, L^1\}_{\psi-\psi_0}\,d\psi, \tag{3.8}$$

where the u-Poisson bracket is computed with ψ frozen and the integral evaluated along the unperturbed separatrix $\bar{u}(\psi)$ and with $\psi \to \psi - \psi_0$. Notice that $M(\psi_0)$ is 2π-periodic with respect to ψ_0.

The geometric interpretation is outlined in Fig. 4. The homoclinic surface of the suspended unperturbed ($\epsilon = 0$) system in (u,ψ) space splits for $\epsilon \neq 0$ into stable W^s and unstable W^u manifolds of hyperbolic fixed points $u^\epsilon_{\pm\infty}$ close to $u^0_{\pm\infty}$. Fix two sections, say $\psi = 0$ and $\psi = 2\pi$. The periodicity implies that the surfaces W^s and W^u intersect the sections in the same pair of curves C^s and C^u. Under the Poincaré map $P^\epsilon_{2\pi}$ a point in $W^s \cap \{\psi = 0\}$ moves to a point in $W^s \cap \{\psi = 2\pi\}$ closer to $u^\epsilon_{+\infty}$ along the curve C^s. By a classical result in dynamical systems,[3] the Poincaré map $P^\epsilon_{2\pi}$: $\{(u,\psi = 0)\} \to \{u,\psi = 2\pi)\}$ will have horseshoes if C^s and C^u intersect transversally.

The Melnikov function measures the gap between C^s and C^u along the normal lines of the unperturbed separatrix at the points $\bar{u}(\psi_0)$, $-\infty < \psi_0 < \infty$, using the level lines of L^0 to compute "distance." More precisely, $M(\psi_0)$ is the coefficient of the first term in Taylor expansion by powers of ϵ of the function $L^0(u_1) - L^0(u_2)$.

Holmes and Marsden showed that (Proposition 6.3)

$$\{L^0, L^1\} = \{H^0, H^1/\Omega\}(1/\Omega)$$

so that, when integration is changed from ψ to t variable, one gets indeed formula (3.5).

We would like to make the following observations:

(i) The presence of the factor Ω in the denominator inside the Poisson bracket is a nuisance, both for practical computations and for possible extensions to more degrees of freedom. On the other hand, using ψ_0 as the independent variable has the advantage of making $M(\psi_0)$ a periodic function. Here ψ_0 corresponds to the time value t_0 such that

$$\psi_0 = \int_0^{t_0}\Omega(t)\,dt \tag{3.9}$$

so it represents merely a regular reparametrization ($\Omega > 0$) of the separatrix $\bar{u}(t_0)$, $-\infty < t_0 < \infty$. Note that $\psi = \psi(t) - \psi_0$ appearing in (3.5) can also be written

$$\psi = \int_{t_0}^{t}\Omega(t)dt. \tag{3.10}$$

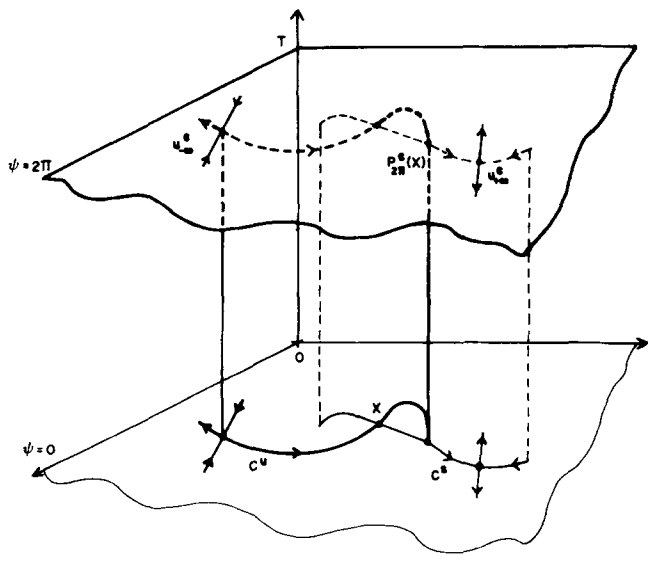


FIG. 4. Geometric interpretation of the Melnikov function.

If (3.5) is evaluated with ψ given by (3.10), then the Melnikov function will appear with the "correct" parametrization $M(t_0)$, but it will no longer be periodic in t_0 (unless Ω is constant), so we believe this will bring no advantage.

(ii) In practice, it is often possible to identify the zeros of $M(\psi_0)$ without having to compute (3.5) explicitly. Then it suffices to check whether $dM/d\psi_0 \neq 0$ whenever $M=0$. We remark that the same proof of (3.5) yields

$$\frac{dM^k}{d\psi_0^k} = \int_{t=-\infty}^{\infty} \left\{ H^0, \frac{\partial^k H^1/\partial\psi^k}{\Omega} \right\} dt, \tag{3.11}$$

where the integral is evaluated in the same manner as indicated before.

(iii) We suggest a slight modification of the Melnikov function, using the level values of $H^0(u,J)$ instead of $L^0(u,h)$. This is natural because after all H^0 is the original Hamiltonian. Following Holmes and Marsden, the crucial step consistsofshowingthat $\{H^0,L^0\} \equiv 0$ in $\{H^0, L^0 + \epsilon L^1 + O(\epsilon^2)\}$, which is indeed the case:

$$0 \equiv \{h, L^0(u)\} = \{H^0(u, L^0(u)), L^0(u)\}$$
$$= \{H^0, L^0\} + \frac{\partial H^0}{\partial J}\{L^0, L^0\}.$$

Since the second term vanishes, so does the first. Now, Proposition (6.3) of Holmes and Marsden produces the equivalent Melnikov function

$$N(\psi_0) = \int_{\psi=-\infty}^{\infty} \left\{ H^0, \frac{H^1}{\Omega} \right\} d\psi$$
$$= \int_{t=-\infty}^{\infty} \left\{ H^0, \frac{H^1}{\Omega} \right\} \Omega\, dt$$
$$= \int_{-\infty}^{\infty} \{H^0, H^1\}\, dt - \int_{-\infty}^{\infty} \frac{\{H^0, \Omega\}}{\Omega} H_1\, dt$$

and

$$\frac{dN^k}{d\psi_0^k}(\psi_0) = \int_{-\infty}^{\infty} \left\{ H^0, \frac{\partial^k H^1}{\partial\psi^k} \right\} dt$$
$$- \int_{-\infty}^{\infty} \frac{\{H^0, \Omega\}}{\Omega} \frac{\partial^k H_1}{\partial\psi^k}\, dt. \tag{3.12}$$

4. EXISTENCE OF HORSESHOES

It is now a tedious, but straightforward, task to write down the Melnikov integral (3.12) with H^0 and H^1 given by (2.2) and (3.3).

For definiteness, we consider the case $\alpha > \beta > \gamma$ in Fig. 3(a). Choosing the initial point $M(0)$ of the separatrix at P, then the functions $M_1(t)$ and $M_3(t)$ are even, while $M_2(t)$ is odd.

After the simplifications which result from dropping the odd integrands, we get

$$N(\theta^0) = r^2 A \sin(2\theta_0), \tag{4.1a}$$

where

$$A = \int_0^{\infty} \left[\frac{-s(2q)M_1}{\alpha^2} + \frac{c(2q)M_2}{\alpha\beta} \right] \times \left[\frac{M_2(M_3 - I)}{\gamma} - \frac{M_2 M_3}{\beta} \right] dt$$
$$+ \int_0^{\infty} \left[\frac{s(2q)M_2}{\beta^2} + \frac{c(2q)M_1}{\alpha\beta} \right] \times \left[\frac{M_1 M_3}{\alpha} - \frac{M_1(M_3 - I)}{\gamma} \right] dt + \frac{1}{2\gamma}\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$$
$$\times \int_0^{\infty} \frac{M_1 M_2 [2M_1 M_2 c(2q)/\alpha\beta + (M_2^2/\beta^2 - M_1^2/\alpha^2)s(2q)]\, dt}{I(1/\gamma + 1/\delta) - M_3/\gamma}. \tag{4.1b}$$

Here

$$\Omega(t) = I\left(\frac{1}{\gamma} + \frac{1}{\delta}\right) - \frac{M_3}{\gamma},$$
$$q(t) = \int_0^t \Omega(t')\, dt' = I\left(\frac{1}{\gamma} + \frac{1}{\delta}\right)t - \frac{1}{\gamma}\int_0^t M_3(t') + dt'. \tag{4.1c}$$

It would be nice if one could prove without effort that $A \neq 0$ for all values but an analytical hypersurface on the space of parameters $J_1, J_2, J_3, K, K_3, m, L, I$. Then the very form of expression (4.1a) shows that the zeros of the Melnikov function are simple, and we would be done.

At any rate, the numerical calculation of (4.1b) for given values of the parameters is always possible, since formulas for M_1, M_2, M_3 can be given in terms of elementary functions. Moreover, using M_3 as variable of integration, we have an (improper) integral over a finite interval.

In principle, the integral could be evaluated by the method of residues, but it seem to us hopeless attempting to do it even with the present capabilities of symbolic computer manipulations.

It would be very disappointing to stop here, so we make a compromise; we are able to show that indeed $A \neq 0$ in a limit case. This is sufficient to guarantee that the analytic function $A\ (J_1, J_2, J_3, K, K_3, m, L, I) \neq 0$ for a generic choice of the parameters. It turns out that the computations, although laborious, are possible and elementary in the limit case. We summarize the result in the following.

Theorem: Choose α, β, γ such that $(1-\gamma/\beta)\times(1-\gamma/\alpha)=\frac{1}{4}$.

Let $I\to 0^-$ and $\delta\to 0$ in such a way that $\omega' = I/\delta$ remains constant. Set $\omega = 2\omega'/(\sqrt{a_1 a_3}\, l)$, where $a_1 = (1/\gamma - 1/\beta) > 0$, $a_2 = (1/\alpha - 1/\gamma) < 0$, $a_3 = (1/\beta - 1/\alpha) > 0$. Then, as $\omega'\to 0^-$, $A \sim k/\sinh(\omega\pi/2) \to -\infty$ where $k = (5\pi/64) l^2 (1-\gamma/\beta)/\alpha\beta$.

The proof, as we said, is a lengthy calculation. We just indicate here the key features.

First, as $I\to 0$, the phase portrait approaches the usual rigid body system, and we can, in the limit, use the well-known formulas for the separatrices,

$$M_1(t) = l\sqrt{a_1/(-a_2)}\,\operatorname{sech}(-\sqrt{a_1 a_3}\, l t),$$
$$M_2 = l\tanh(-\sqrt{a_1 a_3}\, l t), \quad M_3 = \sqrt{a_3/\alpha_1} M_1. \tag{4.2}$$

Secondly, the function $q(t)$ in (4.1) becomes

$$q(t) = \omega' t - \frac{1}{\sqrt{(1-\gamma/\beta)(1-\gamma/\alpha)}} \tan^{-1}(\sin h(\sqrt{a_1 a_3}\, l t)). \tag{4.3}$$

The choice of $(1-\gamma/\beta)(1-\gamma/\alpha) = 1/4$ allows us to simplify the expressions $\sin(2q)$ and $\cos(2q)$ in (4.1) through the trigonometrical formulas for arc sums and the identities

$$\cos(2\tan^{-1}(\sinh x)) = (1-\sinh^2 x)\operatorname{sech}^2 x,$$
$$\sin(2\tan^{-1}(\sinh x)) = 2\sinh x \operatorname{sech}^2 x .$$

The rest of the issue consists in computing the integrals by residues. Details will appear elsewhere.[15]

How is the result to be interpreted? Physically, one always has $\delta = K_3 \neq 0$, and so, as $I\to 0^-$, then $\omega\to 0^-$. Since A gives, roughly, the "size" of the horseshoe (in the sense of Ref. 5, Appendix B), then the smaller $|\dot\theta_0|$ is taken, $\dot\theta < 0$, the more visible would become the horseshoe. One could be tempted to call these objects "wild" horseshoes.

But this runs against our intuition, since as $I\to 0$ the influence of the flywheel disappears.

We believe that the explanation for this "paradox" lies in the fact that, as $\omega\to 0$, the time needed for a complete turn of the flywheel becomes bigger and bigger, and one would have to be very patient to observe the chaotic motions. (In other words, either the horses are wild, but do not have shoes, or they have horseshoes, but are tame...). The mathematician, using the angle θ_0 as a clock, is fooled by his own trick.

5. CONCLUSIONS

We have applied the technique of Holmes and Marsden to the example of the rigid body with an attachment. A small imperfection in the flywheel destroys the integrability of the system, as indicated by the presence of horseshoes in its dynamics.

It is convenient to place the flywheel along a principal axis of the rigid body, so that the separatrices of the unperturbed system are still symmetric with respect to a plane. Then the Melnikov function has the simple form $M(\theta_0) = A\sin(2\theta_0)$. We were able to compute A, exactly, in a limit case. In general, A can be evaluated numerically for each set of values of the parameters.

To finalize, we would like to formulate some related questions that one could try to pursue.

(1) Is it possible to guarantee the splitting of separatrices by a general "abstract non-sense" argument? In that case the form of $M(\theta_0)$ shows automatically that $A \neq 0$, and therefore the splitting is transversal. An indication is the fact that the stable and unstable manifolds always have points in common, by Weinstein's transversal Lagrange manifolds theory.[16]

(2) What is the role played by the symmetry of the separatrix in the process of splitting? In this example, the center of symmetry of the separatrix is a point of transversal intersection. Is this a general fact?

(3) What features would be revealed by a detailed numerical study of the behavior of the amplitude A as parameters vary? Particularly interesting should be the bifurcation values I_{crit} of the unperturbed system. What happens when I becomes > 0? What happens near the values of I where the homoclinic orbit is disappearing?

(4) Is it possible to show the existence of Arnold diffusion[8] when two flywheels are attached?

(5) Holmes and Marsden[5] derived a relationship between the number of iterates of the Poincaré map P necessary to guarantee that P^N has a horseshoe and the perturbation parameter. Are the effects of $A\to\infty$ and $\omega\to 0$ compensating each other, so that one can observe the chaotic motions in reasonable time scale?

We would like to make three final comments. First, as pointed out by Holmes and Marsden,[8] the rigid body with a flywheel has formally the same equations as Euler's elastica, a result due to Kirchhoff.[17] It could be interesting to examine in more detail the symmetry breaking process for the elastica. Secondly, that the model has been used in technological applications: Professor Holmes called our attention to the work of Hubert.[18] Thirdly, some results in the direction of problem (2) were obtained by Carvalho.[19]

ACKNOWLEDGMENTS

This work was done while on a postdoctoral visit to the Department of Mathematics, University of California at Berkeley, from March 1982 to February 1983. We would like to thank Professor Marsden for encouraging us to work out the example. We thank Professor Alan Weinstein for sponsoring our visit and for many conversations. We also thank all of the participants of J. Marsden's seminar on Bifurcations and Hamiltonian systems during the fall quarter and for the friendly and creative atmosphere. Last, but not least, we thank all the wonderful people of Berkeley, particularly the kids, parents, and teachers of the Bancroft nursery school cooperative. This work was financially supported by the Brazilian agency CNPq and by the Federal University of Rio de Janeiro.

[1] See, for example, the annual workshops, "Dynamics days," at the La Jolla Institute and any issue of *Nonlinear Science Abstracts/Physica D.*

[2] H. Poincaré, *Les Methodes Nouvelles de la Mecanique Celeste* (Paris,

1899), (final chapter); G. D. Birkhoff, *Dynamical Systems* (Am. Math. Soc., Providence, RI, 1966).
[3]S. Smale, *The Mathematics of Time* (Springer–Verlag, Berlin, 1980), pp. 25–8, 149–50.
[4]V. K. Melnikov, "On the stability of the center for time periodic perturbations," Trans. Moscow Math. Soc. **12**, 1 (1963); V. I. Arnold, "Instability of dynamical systems with several degrees of freedom," Dokl. Akad. Nauk SSSR **156**, 9 (1964).
[5]P. J. Holmes and J. E. Marsden, "Horseshoes in perturbations of Hamiltonian systems with two degrees of freedom," Commun. Math. Phys. **82**, 533–44 (1982). See, in particular, Appendix B.
[6]P. J. Holmes and J. E. Marsden, "Melnikov's method and Arnold diffusion for perturbations of integrable Hamiltonian systems," J. Math. Phys. **23**(4), 669 (1982).
[7]P. J. Holmes and J. E. Marsden, "A partial differential equation with infinitely many periodic orbits: chaotic oscillations of a forced beam," Arch. Ratl. Mech. Anal. **76**(2), 135–66 (1981).
[8]P. J. Holmes and J. E. Marsden, "Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups," Indiana Univ. Math. J. **32**, 273–310 (1983).
[9]J. Wittenburg, Dynamics of multibody systems, in *IUTAM Conference Proceedings*, edited by F. Rimrott and B. Tabarrok, (North-Holland, Amsterdam, 1980); L. Lilov and M. Lorev, "Dynamic Analysis of Multirigid-Body system based on the Gauss principle," Z. Angew. Math. Mech. **62**, 539–45 (1982).
[10]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978), see Chap. 6.
[11]J. Wittenburg, *Dynamics of Systems of Rigid Bodies* (B. G. Teubner, Stuttgart, 1977), see Sec. 4.6.
[12]J. E. Marsden, T. Ratiu, and A. Weinstein, "Semi-direct products and reduction in mechanics," Preprint PAM-96, University of California, Berkeley, 1982.
[13]W. D. MacMillan, *Dynamics of Rigid Bodies* (Dover, New York, 1960).
[14]S. V. Manakov, "Remark on integration of Euler's dynamic equations for an n-dimensional solid," Funct. Anal. Appl. **10**(4), 93–4 (1976); T. Ratiu,"Euler-Poisson equations on Lie algebras and the N dimensional heavy rigid body," Proc. Natl. Acad. Sci. USA **78**, 1329 (1981), and thesis, University of California, Berkeley, 1980.
[15]J. Koiller, "A mechanical system with chaotic motions," Technical report IM-UFRJ, Universidade Federal de Rio de Janiero.
[16]A. Weinstein, "Lagrangian submanifolds and Hamiltonian systems," Ann. Math. **98**, 377 (1973).
[17]G. R. Kirchhoff, *Vorlesungen über Mathematische Physik* (Teubner, Leipzig, 1896), Vol. 1.
[18]C. H. Hubert, "An attitude acquisition technique for dual-spin spacecraft," Ph.D. thesis, Cornell University, 1980.
[19]S. P. Carvalho, "Orbites heterocliniques des equations differentielles du second ordre," thesis, Université de Dijon.

Solutions of a fourth-order differential equation describing mode conversion in a magnetized plasma

J. E. Maggs
I.G.P.P., UCLA, Los Angeles, California 90024

Alfredo Baños, Jr. and G. J. Morales
Physics Department, UCLA, Los Angeles, California 90024

(Received 1 June 1983; accepted for publication 9 December 1983)

Solutions of the equation $\epsilon^2 u^{(4)} + z u^{(2)} + \alpha u^{(1)} + \beta_1 u - \beta_2^2 z u = 0$, where α, β_1, β_2^2 are constants, $\epsilon^2 \ll 1$, and z is the independent variable, are obtained using the Laplace integral technique. This equation describes the propagation of high frequency electrostatic waves near plasma resonance in a magnetized plasma with a longitudinal density gradient and is a generalization of an equation studied by Wasow and by Rabenstein in the context of boundary layer phenomena. The solutions of this fourth-order equation in which the associated second-order equation (i.e., $\epsilon^2 = 0$) exhibits both a singularity (at $z = 0$) and a turning point (at $z = \beta_1/\beta_2^2$) fall readily into two classes. One class resembles Airy functions and exists only for ϵ^2 not equal to zero. In the other class, the solutions are related to confluent hypergeometric functions and can be viewed as solutions of the second-order equation with small corrections proportional to ϵ^2. Using the integral representations of solutions, it is demonstrated that each class of solutions can generate the other when the independent variable crosses the singular point. This is the physical phenomenon of mode conversion. Asymptotic descriptions of both classes of solutions are given and the form of the solutions near the singular point is expressed as a power series.

PACS numbers: 52.35.Fp, 02.30. + g

1. INTRODUCTION

This paper investigates the solutions of a fourth-order differential equation which arises in the description of high frequency electrostatic waves near plasma resonance in a magnetized plasma with a zero-order density gradient along the magnetic field. The behavior of the electrostatic potential is described by Poisson's equation which can be written as

$$\nabla\cdot\boldsymbol{\kappa}\cdot\nabla\phi = i(\omega/c)\nabla\cdot\boldsymbol{\kappa}\cdot\mathbf{A}\,, \tag{1.1}$$

where $\boldsymbol{\kappa}$ is the plasma dielectric tensor, ϕ the electric potential, $\mathbf{A}$ the vector potential, ω the angular frequency of oscillation, and c the speed of light. A harmonic time dependence of the form $\exp(-i\omega t)$ is assumed. In (1.1) the term containing the vector potential can be viewed as a driving source term, which can be physically identified with an externally launched electromagnetic wave, as might be the case in a laboratory or ionospheric experiment. From this point of view, (1.1) can be solved for interesting physical applications by obtaining the appropriate Green's function, a task which requires knowledge of the solutions of the associated homogeneous equation

$$\nabla\cdot\boldsymbol{\kappa}\cdot\nabla\phi = 0\,. \tag{1.2}$$

When thermal corrections associated with the motion of plasma along the magnetic field are retained, the plasma dielectric tensor becomes a second-order differential operator and (1.2) can be written in dimensionless form as[1]

$$\epsilon^2 u^{(4)} + z u^{(2)} + \alpha u^{(1)} + \beta_1 u - \beta_2^2 z u = 0\,. \tag{1.3}$$

To obtain (1.3), it is assumed that the plasma has a linear density gradient with scale length L along the magnetic field direction. This assumption is appropriate for many physical applications and retains the important physical processes. In (1.3), u is the electric potential suitably normalized, and z is the distance along the magnetic field normalized to the density scale length L. The small parameter

$$\epsilon^2 = (k_D L)^{-2}\,, \tag{1.4}$$

in which k_D is the Debye wave number. For typical ionospheric plasmas the parameter ϵ^2 can be less than 10^{-6}. The other parameters in the plasma have the values

$$\alpha = 1\,, \tag{1.5}$$

$$\beta_1 = (k_\perp L)^2 \Omega_e^2/(\omega^2 - \Omega_e^2)\,, \tag{1.6}$$

$$\beta_2^2 = (\omega/\Omega_e)^2 \beta_1\,, \tag{1.7}$$

where Ω_e is the electron cyclotron angular frequency and $k_\perp$ is the fixed wave number perpendicular to the magnetic field. In obtaining solutions of (1.3) we do not restrict ourselves to the parameter values given in (1.5)–(1.7). We do, however, assume that all parameters are real and that ϵ^2, β_1, and β_2 are positive. These assumptions apply to a plasma in which the wave frequency is larger than the electron gyrofrequency (i.e., $\omega > \Omega_e$).

Equation (1.3) supports two distinct classes of solutions: thermal modes and cold modes. The first of these classes represents short wavelength modes in the sense that these solutions exist only when ϵ^2 (and hence k_D^{-2}) is not zero. The prototype equation for this class is obtained from (1.3) by setting β_1 and β_2^2 equal to zero,

$$\epsilon^2 u^{(4)} + z u^{(2)} + \alpha u^{(1)} = 0\,. \tag{1.8}$$

Equation (1.8) approximates (1.3) whenever the term $\epsilon^2 u^{(4)}$ is large in comparison with $(\beta_1 u - \beta_2^2 z u)$. In this situation the solutions of (1.3) can be obtained from the solutions of (1.8)

 0022-2488/84/051605-14$02.50

by adding corrections proportional to β_1 and β_2^2. In the WKB sense the thermal class solutions are short wavelength because the term $\epsilon^2 u^{(4)}$ in (1.3) can be large in comparison with $\beta_1 u$ and $\beta_2^2 z u$. As shown in Sec. 3, the solutions of (1.8) are related to Airy functions since they are proportional to the $(\alpha-2)$ derivative of Airy functions of negative argument when α is an integer.

The second class of solutions associated with (1.3) is comprised of cold plasma modes in the sense that these solutions exist even when ϵ^2 (and hence k_D^{-2}) is equal to zero. The prototype equation for the cold mode class is obtained from (1.3) by setting $\epsilon^2 = 0$,

$$z u^{(2)} + \alpha u^{(1)} + \beta_1 u - \beta_2^2 z u = 0 . \tag{1.9}$$

The solutions of (1.9) are related to confluent hypergeometric functions. When the term $\epsilon^2 u^{(4)}$ is small, solutions of (1.3) can be obtained from solutions of (1.9) by adding small corrections proportional to ϵ^2. In the WKB sense the solutions in the cold mode class are long wavelength in that $\epsilon^2 u^{(4)}$ is small in comparison with the other terms in (1.3).

Solutions in the two classes are generally distinguished by the disparity in their wavelengths. However, near plasma resonance ($z = 0$), the WKB wavelengths of the two modes become comparable and mode conversion occurs. That is, solutions of one class generate solutions of the other class. This mode conversion process is clearly exhibited by solutions obtained in this study.

The second-order Eq. (1.9) obtained from (1.3) by setting $\epsilon^2 = 0$ exhibits both a singularity at $z = 0$ and a turning point at $z = \beta_1/\beta_2^2$. The existence of the turning point distinguishes (1.3) from an equation previously studied by Wasow[2] and by Rabenstein.[3] The results obtained by these authors can be recovered in the limit $\beta_2 \to 0$. In the plasma application the singularity corresponds to plasma resonance and the turning point to upper hybrid resonance. As shown in Sec. 5, the existence of the turning point profoundly affects the structure of the cold plasma modes and allows for the existence of solutions which exhibit no mode conversion for certain restricted parameter values.

The paper is organized as follows. In Sec. 2 we introduce integral representations for the solutions of (1.3) and describe the contours associated with the solution set. In Sec. 3 the solutions corresponding to the thermal modes are obtained. In Sec. 4 expressions for the solutions corresponding to the cold plasma modes are derived. In Sec. 5, the mode conversion process and the linear independence of the solutions is discussed. Finally, in Sec. 6 the principal findings are summarized.

2. GENERAL PROPERTIES OF SOLUTIONS

Since all of the coefficients of u, and its derivatives in (1.3) are linear in the independent variable z, general solutions may be found in the form of Laplace integrals. Following Coddington and Levinson,[4] but using the kernel $\exp(-sz)$ in the Laplace integral instead of $\exp(sz)$, we obtain solutions of (1.3) in the form

$$u(\alpha,\beta_1,\beta_2,\epsilon^2,z) = \int_C \frac{e^{-sz}}{P(s)} \exp\left[-\int^s \frac{Q(t)}{P(t)}\, dt \right] ds , \tag{2.1}$$

where

$$P(t) = t^2 - \beta_2^2 , \tag{2.2}$$

$$Q(t) = \epsilon^2 t^4 - \alpha t + \beta_1 . \tag{2.3}$$

The contours C in (2.1) are chosen such that, at the end points of the contours, the following condition is satisfied:

$$e^{-sz} \exp\left[-\int^s \frac{Q(t)}{P(t)}\, dt \right] = 0 . \tag{2.4}$$

Using (2.2) and (2.3), we obtain

$$\begin{aligned} \int^s \frac{Q}{P}\, dt &= \int^s \left[\frac{\epsilon^2 t^4 - \alpha t + \beta_1}{t^2 - \beta_2^2} \right] dt \\ &= \epsilon^2 \left(\frac{s^3}{3} + \beta_2^2 s \right) - \frac{\alpha}{2} \ln(s^2 - \beta_2^2) \\ &\quad + \beta \ln\left[\frac{(s-\beta_2)}{(s+\beta_2)} \right] , \end{aligned} \tag{2.5}$$

where

$$\beta = (\epsilon^2 \beta_2^3 + \beta_1/\beta_2)/2 . \tag{2.6}$$

Inserting (2.5) into (2.1), the solutions can be written as

$$u = \int_C (s+\beta_2)^{\alpha_+} (s-\beta_2)^{\alpha_-} \exp\left[\frac{-\epsilon^2 s^3}{3} - s\tilde{z} \right] ds , \tag{2.7}$$

where

$$\tilde{z} = z + \epsilon^2 \beta_2^2 , \tag{2.8}$$

$$\alpha_+ = \alpha/2 - 1 + \beta , \tag{2.9}$$

$$\alpha_- = \alpha/2 - 1 - \beta . \tag{2.10}$$

The end points of the contours C are chosen to satisfy the condition

$$(s+\beta_2)^{\alpha/2+\beta} (s-\beta_2)^{\alpha/2-\beta} \exp[-\epsilon^2 s^3/3 - s\tilde{z}] = 0 . \tag{2.11}$$

It is worth noting that the basic Eq. (1.3) is invariant under the transformation

$$(\alpha,\beta_1,\beta_2,\epsilon^2,z) \to (\alpha, -\beta_1, \beta_2, -\epsilon^2, -z) . \tag{2.12}$$

Thus another family of solutions of (1.3) is given by

$$\begin{aligned} &\phi(\alpha,\beta_1,\beta_2,\epsilon^2,z) \\ &\quad = u(\alpha, -\beta_1, \beta_2, -\epsilon^2, -z) \\ &\quad = \int_{C'} (s+\beta_2)^{\alpha_-} (s-\beta_2)^{\alpha_+} \exp\left[\frac{\epsilon^2 s^3}{3} + s\tilde{z} \right] ds , \end{aligned} \tag{2.13}$$

where the end points of the contours C' satisfy the condition

$$(s+\beta_2)^{\alpha/2-\beta} (s-\beta_2)^{\alpha/2+\beta} \exp[\epsilon^2 s^3/3 + s\tilde{z}] = 0 . \tag{2.14}$$

In the following we investigate the functions u as defined by (2.7) in detail and use the transformation (2.12) to obtain the functions ϕ.

The condition (2.11) that must be met at the end points of the contours can be satisfied at large s by choosing the real part of $\epsilon^2 s^3/3$ to be positive. Since we have chosen ϵ^2 real, this requirement becomes, with $j = 1,2,3$,

$$2\pi(2-j)/3 - \pi/6 < \arg s < \pi/6 + 2\pi(2-j)/3 . \tag{2.15}$$

Thus, as shown in Fig. 1, there are three open sectors in the s-plane of angular width $\pi/3$ centered about $2\pi/3$, 0, and $-2\pi/3$ in which the contour may go to infinity and satisfy

the condition (2.11). Since we have stipulated that α, β_1, β_2, and ϵ^2 are real and positive, the quantity $\alpha/2 + \beta$ is real and positive. Thus the condition (2.11) can also be satisfied at the point $s = -\beta_2$ because the quantity $(s+\beta_2)^{\alpha/2+\beta}$ is then always zero. Since the quantities α_+ and α_-, while real, are not necessarily integers, the points $s = \pm\beta_2$ are, in general, branch points. If the s-plane is cut from β_2 to positive infinity and from $-\beta_2$ to negative infinity along the real axis as shown in Fig. 1, the multivalued function $(s-\beta_2)^{\alpha_-}(s+\beta_2)^{\alpha_+}$ in the integrand of (2.7) can be written as

$$(s-\beta_2)^{\alpha_-}(s+\beta_2)^{\alpha_+} = (r_1 e^{i\theta_1})^{\alpha_-}(r_2 e^{i\theta_2})^{\alpha_+}$$
$$= \exp[\alpha_- \ln r_1 + i\alpha_-(\theta_1 + 2\pi n) + \alpha_+ \ln r_2 + i\alpha_+(\theta_2 + 2\pi m)], \quad (2.16)$$

with $-\pi < \theta_2 < \pi$ and $0 < \theta_1 < 2\pi$. In (2.16) the various branches of the ln functions are represented by the integers m and n, which denote various sheets of the Riemann surface. Each sheet of the Riemann surface can thus be labeled by the pair of integers (m,n). The s-plane represented in Fig. 1 corresponds to the principal branches of the ln functions $m = n = 0$ and is labeled by (0,0).

The end points for contours corresponding to solutions of (1.3) must be chosen to satisfy the relation given in (2.11). This condition can be met by choosing the contours that begin and end at infinity in any of the sectors $j = 1,2,3$ or contours that begin at the branch point $s = -\beta_2$ and end at $s = -\beta_2$ or, alternatively, proceed to infinity within the numbered sectors. Contours which correspond to solutions in the thermal mode class are shown in Fig. 2. These contours begin at infinity in one sector and end at infinity in another sector. The solutions obtained from (2.7) by integrating along these contours are labeled A_j, where j refers to the sector opposite the contour. For example, the contour beginning in sector 2 and ending in sector 1 corresponds to the solution labeled A_3. The contour for the solution A_2 begins in sector 1 and crosses the branch cut along the negative real axis before proceeding to infinity in sector 3. Thus part of the contour for the solution A_2 lies in the (1,0) Riemann sheet (i.e., the sheet with $m = 1$, $n = 0$) and is thus shown as a broken line in Fig. 2.

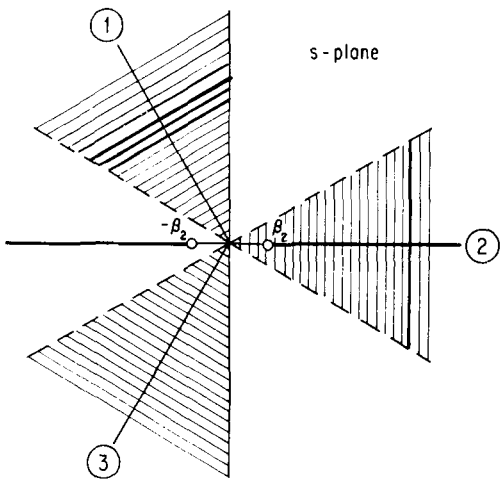


FIG. 1. Contours corresponding to solutions of (1.3) must proceed to infinity in the shaded open sectors labeled 1, 2, or 3. The integration plane is cut along the real axis from β_2 to positive infinity and from $-\beta_2$ to negative infinity. The plane of the paper represents the principal sheet with $m = 0$, $n = 0$ in (2.16).

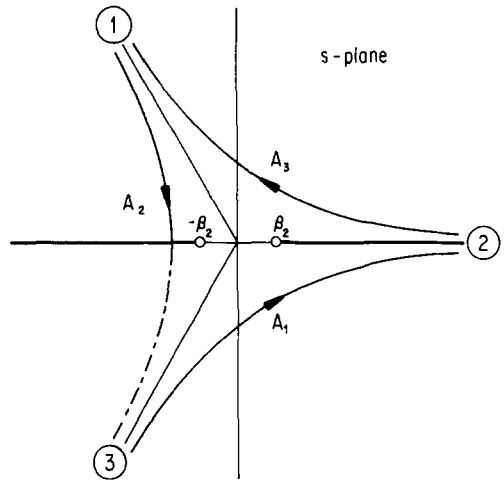
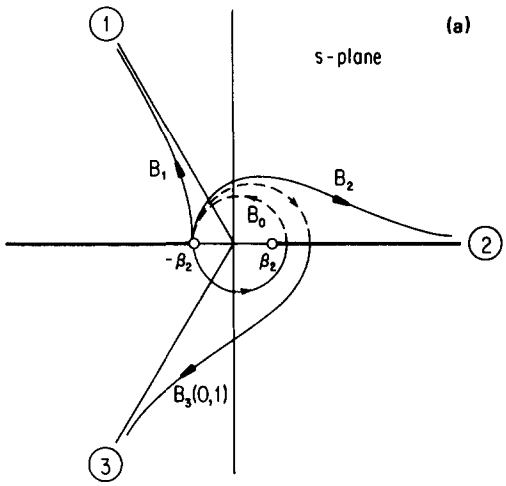


FIG. 2. The contours for the solutions A_1, A_2, and A_3 begin and end at infinity. The contour for the solution A_2 passes onto the adjacent Riemann sheet with $m = 1$, $n = 0$ in (2.16) when crossing the branch cut.

Contours corresponding to solutions in the cold plasma mode class have at least one end point at $s = -\beta_2$ and are shown in Figs. 3(a) and 3(b). The contour corresponding to the solution labeled B_0 starts at the point $s = -\beta_2$, encircles the point $s = \beta_2$ in the counterclockwise direction, and ends again at $s = -\beta_2$. The contour for B_0 crosses the branch line along the positive real axis and thus crosses onto the (0,1) Riemann sheet (i.e., $m = 0$, $n = 1$). The portion of the contour lying in the (0,1) sheet is shown as a dashed line. Contours corresponding to solutions labeled B_j where j corresponds to the sector in which they proceed to infinity are illustrated in Fig. 3(a). These contours start at the point $s = -\beta_2$ and proceed to infinity in sector j in such a fashion that the radius vector from $s = \beta_2$ to a point on the contour moves in the *clockwise* direction as the point proceeds to infinity. Thus the contour for the solution β_2 passes above the point $s = \beta_2$. Contours for solutions labeled $\tilde{B}_j$ are shown in Fig. 3(b). These contours also start at $s = -\beta_2$, but the radius vector from $s = \beta_2$ to a point on the contour moves in the *counterclockwise* direction as the point proceeds to infinity in sector j. Thus the contour for $\tilde{B}_2$ passes below the point $s = \beta_2$. Note that the contour for $\tilde{B}_1$ shown in Fig. 3(b) crosses the branch cut along the positive real axis and thus passes onto the (0,1) Riemann sheet. The portion of the contour on the (0,1) sheet is again shown as a dashed line. While not illustrated in Fig. 3(a), the contour for the solution B_3 would similarly cross the branch cut along the positive real axis but in a clockwise direction and thus would pass onto the (0, − 1) Riemann sheet. The contours for the solutions B_j and $\tilde{B}_j$ ($j = 1,2,3$) proceed to infinity along the same asymptotic direction in sector j.

Although illustrated only for the principal sheet, the contours corresponding to solutions B_j, $\tilde{B}_j$, B_0, and A_j may begin on any sheet of the Riemann surface. In order to distin-

guish among functions corresponding to contours on different sheets of the Riemann surface, we introduce the notation $X(m,n;p)$, where X is any of the solutions A_j, B_0, B_j, $\tilde{B}_j$; m and n are integers specifying the sheet, p denotes dependence on the parameters $\alpha, \beta_1, \beta_2, \epsilon^2$, and z. In this notation the values m and n indicate the sheet on which the contour begins. As described above, some contours begin on one sheet and end on an adjacent sheet. Thus $A_2(0,0;p)$ indicates the contour beginning in the sheet with $m = 0$, $n = 0$, but, as shown in Fig. 2, it ends in the sheet with $m = 1$, $n = 0$. For brevity of notation we omit the dependence on m and n or other parameters unless they are needed to clarify the discussion. Furthermore, if the values of m and n are not explicitly indicated, the principal values $m = 0$, $n = 0$ are to be assumed. As an illustration, we have drawn in Fig. 3(a) the contour corresponding to the function $B_3(0,1)$. The contour begins on the sheet (0,1) (and is thus shown dashed) and passes onto the principal sheet when it crosses the branch cut



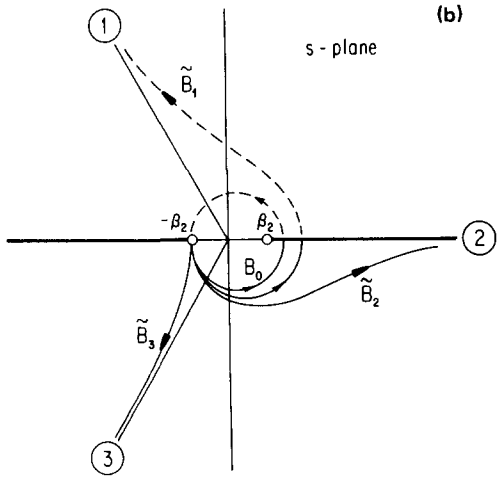


FIG. 3. (a) The contours for the solutions B_0 and B_j have at least one end point at $s = -\beta_2$. Contours that cross the branch cut pass onto the adjacent Riemann sheet on which $m = 0$ and $n = 1$ in (2.16) and are shown dashed. The contour for the solution $B_3(0,1)$ starts on the sheet (0,1) and crosses onto the principal sheet. (b) The same as (a) but for the contours corresponding to the solutions $\tilde{B}_j$.

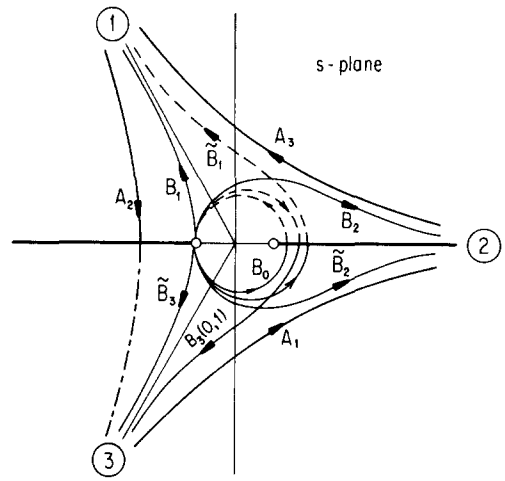


FIG. 4. Contours for all of the solutions shown in Figs. 2 and 3 are combined to facilitate the derivation of the relations (2.18)–(2.23).

extending from β_2 to positive infinity. Finally, functions corresponding to contours on different sheets are simply related. Employing the notation described above and using (2.16), one can write

$$X(m,n) = e^{i2\pi m\alpha} e^{i2\pi n\alpha} X(0,0) . \tag{2.17}$$

Since the integrand in (2.7) is analytic throughout the entire s-plane except at the points $s = \pm\beta_2$, Cauchy's theorem can be used to establish relationships among the various solutions A_j, B_0, B_j, and $\tilde{B}_j$ by using combinations of the appropriate contours. In order to facilitate the derivation of relations among the solutions all of the contours in Figs. 2 and 3 have been combined into Fig. 4. Referring to Fig. 4 and using Cauchy's theorem, the following relationships can be established

$$B_2 = B_1 - A_3 , \tag{2.18}$$

$$\tilde{B}_1 = \tilde{B}_2 + A_3(0,1) , \tag{2.19}$$

$$B_1 = \tilde{B}_3(1,0) - A_2 , \tag{2.20}$$

$$B_3(0,1) = B_2(0,1) - A_1 , \tag{2.21}$$

$$\tilde{B}_3 = \tilde{B}_2 - A_1 , \tag{2.22}$$

$$\tilde{B}_j - B_j(0,1) = B_0 \quad (j = 1,2,3) . \tag{2.23}$$

The relations (2.18), (2.20), and (2.22) can be readily verified by referring to Fig. 4. To verify the remaining relations, it is helpful to picture the contours for $\tilde{B}_2$ or $B_2(0,1)$ as crossing the branch cut which extends from β_2 to infinity. The relations (2.18)–(2.23) prove useful in obtaining, among other things, analytic continuations of the functions B_j and $\tilde{B}_j$. Having determined the set of contours which yield solutions of (1.3) in the integral representation (2.7), we next proceed to evaluate these solutions in detail. We first examine the solutions A_j before proceeding to investigate the solutions B_0, B_j, and $\tilde{B}_j$.

3. THE SOLUTIONS A_j

The functions $A_j(z)$ are defined by the integral representation (2.7) in which the paths of integration are given by

the contours $C(A_j)$ as illustrated in Fig. 2. In this section we first discuss the asymptotic behavior of the functions $A_j(z)$ for $|z| \gg \epsilon^{2/3}$ and then derive a power series expansion which is particularly useful for small argument, $z \to 0$. In both cases it is convenient to introduce the transformation

$$\sigma = \epsilon^{2/3} s\,, \quad \eta = \epsilon^{-2/3} \tilde{z}\,, \quad \sigma_0 = \sigma(\beta_2) = \epsilon^{2/3} \beta_2\,, \tag{3.1}$$

where $\tilde{z} = z + \epsilon^2 \beta_2^2 \simeq z$ when $\epsilon^2 \ll 1$. Making use of (3.1) in (2.7), we obtain the integral representation

$$A_j(\eta) = \epsilon^{-2(\alpha-1)/3} \int_{C'(A_j)} (\sigma - \sigma_0)^{\alpha_-} (\sigma + \sigma_0)^{\alpha_+} \times \exp\left[-\left(\tfrac{1}{3}\sigma^3 + \sigma\eta\right)\right] d\sigma\,, \tag{3.2}$$

where $C'(A_j)$ is the image of the contour $C(A_j)$ under the transformation (3.1).

To evaluate (3.2) asymptotically, we employ the saddle point method of integration.[5] We note that the derivative of the exponent $f(\sigma) = (\frac{1}{3}\sigma^3 + \sigma\eta)$, namely, $f'(\sigma) = -(\sigma^2 + \eta)$, when equated to zero yields two roots, or saddle points, at $\sigma_s = \pm i\eta^{1/2}$. For our purpose we discard the minus sign and choose the saddle point

$$\sigma_s = i\eta^{1/2}\,, \quad \eta = |\eta| e^{i\theta}\,, \quad |\eta| \gg 1\,, \tag{3.3}$$

from which we deduce that, when $\theta = 0$, $\sigma_s = i\eta^{1/2}$ with η real and positive; when $\theta = \pi$, $\sigma_s = e^{i\pi}|\eta|^{1/2} = -(-\eta)^{1/2}$ with η real and negative, and finally, when $\theta = 2\pi$, $\sigma_s = e^{3\pi i/2}|\eta|^{1/2} = -i\eta^{1/2}$, with η again real and positive. These values of σ_s represent the principal saddle points and their location in the σ-plane is illustrated in Fig. 5. The principal saddle points at $\theta = 0$, π, and 2π correspond, respectively, to the contours for A_3, A_2, and A_1. The other choice of sign, $\sigma_s = -i\eta^{1/2}$, yields nothing new.

The contours $C'(A_j)$ in (3.2), as shown in Fig. 5, are asymptotic to the rays with phase 0, $2\pi/3$, and $4\pi/3$. Accordingly, making use of (3.3), one has the following argument ranges:

$$\begin{aligned} &2\pi(2-j)/3 + 2\pi/3 < \arg\sigma_s < 4\pi/3 + 2\pi(2-j)/3\,, \\ &4\pi(2-j)/3 + \pi/3 < \theta < 5\pi/3 + 4\pi(2-j)/3\,. \end{aligned} \tag{3.4}$$

As shown in Fig. 5, we can use (3.4) to trace the path of the saddle point with $|\eta|$ fixed as θ varies from $-\pi$ to 3π, or $\arg\sigma_s$ varies from 0 to 2π. It is seen from Fig. 5 that the chosen saddle point σ_s in (3.3) corresponds to the functions A_3, A_2, and A_1 and traces a circle of radius $\sigma = |\eta|^{1/2}$. However, we note that σ_s starts just above the right-hand branch cut of Fig. 1 and traces a semicircle in the upper half of the σ-plane, at which point σ_s crosses the left-hand branch cut from the principal sheet onto the adjacent sheet of the Riemann surface, $m = 1$, $n = 0$ in accordance with (2.16), and traces a semicircle on the lower half of the (1,0) sheet.

We observe from (3.4) that the path of steepest descents for A_3 lies entirely on the principal sheet of the Riemann surface and therefore the saddle point integration of (3.2) is independent of the branch cuts, which is to say that the multivalued factors $(\sigma - \sigma_0)^{\alpha_-}$ and $(\sigma + \sigma_0)^{\alpha_+}$ in the integrand of (3.2) are assigned their principal values. For A_2 the path of steepest descents starts at infinity on the principal sheet and terminates at infinity on the adjacent $m = 1$, $n = 0$ sheet

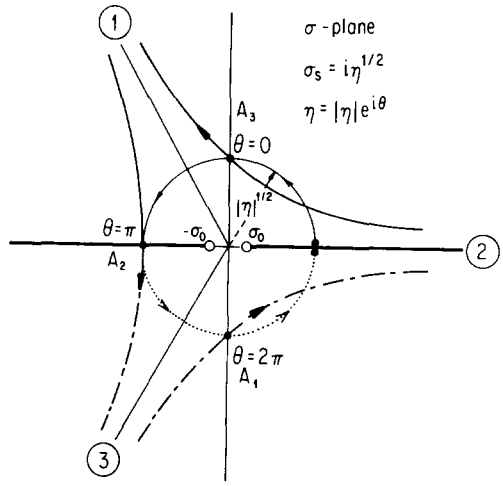


FIG. 5. The position of the saddle point used to obtain an asymptotic representation for the solutions A_j is shown as a function of $\theta = \arg(\eta)$. The saddle point starts just above the right-hand branch cut for $\theta = -\pi$ and traces out a circular path, crossing onto the adjacent $m = 1$, $n = 0$ Riemann sheet to end just below the branch cut at $\theta = 3\pi$. The contour for the solution A_1 lies in the (1,0) Riemann sheet.

thus crossing the left-hand branch cut. However, since the saddle point also crosses onto the (1,0) sheet, the saddle point integration is carried out, assuming that the multivalued factors in (3.2) vary continuously as the saddle point crosses the branch cut. Finally, as regards the path of steepest descents for A_1, we note that it is also independent of the branch but now lies entirely on the $m = 1$, $n = 0$ sheet and, therefore, the saddle point integration of (3.2) evaluates $A_1(1,0)$. The multivalued factor is evaluated according to (2.16) with $m = 1$, $n = 0$.

Next, we wish to determine the direction of traversal through the chosen saddle point σ_s. In the vicinity of σ_s we can write $-x^2 = \frac{1}{2} f''(\sigma_s)(\sigma - \sigma_s)^2 + \cdots$, where x is real and positive ($x > 0$) after passing through the saddle point. In the present instance $-\frac{1}{2} f''(\sigma_s) = \sigma_s = i\eta^{1/2}$. Hence, extracting the square root and putting $w = \sigma - \sigma_s$, we have $x = \pm\left[-\frac{1}{2} f''(\sigma_s)\right]^{1/2} w = \pm(\sigma_s)^{1/2} w$. Ignoring the plus sign, writing $-1 = e^{-i\pi}$, and putting $\arg x = 0$ yields

$$\arg w = \pi - \arg(\sigma_s)^{1/2} = 3\pi/4 - \theta/4\,, \tag{3.5}$$

which says that, for A_3 and $\theta = 0$, $\arg w = 3\pi/4$; for A_2 and $\theta = \pi$, $\arg w = \pi/2$, and, for A_1 and $\theta = 2\pi$, $\arg w = \pi/4$. We note that (3.5) gives, for A_3 and A_1, the correct direction of traversal through the saddle point in accordance with the arrows drawn in Figs. 2 and 5. However, for A_2 the direction of traversal is opposite to the direction of the arrow, which gives a minus sign to be attached to the final result for A_2.

Asymptotically, $|z| \gg \epsilon^{2/3}$ and with $\epsilon^2 \ll 1$ yields $\tilde{z} = z + \epsilon^2 \beta_2^2 \simeq z$ and $\eta \simeq \epsilon^{-2/3} z$. Furthermore, from (2.6), $\beta = \epsilon^2 \beta_2^3/2 + \beta_1/2\beta_2 \simeq \beta_1/2\beta_2 \equiv \beta_0$, from which instead of the exponents α_+ and α_- defined by (2.9) and (2.10) we introduce

$$a_+ = \alpha/2 - 1 + \beta_0\,, \quad a_- = \alpha/2 - 1 - \beta_0\,. \tag{3.6}$$

After this preamble, we can now apply the familiar leading term formula of the method of steepest descents to compute an asymptotic representation for the functions $A_j(\eta)$ given

by the integral representations (3.2). Thus we obtain, for the saddle point $\sigma_s = i\eta^{1/2}$ and $j = 1,2,3$, the leading terms:

$$A_j(\bar{\gamma},0;\eta) \sim \sqrt{\pi}(-)^j \epsilon^{-2(\alpha-1)/3} (\sigma_s - \sigma_0)^{\alpha_-} (\sigma_s + \sigma_0)^{\alpha_+} \times \{\exp[f(\sigma_s)]/[-f''(\sigma_s)/2]^{1/2}\} \{1 + O(\sigma_s^{-3})\}$$
$$= \sqrt{\pi}(-)^j \epsilon^{-2(\alpha-1)/3} (i\eta^{1/2} - \sigma_0)^{\alpha_-} (i\eta^{1/2} + \sigma_0)^{\alpha_+} \times (e^{-i\zeta}/[i\eta^{1/2}]^{1/2}) \{1 + O(1/\zeta)\}, \quad (3.7)$$

where

$$\zeta \equiv \tfrac{2}{3}\eta^{3/2} = \tfrac{2}{3}\epsilon^{-1}z^{3/2}, \quad \bar{\gamma} = \tfrac{1}{2}[1 - (-)^j]\gamma, \quad (3.8)$$

and

$$\gamma \equiv \begin{cases} 0, & \mathrm{Im}(\sigma_s) > 0, \\ 1, & \mathrm{Im}(\sigma_s) \leqslant 0. \end{cases} \quad (3.9)$$

We have introduced the factor $\bar{\gamma}$ to insure the continuity of the function A_2 as the saddle point crosses the branch cut along the negative real axis. The general solution $A_j(m,n;\eta)$ can be found from (3.7) by using (2.17).

From (3.8) we deduce that, for $j = 3$ and $\theta = 0$, $-i\zeta = -(2/3)i\eta^{3/2}$ with η real and positive; for $j = 2$ and $\theta = \pi$, $-i\zeta = -(2/3)(-\eta)^{3/2}$ with η real and negative, and finally, for $j = 1$ and $\theta = 2\pi$, $-i\zeta = (2/3)i\eta^{3/2}$ with η again real and positive. Finally, to conclude our asymptotic leading term presentation, we translate (3.7) into a function of z. Recalling that $\sigma_0 = \epsilon^{2/3}\beta_2$, we obtain, after some algebraic manipulations,

$$A_j(\bar{\gamma},0;z) = (-)^j (iz^{1/2}/\epsilon - \beta_2)^{\alpha_-} (iz^{1/2}/\epsilon + \beta_2)^{\alpha_+} \times (\sqrt{\pi} e^{-i\zeta}/[i\epsilon z^{1/2}]^{1/2}) \{1 + O(1/\zeta)\}, \quad (3.10)$$

where the multivalued factor is evaluated as discussed above.

In (3.7) we have retained the term σ_0 even though we have assumed $|\sigma_s| \gg \sigma_0$ (i.e., $|z| \gg \epsilon^2\beta_2^2$) in order to preserve the topological structure of the cut σ-plane. It is useful to obtain a form for A_j in which σ_0 is ignored in comparison with σ_s, but this procedure will alter the topology of the cut σ-plane. Ignoring the σ_0 term in (3.7) is formally equivalent to setting $\sigma_0 = 0$ so that there is a single branch point at the origin rather than two branch points. The multivalued factor in (3.7) becomes $\sigma_s^{\alpha_-}\sigma_s^{\alpha_+} = \sigma_s^{\alpha-2}$. Taking the branch cut to lie along the positive real axis and the principal sheet to have argument range $0 < \arg \sigma_s < 2\pi$, we see that the argument range for the multivalued factor $(\sigma_s - \sigma_0)^{\alpha_-}(\sigma_s + \sigma_0)^{\alpha_+}$ in (3.7) exactly coincides with that of $\sigma_s^{\alpha_-}\sigma_s^{\alpha_+}$ when $\sigma_0 = 0$. Noting from (2.17) that $A_1(1,0;\eta) = e^{i2\pi\alpha_+}A_1(\eta)$, we can then write

$$A_j(\eta) \sim (-)^j \sqrt{\pi} p_j \epsilon^{-2(\alpha-1)/3} (i\eta^{1/2})^{\alpha-5/2} e^{-i\zeta}, \quad (3.11)$$

where the phase factors p_j are

$$p_3 = p_2 = 1, \quad p_1 = e^{i2\pi\alpha_+}. \quad (3.12)$$

In (3.11) the argument range of $(i\eta^{1/2})$ is from 0 to 2π.

As mentioned previously, the functions $A_j(\eta)$ are closely related to Airy functions. This relationship is best elucidated if we consider the case $\sigma_0 = 0$, which is equivalent to setting $\beta_2 = \beta_1 = 0$, with the result that (1.3) reduces to (1.8). Putting $\eta = \epsilon^{2/3}z$ in accordance with (3.1) and writing $\chi(\eta)$ instead of $u(z)$, we obtain from (1.8) the equation $\chi^{(3)} + \eta\chi^{(1)} + (\alpha - 1)\chi = 0$. When α is an integer, it can be shown that $\chi(\eta)$ is proportional to the $(\alpha - 2)$ derivative of any Airy function of negative argument, i.e., any solution of $v^{(2)} + \eta v = 0$. When $(\alpha - 2)$ is a negative integer, the above statement must be interpreted as the $|\alpha - 2|$ integral of $v(\eta)$. The same conclusions follow from the integral representation (3.2) upon setting $\sigma_0 = 0$; that is, $A_j(\eta)$ in (3.11), when α is an integer, is proportional to the $(\alpha - 2)$ derivative of the asymptotic leading term of an Airy function of negative argument, and hence is a solution of our basic Eq. (1.3).

We now proceed to discuss the small argument behavior of the function $A_j(\eta)$. For this purpose we make use again of the scale transformation (3.1) and the integral representation (3.2) wherein α_+ and α_- have their original definitions (2.9) and (2.10). To evaluate (3.2), we first construct the power series expansion

$$(\sigma - \sigma_0)^{\alpha_-}(\sigma + \sigma_0)^{\alpha_+} = \sigma^{\alpha-2} \sum_{m=0}^{\infty} c_m w^m, \quad |w| < 1, \quad (3.13)$$

where $w = \sigma_0/\sigma$ and the expansion coefficients

$$c_m = \sum_{n=0}^{\infty} (-)^n \binom{\alpha_-}{n} \binom{\alpha_+}{m-n} \quad (3.14)$$

are given in terms of the familiar binomial coefficients. The expansion on the right of (3.13) is absolutely convergent for $|w| < 1$ because it arises from the multiplication of the two absolutely convergent binomial expansions corresponding to the factors $(\sigma - \sigma_0)^{\alpha_-}$ and $(\sigma + \sigma_0)^{\alpha_+}$. This result follows from Cauchy's theorem on the multiplication of absolutely convergent series.[6] Since the power series (3.13) is absolutely convergent for $|w| = 1 - \delta$, where δ is an arbitrarily small number, it also follows that the series in question is uniformly convergent[7] for $|w| \leqslant 1 - \delta$. Hence, we can replace the product $(\sigma - \sigma_0)^{\alpha_-}(\sigma + \sigma_0)^{\alpha_+}$ in the integrand of (3.2) with the uniformly convergent series (3.13) and integrate term by term to obtain, for $j = 1,2,3$,

$$A_j(\eta) = p_j \epsilon^{-2(\alpha-1)/3} \sum_{m=0}^{\infty} c_m \sigma_0^m g_j(\eta, \alpha - m), \quad (3.15)$$

where the phase factors p_j are given by (3.12) and where

$$g_j(\eta,\alpha) = \int_{C'(A_j)} d\sigma\, \sigma^{\alpha-2} e^{-\sigma^3/3 - \sigma\eta} \quad (3.16)$$

are the same functions introduced by Rabenstein.[8] In (3.16) the contour $C'(A_j)$ is chosen such that $|\sigma| > \sigma_0$ everywhere along the contour, which is the condition for the convergence of the series (3.13).

A convenient expression for the functions $g_j(\eta,\alpha)$, $j = 1,2,3$ is obtained by expanding the factor $\exp[-\sigma\eta]$ into a power series

$$e^{-\sigma\eta} = \sum_{k=0}^{\infty} \frac{(-)^k}{k!} \sigma^k \eta^k, \quad (3.17)$$

which is uniformly convergent in the finite σ plane, $|\sigma| < \infty$. Introducing (3.17) into (3.16) and integrating term by term, we obtain

$$g_j(\eta,\alpha) = \sum_{k=0}^{\infty} \frac{(-)^k}{k!} \eta^k \int_{C'(A_j)} d\sigma\, \sigma^{\alpha-2+k} e^{-\sigma^3/3}. \quad (3.18)$$

Putting $y = \sigma^3/3$, the integrals in (3.18) become

$$I_j(\alpha,k) = \int_{C'(A_j)} d\sigma\, \sigma^{\alpha-2+k} e^{-\sigma^3/3}$$

$$= 3^{(\alpha+k-4)/3} \int_{H_j} dy\, e^{-y} y^{(\alpha+k-4)/3}, \qquad (3.19)$$

where the contour H_j is the image of $C'(A_j)$ under the transformation. To ascertain the shape of H_j consider, for example, $C'(A_3)$ which begins at infinity with $\arg\sigma = 0$ and proceeds to infinity with $\arg\sigma = 2\pi/3$. Under the transformation, $y = \sigma^3/3$, the contour H_3 begins at infinity with $\arg y = 0$, encircles the origin in the counterclockwise direction, and proceeds to infinity with $\arg y = 2\pi$. Therefore, H_3 is simply Hankel's contour, and we can write

$$I_3(\alpha,k) = 3^{(\alpha+k-4)/3} \int_{\infty}^{(0+)} dy\, e^{-y} y^{(\alpha+k-4)/3}, \qquad (3.20)$$

which can be evaluated in terms of gamma functions.

Proceeding in an entirely analogous fashion, we ascertain that the contour H_2 starts at infinity with $\arg y = 2\pi$, encircles the origin, and terminates at infinity with $\arg y = 4\pi$, whereas the contour H_1 begins with $\arg y = 4\pi$ and ends with $\arg y = 6\pi$. Thus, introducting the abbreviation

$$\rho = \rho(\alpha,k) = (4-\alpha-k)/3, \qquad (3.21)$$

and making use of a generalization of Hankel's integral representation,[9]

$$I_j(\alpha,k) = 2\pi i \exp[\pi i(1-\rho)(7-2j)]/\{3^{\rho}\Gamma(\rho)\}. \qquad (3.22)$$

Making use of (3.18) and (3.19), we obtain the expansion

$$g_j(\eta,\alpha) = \sum_{k=0}^{\infty} \frac{(-)^k}{k!} I_j(\alpha,k)\eta^k, \qquad (3.23)$$

where the functions $I_j(\alpha,k)$ are given by (3.22). Finally, using (3.15), we obtain the power series expansions

$$A_j(\eta,\alpha)$$

$$= p_j \epsilon^{-2(\alpha-1)/3} \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} c_m \sigma_0^m \frac{(-)^k}{k!} I_j(\alpha-m,k)\eta^k, \qquad (3.24)$$

where the p_j are the phase factors given in (3.12). These functions, for $j = 1,2,3$, correspond to their asymptotic counterparts $A_j(\eta)$ given by (3.7) and are subject to the same argument ranges for η as given by (3.4).

4. THE SOLUTIONS B_j, $\tilde{B}_j$, AND B_0

The functions B_j, $\tilde{B}_j$, and B_0 are solutions of (1.3) and are given by the integral representation (2.7), where the corresponding paths of integration are shown in Figs. 3(a) or 3(b). Since we are unable to directly evaluate these integrals, we resort to a perturbation expansion based on the fact that $\epsilon^2 \ll 1$. Thus, expanding (2.7) in a power series in ϵ^2, we show in the Appendix that, to first order in ϵ^2,

$$u(\alpha,\beta_1,\beta_2,\epsilon^2;z) = u(\alpha,\beta_1,\beta_2,0;z) + \epsilon^2(\partial_{\epsilon^2} u)|_{\epsilon^2=0}$$

$$= u_0 - \epsilon^2[h(u_0)/4\beta_2], \qquad (4.1)$$

where $u(\alpha,\beta_1,\beta_2,\epsilon^2;z)$ is any of the solutions B_0, B_j, or $\tilde{B}_j$ and u_0 is the corresponding solution evaluated at $\epsilon^2 = 0$. The expression for $h(u_0)$ is given in (A15) in terms of u_0 evaluated at shifted values of the parameters α and β_1.

We observe that the functions $u_0(z)$ are solutions of the reduced ($\epsilon^2 = 0$) second-order differential Eq. (1.9) and are given by the integral representation

$$u_0(z) = \int_C (s-\beta_2)^{a_-}(s+\beta_2)^{a_+} e^{-sz}\, ds, \qquad (4.2)$$

where a_+ and a_- are given by (3.6) in terms of α and $\beta_0 = \beta_1/2\beta_2$. The contours of integration in (4.2) are those associated with B_j, $\tilde{B}_j$, and B_0 and, respectively, yield functions denoted by b_j, $\tilde{b}_j$, and b_0. Finally, if we introduce the transformation $u(x) = g(x)\exp(-x/2)$ with $x = 2\beta_2 z$ into (1.9), we obtain Kummer's differential equation[10]

$$xg^{(2)} + (\alpha - x)g^{(1)} - (\alpha/2 - \beta_0)\, g(x) = 0, \qquad (4.3)$$

which has two independent solutions, which, for our purpose, we write as

$$g_1(x) = M(\alpha/2 - \beta_0, \alpha; x), \qquad (4.4)$$

$$g_2(x) = U(\alpha/2 - \beta_0, \alpha; x), \qquad (4.5)$$

where M and U are Kummer functions. Since the properties of these functions are well known, we proceed to obtain expressions for b_j, $\tilde{b}_j$, and b_0 in terms of Kummer functions.

The function b_0 obtained from (4.2) by integrating along the contour associated with B_0 is related to the Kummer function $M(b,c;z)$. This can be demonstrated by using the transformation

$$s = 2\beta_2(t - \tfrac{1}{2}) \qquad (4.6)$$

in (4.2) to obtain

$$b_0 = (2\beta_2)^{\alpha-1} e^{\beta_2 z} \int_0^{(1+)} (t-1)^{a_-} t^{a_+} e^{-2\beta_2 zt}\, dt. \qquad (4.7)$$

The contour in (4.7) is the image of $C(B_0)$ under the transformation (4.6). It starts at the origin, circles the point $t = 1$ in the positive direction, and returns to the origin. The integrand in (4.7) is bounded everywhere along the contour for $|z| < \infty$ so that an expression can be obtained for b_0 that is valid for the entire z-plane, as is shown in the following.

The integral representation for the Kummer function $M(b,c;z)$ for $\mathrm{Re}(b) > 0$ is[11]

$$M(b,c;z)$$

$$= \frac{\Gamma(c)\Gamma(b-c+1)}{2\pi i \Gamma(b)} \int_0^{(1+)} e^{zt} t^{b-1} (t-1)^{c-b-1}\, dt. \qquad (4.8)$$

From (3.6) one has $a_+ = \alpha/2 + \beta_0 - 1$, where $\beta_0 = \beta_1/2\beta_2$ is β evaluated at $\epsilon^2 = 0$, and, since α, β_1, β_2 are taken as real and positive, $\mathrm{Re}(\alpha/2 + \beta_0) > 0$ so that (4.8) can be used in (4.7) to obtain

$$b_0 = (2\beta_2)^{\alpha-1} e^{\xi/2} \left\{ \frac{2\pi i \Gamma(\alpha-a)}{\Gamma(\alpha)\Gamma(1-a)} M(\alpha-a,\alpha;-\xi) \right\}, \qquad (4.9)$$

where $a = \alpha/2 - \beta_0$ and $\xi = 2\beta_2 z$. Finally, using the Kummer transformation

$$M(b,c;z) = e^z M(c-b,c;-z) \qquad (4.10)$$

yields

$$b_0 = (2\beta_2)^{\alpha-1} e^{-\xi/2} \left\{ \frac{2\pi i \Gamma(\alpha-a)}{\Gamma(\alpha)\Gamma(1-a)} M(a,\alpha;\xi) \right\}. \qquad (4.11)$$

An expression for B_0, accurate to order ϵ^2, can thus be obtained by combining (4.1), (4.11), and (A15):

$$B_0 = (2\beta_2)^{\alpha-1}\frac{2\pi i\Gamma(\alpha-a)}{\Gamma(\alpha)\Gamma(1-a)}e^{-\xi/2}\left\{\left[1-(\epsilon^2\beta_2^4)\partial_{\beta_1}\right]\right.$$
$$\times M(a,\alpha;\xi)-\frac{\epsilon^2}{12\beta_2}\left[M(a+1,\alpha+4;\xi)\right.$$
$$\left.-M(a+3,\alpha+4;\xi)\right]-\left(\frac{\epsilon^2}{12\beta_2}+\frac{\epsilon^2\beta_2}{4}\right)$$
$$\left.\times\left[M(a+2,\alpha+2;\xi)-M(a,\alpha+2;\xi)\right]\right\}. \quad (4.12)$$

From (4.12), the value of B_0 can be found to order ϵ^2 for any z with $|z|<\infty$.

The solutions B_j and $\tilde{B}_j$ are related to a linear combination of the Kummer functions $M(b,c;z)$ and $U(b,c;z)$. The contours associated with the solutions B_j and $\tilde{B}_j$ go to infinity in the sector j in accordance with (2.15), and the integrand of u_0 as given in (4.2) has the asymptotic value $s^{\alpha-2}e^{-zs}$. In order to insure that the integrand remains bounded for all α, one must require that $|z|>|z_0|>0$ (for some given z_0), and

$$-\pi/2<\arg(sz)<\pi/2. \quad (4.13)$$

However, the subscript j already implies an argument range for the variable of integration s along the contours associated with the solutions B_j and $\tilde{B}_j$. As given by (2.15) these ranges are, for the solutions B_j,

$$2\pi(2-j)/3-\pi/6<\arg s<\pi/6+2\pi(2-j)/3. \quad (4.14)$$

From (4.13) and (4.14) one finds that the solutions b_j are well defined only over certain regions of the z-plane, namely,

$$2\pi(j-2)/3-2\pi/3<\arg z<2\pi/3+2\pi(j-2)/3. \quad (4.15)$$

Using (4.2) and the transformation

$$s=2\beta_2(t+\tfrac{1}{2}) \quad (4.16)$$

yields

$$b_j=(2\beta_2)^{\alpha-1}e^{-\xi/2}\int_{-1}^{\infty e^{i\theta}}t^{a-}(t+1)^{a+}e^{-\xi t}\,dt, \quad (4.17)$$

where the angle θ lies in the sector j. The paths of integration given in (4.17) are the image of the contours corresponding to the solutions B_j under the transformation (4.16). Similarly, from the definition of the $\tilde{B}_j$ contours, we have

$$\tilde{b}_j=(2\beta_2)^{\alpha-1}e^{-\xi/2}\int_{-1}^{\infty e^{i(\theta+2\pi)}}t^{a-}(t+1)^{a+}e^{-\xi t}\,dt. \quad (4.18)$$

The integral representation for the Kummer function U is[12]

$$U(b,c;y)=\frac{1}{2\pi i}\Gamma(1-b)e^{-b\pi i}\times\int_{\infty e^{i\theta}}^{(0+)}t^{b-1}(1+t)^{c-b-1}e^{-yt}\,dt. \quad (4.19)$$

In the integral of (4.19) the contour starts at infinity with argument θ, circles the origin in the positive direction (i.e., cuts the negative t-axis), and returns to infinity with argument $\theta+2\pi$. Since the integrand of (4.19) is analytic in the region between the origin and $t=-1$ the contour can be extended to include the point $t=-1$; hence

$$\int_{\infty e^{i\theta}}^{(0+)}=\int_{\infty e^{i\theta}}^{-1}+\int_{-1}^{\infty e^{i(\theta+2\pi)}}. \quad (4.20)$$

Using (4.20) and (4.17)–(4.19) yields

$$\tilde{b}_j-b_j=(2\beta_2)^{\alpha-1}\frac{2\pi ie^{i\pi a}}{\Gamma(1-a)}e^{-\xi/2}U(a,\alpha;\xi). \quad (4.21)$$

Using (2.23) together with (2.17), one finds that

$$B_j=[e^{i2\pi a}-1]^{-1}\{(\tilde{B}_j-B_j)-B_0\}, \quad (4.22)$$

which, for $\epsilon^2=0$, becomes

$$b_j=[e^{i2\pi a}-1]^{-1}\{(\tilde{b}_j-b_j)-b_0\}. \quad (4.23)$$

Using (4.21) to replace the combination $\tilde{b}_j-b_j$ in (4.23) and employing the identity $\Gamma(a)\Gamma(1-a)=\pi/\sin(a\pi)$ gives

$$b_j=(2\beta_2)^{\alpha-1}\Gamma(a)e^{-\xi/2}U(a,\alpha;\xi)-e^{-i\pi a}b_0/2i\sin(\pi a). \quad (4.24)$$

The relation (2.23) for $\epsilon^2=0$ gives

$$\tilde{b}_j=b_j(0,1)+b_0=e^{i2\pi a}b_j+b_0. \quad (4.25)$$

Using (4.25) to replace b_j by $\tilde{b}_j$ in (4.24) gives

$$\tilde{b}_j=(2\beta_2)^{\alpha-1}e^{i2\pi a}\Gamma(a)e^{-\xi/2}U(a,\alpha;\xi)-e^{-i\pi a}b_0/2i\sin(\pi a). \quad (4.26)$$

With (4.11) the relations (4.24) and (4.26) can be used to express b_j and $\tilde{b}_j$ entirely in terms of the Kummer functions U and M,

$$b_j(\xi)=(2\beta_2)^{\alpha-1}e^{-\xi/2}\left\{\Gamma(a)U(a,\alpha;\xi)-\frac{\pi\exp(-i\pi a)\Gamma(\alpha-a)}{\sin(\pi a)\Gamma(\alpha)\Gamma(1-a)}M(a,\alpha;\xi)\right\} \quad (4.27)$$

and

$$\tilde{b}_j(\xi)=(2\beta_2)^{\alpha-1}e^{-\xi/2}\left\{\Gamma(a)\exp(2\pi ia)U(a,\alpha;\xi)-\frac{\pi\exp(-i\pi a)\Gamma(\alpha-a)}{\sin(\pi a)\Gamma(\alpha)\Gamma(1-\alpha)}M(a,\alpha;\xi)\right\}, \quad (4.28)$$

where again $a=\alpha/2-\beta_0$ and $\xi=2\beta_2 z$.

Expressions for B_j and $\tilde{B}_j$ valid to order ϵ^2 can be obtained by using (4.27) and (4.28) in the expansion (4.1) in the same fashion that the expression (4.12) for B_0 is obtained from (4.1) using $u_0=b_0$. We do not write out the corresponding expressions for B_j and $\tilde{B}_j$ here but only note that the functions b_j and $\tilde{b}_j$ deduced from (4.1) are obtained under the restriction $|z|>|z_0|>0$ and within the argument range (4.15). In other words, B_j and $\tilde{B}_j$ cannot be evaluated at the point $z=0$ using the expansion in ϵ^2 technique. The problem of applying (4.1) at the origin for the solutions B_j and $\tilde{B}_j$ arises because the function $U(a,\alpha;\xi)$ in the expressions for b_j

and $\tilde{b}_j$ is singular at the origin ($z=0$) for $\alpha \geqslant 1$. However, the behavior of the solutions B_j and $\tilde{B}_j$ for $|z| \to 0$ for all values of α can be found by employing the relations (2.18)–(2.23). For example, multiplying (2.22) by $e^{i2\pi\alpha_-}$ and using (2.9), (2.10), and (2.17) yields

$$A_2(0,1) = -B_1(0,1) + e^{i2\pi\alpha}\tilde{B}_3 . \tag{4.29}$$

Using (2.23) with $j=1$ in (4.29) and (2.19) and (2.22) in the resulting expression gives

$$\tilde{B}_3 = [e^{i2\pi\alpha} - 1]^{-1}\{A_2(0,1) + A_1 + A_3(0,1) - B_0\} . \tag{4.30}$$

The expression (4.12) can be used to evaluate B_0, for all $|z| < \infty$, while the power series expansion (3.24) can be used to evaluate the functions A_j for small argument ($|z| \ll 1$). Thus, expressions such as (4.30) are particularly useful in evaluating B_j and $\tilde{B}_j$ in the neighborhood of the origin. For large values of $|z|$, on the other hand B_j and $\tilde{B}_j$ are more conveniently expressed using the ϵ^2 expansion of (4.1) in terms of b_j and $\tilde{b}_j$, respectively.

Since the solutions related to the Kummer function $U(a,\alpha;\xi)$ contain the combination $\tilde{b}_j - b_j$ [cf. (4.21)], it is useful to construct the combination $\tilde{B}_3 - B_3$ in a form that can be evaluated as $|z| \to 0$. To do this, we multiply (2.23) by $e^{-i2\pi\alpha_-}$ and use (2.17) to obtain

$$B_3 = \tilde{B}_3(0,-1) - B_0(0,-1) . \tag{4.31}$$

Multiplying (4.30) by $e^{-i2\pi\alpha_-}$ and using (4.31) in the resulting relation gives

$$B_3 = [e^{i2\pi\alpha} - 1]^{-1}\{A_2 + A_1(0,-1) + A_3 - B_0(1,0)\} . \tag{4.32}$$

Finally, combining (4.30) and (4.32) yields the desired combination

$$\tilde{B}_3 - B_3 = [e^{i2\pi\alpha} - 1]^{-1}\{(e^{i2\pi\alpha} - 1) \times [A_2 + A_1(0,-1) + A_3] - (1 - e^{i2\pi\alpha_+})B_0\} . \tag{4.33}$$

Expressions similar to (4.33) involving B_2 and B_1 rather than B_3 can be found by using (2.18)–(2.22). The expressions (4.33) can be evaluated at integer values of α by assuming that $\tilde{B}_3 - B_3$ is an entire function of α and then using L'Hospital's rule to obtain

$$\tilde{B}_3 - B_3 = \{[A_2(0,1) + A_3(0,1) + A_1 + B_0(1,0)]/2 + (1/2\pi i)(e^{i2\pi\alpha_-} - 1)[\partial_\alpha A_2 + \partial_\alpha A_1(0,-1) + \partial_\alpha A_3] + (1/2\pi i)(e^{i2\pi\alpha_+} - 1)\partial_\alpha B_0\}\,|_{\alpha = n} , \tag{4.34}$$

where $n = 0, \pm 1, \pm 2, \ldots$. The expressions (4.33) and (4.34) show that the combination $\tilde{B}_3 - B_3$ is finite and nonsingular at the origin for all values of α. On the other hand, the expression (4.21) with $j=3$ indicates that $\tilde{b}_3 - b_3$ is singular at the origin for $\alpha \geqslant 1$. Thus the exact solutions of (1.3) are rigorously well behaved at the origin, although the restricted expressions obtained by using an expansion in ϵ^2 indicate otherwise, which is a consequence of the singular perturbation character of this problem. Expressions for $\tilde{B}_2 - B_2$ and $\tilde{B}_1 - B_1$ are readily obtained from (4.33) using (2.18)–(2.23).

Finally, it is instructive to consider a special case in which the parameters α and β have the values

$$\alpha = \nu, \quad \beta = \nu/2 + \mu , \tag{4.35}$$

where ν is a positive integer and $\mu = 0,1,2,\ldots$. In this case

$$\alpha_+ = \nu - 1 + \mu \tag{4.36}$$

and

$$\alpha_- = -(\mu + 1) . \tag{4.37}$$

In the integral representation (2.7) there is now only an isolated pole of order $\mu + 1$ at the point $s = \beta_2$. The points $s = \pm\beta_2$ are no longer branch points, and the s-plane is no longer cut by branch lines, so that $B_j(0,1) = B_j$ and (2.23) then gives

$$\tilde{B}_j - B_j = B_0 . \tag{4.38}$$

The contour B_0 now encloses the pole at $s = \beta_2$ and from Cauchy's theorem we immediately have

$$B_0 = 2\pi i R(\beta_2) , \tag{4.39}$$

where $R(\beta_2)$ is the residue of the pole at $s = \beta_2$. Using (4.36) and (4.37), one obtains from the integrand of (2.7) the expression

$$R(\beta_2) = \left| \frac{1}{\mu!}\frac{d^{\mu}}{ds^{\mu}}[(s+\beta_2)^{\nu-1+\mu}e^{-\epsilon^2 s^3/3}e^{-s\tilde{z}}] \right|_{s=\beta_2} . \tag{4.40}$$

Making use of Leibniz's theorem, the expression (4.40) can also be written as

$$R(\beta_2) = \frac{1}{\mu!}\sum_{j=0}^{\mu}\binom{\mu}{j}E_j\frac{d^{\mu-j}}{ds^{\mu-j}}[(s+\beta_2)^{\nu-1+\mu}e^{-s\tilde{z}}]\Big|_{s=\beta_2} , \tag{4.41}$$

where $\binom{\mu}{j}$ is the usual binomial coefficient and

$$E_j = \frac{d^j}{ds^j}(e^{-\epsilon^2 s^3/3})\Big|_{s=\beta_2} . \tag{4.42}$$

By defining the small quantity $\sigma_0 \equiv \epsilon^{2/3}\beta_2$ and letting $x = (s+\beta_2)\tilde{z}$, (4.41) becomes

$$R(\beta_2) = \frac{e^{-\sigma_0^3/3}}{\mu!}e^{\beta_2\tilde{z}}\sum_{j=0}^{\mu}\binom{\mu}{j}\left[\frac{\sigma_0}{\beta_2}\right]^j\tilde{z}^{1-\nu-j}T_j \times\frac{d^{\mu-j}}{dx^{\mu-j}}[x^{\nu-1+\mu}e^{-x}]\Big|_{s=2\beta_2\tilde{z}} , \tag{4.43}$$

where the polynomials T_j are given by

$$T_j = e^{\sigma_0^3/3}\frac{d^j}{dt^j}(e^{-t^3/3})\Big|_{t=\sigma_0} \tag{4.44}$$

and $T_0 = 1$. The expression (4.43) can be further simplified by using Rodrigues' formula[13] for the Laguerre polynomials

$$L_{\mu-j}^{\nu-1+j}(x) = \frac{1}{(\mu-j)!}e^x x^{1-\nu+j}\frac{d^{\mu-j}}{dx^{\mu-j}}(x^{\nu-1+\mu}e^{-x}) . \tag{4.45}$$

Using (4.45) in (4.43), one obtains from (4.39)

$$\tilde{B}_j - B_j = B_0 = 2\pi i(2\beta_2)^{\nu-1}e^{-\beta_2\tilde{z}}e^{-\sigma_0^3/3} \times\left\{\sum_{j=0}^{\mu}\frac{(2\sigma_0)^j}{(j)!}T_j L_{\mu-j}^{\nu-1+j}(2\beta_2\tilde{z})\right\} \tag{4.46}$$

for $\mu = 0,1,2,...$, and where $\tilde{z} = z + \sigma_0^3/\beta_2$. Using (4.44) to evaluate terms in (4.46) results in

$$B_0 = 2\pi i(2\beta_2)^{\nu-1} e^{-\beta_2 z} L_\mu^{\nu-1}(2\beta_2 z) + O(\sigma_0^3) . \quad (4.47)$$

Note that $\sigma_0^3 = \epsilon^2 \beta_2^3$, so that the correction terms to B_0 as given by (4.47) are of order ϵ^2.

5. APPLICATION TO PHYSICAL PROBLEMS

In applying the solutions of (1.3) to a physical problem one needs to evaluate the solutions A_j, B_0, B_j, and $\tilde{B}_j$ in the vicinity of the real axis (Im $z \to 0+$). However, the expressions given in Secs. 3 and 4 for some of the solutions are valid only in certain sectors of the complex z-plane. Since some of these sectors contain only a portion of the real axis, the expressions given must be extended by means of analytic continuation to include the entire region of physical interest. The process of analytic continuation leads to a mixing of the solutions of the thermal and cold plasma classes. This mixing embodies the physical phenomenon of mode conversion.

In constructing analytic continuations of solutions we will take the imaginary part of the independent variable z to be small and positive. This assumption leads to exponential decay of waves in the direction of propagation and is justified on physical grounds because it corresponds to adiabatic switch-on of the exciter at frequency ω. The analytic continuation of the solutions of (1.3) can be accomplished by using the relations (2.18)–(2.23) which are derived by applying Cauchy's theorem to the contours defining the various solutions.

The asymptotic expressions for the solutions A_j as given by (3.7) are valid in sectors of aperture $4\pi/3$ as given by (3.4). Since the solution A_1 is defined in the sector $5\pi/3 < \theta < 3\pi$ ($z = |z|e^{i\theta}$), which includes the entire region of physical interest, it needs no analytic continuation. Just above the negative real axis $\theta = 3\pi - \Delta$, where Δ is a positive infinitesimal, the term $\exp(-i\zeta)$ in the asymptotic expression (3.7) for A_1 varies as $\exp|\zeta|$. Thus A_1 is exponentially growing near the negative real axis. Along the positive real axis where $\theta = 2\pi + \Delta$ the term $\exp(-i\zeta)$ varies as $\exp i|\zeta|$ so that A_1 represents an outward propagating wave in the WKB sense. The solution A_2 is defined in the region $\pi/3 < \theta < 5\pi/3$ which includes the region of physical interest only near the negative real axis where $\theta = \pi - \Delta$. The term $\exp(-i\zeta)$ in (3.4) for $j = 2$ then varies as $\exp(-|\zeta|)$ so that A_2 is exponentially decaying for Re$(z) < 0$. To analytically continue A_2, we use the relations (2.18), (2.20), and (2.22) to obtain (see Fig. 4)

$$A_2 = -A_3 - B_2 + \tilde{B}_2(1,0) - A_1(1,0) , \quad (5.1)$$

for Re$(z) > 0$. All of the functions on the right-hand side of (5.1) are defined in sectors that include the region just above the positive real axis as can be verified by (3.4) and (4.15). The relation (5.1) indicates that the solution A_2 undergoes mode conversion when the Re(z) changes sign. In this paper we use the term mode conversion to denote a process wherein a solution of one class produces a solution of the other class. For example, as indicated by (5.1), a solution of the thermal class, A_2, near the negative real axis, leads to solutions of the cold mode class as well as thermal mode class solutions near the positive real axis. Finally, the domain in which the solution A_3 is defined includes the region above the positive real axis but not the region above the negative real axis. Near the positive real axis $\theta = \Delta$ for $j = 3$ and A_3 varies as $e^{-i|\zeta|}$, thus representing, in the WKB sense, an inward propagating wave. Using (2.18), (2.21), and (2.20) to analytically continue A_3 we find that (see Fig. 4)

$$A_3 = -A_1(0,-1) - B_3 + \tilde{B}_3(1,0) - A_2 , \quad (5.2)$$

where all of the functions on the right-hand side of (5.2) are defined in the region just above the negative real axis.

The expressions derived for the solutions B_j and $\tilde{B}_j$ are also valid only in certain sectors of the z-plane as indicated by (4.15). The solutions B_3 and $\tilde{B}_3$ are defined for the entire region above the real axis and need no analytic continuation. The solutions B_2 and $\tilde{B}_2$ are defined only in the region above the positive real axis and must be continued by using (2.21) and (2.22) from which we obtain

$$B_2 = B_3 + A_1(0,-1) \quad (5.3a)$$

and

$$\tilde{B}_2 = \tilde{B}_3 + A_1 . \quad (5.3b)$$

Both functions on the right-hand side of (5.3a) and (5.3b) are defined in the region just above the negative real axis. B_2 and $\tilde{B}_2$ both undergo mode conversion when Re(z) changes sign as indicated by (5.3a) and (5.3b) because they generate a solution of the thermal mode class in addition to a solution of the cold mode class. The solutions B_1 and $\tilde{B}_1$ are defined for the region just above the negative real axis and can be continued to include the region just above the positive real axis by using (2.18) and (2.19) to obtain

$$B_1 = B_2 + A_3 \quad (5.4a)$$

and

$$\tilde{B}_1 = \tilde{B}_2 + A_3(0,1) . \quad (5.4b)$$

Again the solutions B_1 and $\tilde{B}_1$ undergo mode conversion when Re(z) changes sign.

In addition to having expressions for the solutions in the region above the real axis, a useful set of four linearly independent solutions must be selected from the general set in order to solve a well posed problem of physical interest. In particular, we have in mind applying the solutions of (1.3) to the inhomogeneous problem in which a driving source is present in the plasma. We thus shall discuss four linearly independent solutions that are convenient in constructing a Green's function for (1.3). The linear independence of a solution set can be formally established by evaluating the system Wronskian. Rather than perform this calculation we defer it to a later paper and only present here a heuristic argument for linear independence.

A Green's function must satisfy certain boundary conditions, and we impose these beforehand in choosing our solution set. The boundary conditions we impose are: (1) wavelike solutions must correspond to transport of energy away from the source, and (2) solutions must be bounded as $|\text{Re}(z)| \to \infty$. Since the thermal mode and cold plasma mode classes of solutions have vastly different scale sizes away from plasma resonance, our strategy for identifying a linearly independent set is to select one pair of linearly indepen-

dent solutions from each class. We are then assured of the linear independence of pairs of solutions from different classes because of their different scale sizes. From the cold mode class of solutions we choose the pair $B_2, \tilde{B}_2$ while from the thermal class of solutions we choose the pair A_1, A_2.

From the solution pair $B_2, \tilde{B}_2$ we next construct a pair of linearly independent solutions, one of which is bounded as $z \to \infty$, while the other is bounded for $z \to -\infty$. This solution pair is

$$B_R = -[(2\beta_2)^{1-\alpha}/2\pi i]\Gamma(-\alpha_-)e^{-i\pi\alpha_-} \times\{\tilde{B}_2(\alpha,\beta_1,\beta_2,\epsilon^2,\xi) - B_2(\alpha,\beta_1,\beta_2,\epsilon^2,\xi)\}, \quad (5.5)$$

$$B_L = -[(2\beta_2)^{1-\alpha}/2\pi i]\Gamma(-\alpha_+)e^{-i\pi\alpha_+} \times\{\tilde{B}_2(\alpha,-\beta_1,\beta_2,-\epsilon^2,-\xi) + B_2(\alpha,-\beta_1,\beta_2,-\epsilon^2,-\xi)\}. \quad (5.6)$$

Note that in (5.6) we have used the transformation (2.12) so that we are assured B_L is a solution to (1.3). The normalization coefficients in (5.5) and (5.6) have been determined using (4.27) and (4.28) and are chosen so that the expressions reduce, when $\epsilon^2 = 0$, to the simple forms given below,

$$B_R = e^{-\xi/2}U(a,\alpha;\xi), \quad (5.7)$$

$$B_L = e^{+\xi/2}U(\alpha-a,\alpha;-\xi), \quad (5.8)$$

where $a = \alpha/2 - \beta_0$. In the notation of Slater[14] these two solutions are then

$$B_R = e^{-\xi/2}y_5, \quad (5.9)$$

$$B_L = e^{-\xi/2}y_7, \quad (5.10)$$

so that, using (5.9) and (5.10) together with (4.1), we find that the Wronskian of B_R and B_L is

$$W(B_R,B_L) = e^{i\pi(\alpha/2+\beta_0)}(\xi)^{-\alpha} + O(\epsilon^2), \quad (5.11)$$

where $\xi = 2\beta_2 z$. In (5.11) we have again assumed that $\mathrm{Im}(z)$ is positive and have ignored contributions to the Wronskian arising from mode conversion. Mode conversion contributes terms to the Wronskian that are a product of cold and thermal terms and thus are rapidly oscillating. The function B_R is bounded as z approaches positive real infinity while B_L is bounded as z approaches negative real infinity.

The asymptotic form of the two thermal modes A_1 and A_2 has been discussed before. The function A_2 is exponentially decaying along the negative real axis while A_1 is exponentially growing. Thus A_2 and A_1 are clearly linearly independent. Near the positive real axis A_1 represents an outward propagating wave while from (5.1) we see that A_2, while undergoing mode conversion, contains an A_3 term which represents an inward propagating wave so that A_1 and A_2 are again linearly independent.

The solution set B_R, B_L, A_2, A_1 is convenient for constructing a Green's function. If we denote the source location by z', then the solution pair B_L, A_2 can be used exclusively for $\mathrm{Re}(z-z') < 0$ because they are the only pair bounded at negative real infinity, while the remaining pair B_R, A_1 can be used exclusively for $\mathrm{Re}(z-z') > 0$ because they are the only pair bounded at positive real infinity. While we do not construct a Green's function here, we can use the solution set B_R, A_1, B_L, A_2 to illustrate the type of mode conversion that may occur in a plasma with a driving source.

As an illustration of mode conversion in a physical problem, consider the solution B_R for $\mathrm{Re}(z) > 0$. This cold mode solution as given by (5.5) depends upon B_2 and $\tilde{B}_2$ which can be analytically continued to the region just above the negative real axis by using (5.3a) and (5.3b) from which we obtain, for $\mathrm{Re}(z) < 0$,

$$B_R = -[(2\beta_2)^{1-\alpha}/2\pi i]\Gamma(-\alpha_-)e^{-i\pi\alpha_-} \times\{\tilde{B}_3(\alpha,\beta_1,\beta_2,\epsilon^2,\xi) - B_3(\alpha,\beta_1,\beta_2,\epsilon^2,\xi)\} + [(2\beta_2)^{1-\alpha}/\Gamma(1+\alpha_-)]e^{-i2\pi\alpha_-} \times A_1(\alpha,\beta_1,\beta_2,\epsilon^2,\xi). \quad (5.12)$$

We have already indicated that A_1 is exponentially decaying near the negative real axis while the combination, $\tilde{B}_3 - B_3$, is defined in the region above the negative real axis and according to (4.21) is proportional to the hypergeometric function $U(a,\alpha;\xi)$ for $\epsilon^2 = 0$. The expression (5.12) illustrates mode conversion in that a solution of the cold class on one side of plasma resonance consists of a combination of cold and thermal modes on the other side of resonance.

We note that in this particular example the amount of mode converted thermal mode is proportional to $1/\Gamma(1+\alpha_-)$, and thus can be zero when

$$1 + \alpha_- = \alpha/2 - \beta = -l, \quad (5.13)$$

where $l = 0,1,2,\ldots$. When α is an integer, as is the case for a plasma, the condition (5.13) is identical to (4.35). In Sec. 4 we found that under these conditions the cold mode solutions were everywhere proportional to B_0 which in turn could be expressed as a sum involving Laguerre polynomials [see (4.46)]. Thus at certain special values of the parameter β the phenomenon of mode conversion does not occur in the sense that the amplitude of the thermal mode is zero, and, furthermore, at these values of β the cold solutions are proportional to Laguerre polynomials for $\epsilon^2 = 0$.

The quenching of the mode conversion process does not occur for all possible solutions, e.g., the solutions B_L for $\mathrm{Re}(z) < 0$, corresponding to a source located in the region $z > 0$. Using (5.6) and the relations (5.3a) and (5.3b) with the proper parameter values, we obtain the analytic continuation, for $\mathrm{Re}(z) > 0$,

$$B_L = -[(2\beta_2)^{1-\alpha}/2\pi i]\Gamma(-\alpha_+)e^{-i\pi\alpha_+} \times\{\tilde{B}_3(\alpha,-\beta_1,\beta_2,-\epsilon^2,-\xi) - B_3(\alpha,-\beta_1,\beta_2,-\epsilon^2,-\xi)\} + [(2\beta_2)^{1-\alpha}/\Gamma(1+\alpha_+)]e^{-i2\pi\alpha_+} \times A_1(\alpha,-\beta_1,\beta_2,-\epsilon^2,-\xi). \quad (5.14)$$

Since α and β are positive, $1/\Gamma(1+\alpha_+) = 1/\Gamma(\alpha/2+\beta)$ is never zero and mode conversion always occurs in this case. Furthermore, the mode conversion process is not restricted to the generation of thermal modes by cold modes. The thermal modes also undergo the mode conversion process and generate cold modes. Although the solution A_1 does not undergo mode conversion when $\mathrm{Re}(z)$ changes sign, the solution A_2 does as is illustrated by the analytic continuation of A_2 given in (5.1).

6. CONCLUSIONS

By applying the Laplace integral technique we have obtained integral representations of the solutions of the fourth-order differential Eq. (1.3). The solutions are distinguished by the contours along which the integral is evaluated. The solutions A_j are obtained by integrating along contours having both end points at infinity and have been identified as belonging to a class of solutions characterized by short scale lengths and thus are referred to as thermal modes. The solutions B_0, B_j, and $\tilde{B}_j$ are obtained by integrating along contours with at least one end point at the branch point $s = -\beta_2$ and belong to a class of solutions characterized by long scale lengths and are referred to as cold modes. The properties of the exact solutions A_j, B_0, B_j, and $\tilde{B}_j$ have been elucidated by expressing them in terms of more familiar functions using power series and asymptotic expansions.

The solutions A_j have been evaluated for large and small values of the independent variable z by using different techniques. A power series expansion in z has been derived in (3.24) in order to determine the solutions A_j in the neighborhood of the origin. The solutions A_j are clearly finite and well behaved in the vicinity of the origin. For large values of z the functional form of the solutions A_j has been determined in (3.11) using the saddle point method of integration to obtain the leading term in an asymptotic expansion. For real values of the independent variable these asymptotic expressions represent either exponentially decaying or growing solutions or propagating wave-like solutions.

Expressions for the solutions B_0, B_j, and $\tilde{B}_j$ have been obtained by expanding the integral representation in powers of ϵ^2. This technique allows an expression to be obtained for the exact solutions B_0, B_j, $\tilde{B}_j$ in terms of the corresponding solutions b_0, b_j, $\tilde{b}_j$ of the second-order differential equation obtained from (1.3) by setting $\epsilon^2 = 0$. The expansion (4.12) obtained in this manner for the solution B_0 involves the Kummer function $M(b,c;z)$ and is valid in the entire z-plane ($|z| < \infty$). In particular, the solution B_0 can be evaluated in the neighborhood of the origin using (4.12) and thus is useful in evaluating other members of the cold mode solution class near the origin.

An expansion in powers of ϵ^2 can be obtained for the solutions B_j and $\tilde{B}_j$ but involves the restriction $|z| > |z_0| > 0$. The expansion involves the corresponding solutions b_j and $\tilde{b}_j$ which contain the Kummer function $U(b,c;z)$ which is not bounded at the origin for $\mathrm{Re}(c) \geqslant 1$. The ϵ^2 expansion is helpful in evaluating the solutions B_j and $\tilde{B}_j$ for large values of z. The solutions B_j and $\tilde{B}_j$ can be evaluated in the neighborhood of the origin by using Cauchy's theorem to establish relationships among the various solutions as given in (2.18)–(2.23). Thus the solutions B_j and $\tilde{B}_j$ can be expressed entirely in terms of various combinations of the solutions A_j and B_0 as shown for $\tilde{B}_3$ in (4.30). Since expressions for the solutions A_j and B_0 valid in the neighborhood of the origin have been obtained, the solutions B_j and $\tilde{B}_j$ can also be, in principle, evaluated there. Unlike the associated solutions to the second-order differential equation, b_j and $\tilde{b}_j$, which can have divergent behavior at the origin for $\alpha \geqslant 1$, the solutions B_j and $\tilde{B}_j$ are always finite at the origin. The physical interpretation of this result is that the inclusion of thermal effects keeps the amplitude of the electric field finite at plasma resonance through the production of short scale thermal modes. The production of short scale waves near plasma resonance involves the process of mode conversion.

For certain values of the parameters, α and β, namely, α a positive integer and $\beta = \epsilon^2\beta_2^3/2 + \beta_1/2\beta_2$ a positive half-integer, the topology of the integration plane is greatly simplified. While the sector structure in the integration plane remains, there are no branch points but only an isolated pole. The quantities α_+ and α_- have integer values under these conditions so that the phase factors $\exp(i2\pi m\alpha_+)$ and $\exp(i2\pi\alpha_-)$ are unity and as indicated by (2.17) all solutions $X(m,n)$ are equivalent to the solution X evaluated on the principal sheet. The combination of solutions $\tilde{B}_j - B_j$ is equivalent to the solution B_0 as indicated in (4.38). Furthermore, the solution B_0 can be expressed as a series of generalized Laguerre polynomials as given in (4.46) which is convenient for numerically evaluating quantities for physical applications. Finally, we note that as shown in (4.34) the combination $\tilde{B}_3 - B_3$ generally contains an admixture of the thermal mode solutions A_j. For these special parameter values, however, the combinations $\tilde{B}_3 - B_3 = B_0$; hence it contains no thermal mode solutions. Physically, this indicates that mode conversion does not occur at these special parameter values.[15]

The asymptotic expressions given for the solutions A_j, B_j, and $\tilde{B}_j$ are defined in certain sectors of the z-plane as given by (3.4) and (4.15). Not all of these sectors contain the region just above the real axis which is the region of interest in physical applications. The solutions can be analytically continued, however, and it is this process which gives rise to the phenomenon of mode conversion. A solution of either the thermal or cold mode class on one side of plasma resonance gives rise to a combination of both classes on the other side of resonance. Physically, the mode conversion process serves to limit the amplitude of the solutions at plasma resonance. Specifically, the solutions of the purely cold plasma ($\epsilon^2 = 0$) which exhibit singularities at the origin now correspond to solutions which are finite at the origin but produce thermal modes through the mode conversion process which carry wave energy away from the resonance region. In addition, for special values of the parameters, it is possible to construct cold mode class solutions which are finite at the origin without generating thermal mode class solutions. This quenching of the mode conversion process corresponds to a change in the topology of the integration plane in which the branch points become an isolated pole.

ACKNOWLEDGMENTS

We are grateful to Professor Earl A. Coddington for illuminating discussions concerning integral representations. This work has been supported by the Office of Naval Research.

APPENDIX

1. General expression for the derivatives of $u(z)$

The nth derivative of the function $u(\alpha,\beta_1,\beta_2,\epsilon^2,z)$ with respect to z, which we denote as $u^{(n)}$, can be obtained from the integral representation (2.7):

$$u^{(n)}(\alpha,\beta_1,\beta_2,\epsilon^2,z) = (-)^n \int_C s^n (s-\beta_2)^{\alpha_-}(s+\beta_2)^{\alpha_+} \exp(-s^3\epsilon^2/3 - sz)\, ds, \tag{A1}$$

where C is any of the contours discussed in Sec. 2. To find an expression for the first derivative, we note that

$$s = (1/4\beta_2)\,[(s+\beta_2)^2 - (s-\beta_2)^2]\,, \tag{A2}$$

together with

$$u(\alpha+2,\beta_1+2\beta_2,\beta_2,\epsilon^2,z) = \int_C (s+\beta_2)^2 (s+\beta_2)^{\alpha_+}(s-\beta_2)^{\alpha_-} \exp(-s^3\epsilon^2/3 - zs)\, ds \tag{A3}$$

and

$$u(\alpha+2,\beta_1-2\beta_2,\beta_2,\epsilon^2,z) = \int_C (s-\beta_2)^2 (s+\beta_2)^{\alpha_+}(s-\beta_2)^{\alpha_-} \exp(-s^3\epsilon^2/3 - zs)\, ds\,. \tag{A4}$$

Using (A1) with $n = 1$, the expression (A2) together with (A3) and (A4) gives

$$u^{(1)}(\alpha,\beta_1,\beta_2,\epsilon^2,z) = (-1/4\beta_2)[u(\alpha+2,\beta_1+2\beta_2,\beta_2,\epsilon^2,z) - u(\alpha+2,\beta_1-2\beta_2,\epsilon^2,z)]\,. \tag{A5}$$

The second derivative of u can be found in a similar fashion by using the expression

$$s^2 = (s-\beta_2)(s+\beta_2) + \beta_2^2\,; \tag{A6}$$

we then obtain

$$u^{(2)}(\alpha,\beta_1,\beta_2,\epsilon^2,z) = u(\alpha+2,\beta_1,\beta_2,\epsilon^2,z) + \beta_2^2 u(\alpha,\beta_1,\beta_2,\epsilon^2,z)\,. \tag{A7}$$

Repeated use of the expression (A7) yields

$$u^{(2n)}(\alpha,\beta_1,\beta_2,\epsilon^2,z) = \sum_{j=0}^{n} \binom{n}{j} \beta_2^{2(n-j)}\, u(\alpha+2j,\beta_1,\beta_2,\epsilon^2,z)\,, \tag{A8}$$

where $n = 0,1,2,3,\ldots$ and $\binom{n}{j} = n!/(n-j)!j!$, are the binomial coefficients. Application of (A5) to (A8) then yields

$$u^{(2n+1)}(\alpha,\beta_1,\beta_2,\epsilon^2,z) = \frac{1}{4}\sum_{j=0}^{n} \binom{n}{j} \beta_2^{2(n-j)-1} \times [u(\alpha+2(j+1),\beta_1-2\beta_2,\beta_2,\epsilon^2,z) - u(\alpha+2(j+1),\beta_1+2\beta_2,\beta_2,\epsilon^2,z)]\,. \tag{A9}$$

2. A series expansion of u in powers of ϵ^2

In this appendix we outline the procedure for obtaining a series expansion of $u(\alpha,\beta_1,\beta_2,\epsilon^2,z)$ in powers of ϵ^2 and explicitly calculate the first two coefficients of this series. We first assume that the integrand of (2.7) is an analytic function of ϵ^2 in the vicinity of $\epsilon^2 = 0$ and expand a portion of the integrand in a Taylor's series about the point $\epsilon^2 = 0$. We will discuss the validity of this assumption later. The integrand of (2.7) depends on ϵ^2 indirectly through the parameters α_+, α_- and directly as $\exp[-\epsilon^2(s^3/3 + \beta_2^2 s)]$. Thus part of the integrand in (2.7) can be written as

$$(s+\beta_2)^{\alpha_+}(s-\beta_2)^{\alpha_-} \exp[-\epsilon^2(s^3/3 + \beta_2^2 s)]$$
$$\simeq (s+\beta_2)^{a_+}(s-\beta_2)^{a_-}\{1 - \epsilon^2\{s^3/3 + \beta_2^2 s - (\beta_2^3/2)\ln[(s+\beta_2)/(s-\beta_2)]\}\} + \cdots, \tag{A10}$$

where we have explicitly shown only the first two terms of the Taylor series expansion. The parameters a_+ and a_- are simply the values of α_+ and α_- at $\epsilon^2 = 0$, namely,

$$a_+ = \alpha_+(\epsilon^2 = 0) = \alpha/2 - 1 + \beta_1/2\beta_2\,, \tag{A11}$$

$$a_- = \alpha_-(\epsilon^2 = 0) = \alpha/2 - 1 - \beta_1/2\beta_2\,. \tag{A12}$$

Using (A10) in (2.7), we obtain

$$u(\alpha,\beta_1,\beta_2,\epsilon^2,z) = \int_C (s-\beta_2)^{a_-}(s+\beta_2)^{a_+} e^{-zs}\, ds$$
$$- \epsilon^2\left\{\beta_2^2 \int_C s(s-\beta_2)^{a_-}(s+\beta_2)^{a_+} e^{-zs}\, ds + \frac{1}{3}\int_C s^3 (s-\beta_2)^{a_-}(s+\beta_2)^{a_+} e^{-zs}\, ds\right.$$
$$\left. - \frac{\beta_2^3}{2}\int_C \ln[(s+\beta_2)/(s-\beta_2)](s+\beta_2)^{a_+}(s-\beta_2)^{a_-} e^{-zs}\, ds\right\} + \cdots. \tag{A13}$$

Noting that

$$\int_C \ln[(s+\beta_2)/(s-\beta_2)](s+\beta_2)^{a_+}(s-\beta_2)^{a_-} e^{-zs}\, ds = 2\beta_2 \frac{\partial}{\partial\beta_1}\int_C (s+\beta_2)^{a_+}(s-\beta_2)^{a_-} e^{-zs}\, ds\,, \tag{A14}$$

and, using (A1) together with (A9), we obtain

$$u(\alpha,\beta_1,\beta_2,\epsilon^2,z) = u(\alpha,\beta_1,\beta_2,0,z)$$
$$-\frac{\epsilon^2}{4\beta_2}\left\{\left[\frac{1}{3}\right]u(\alpha+4,\beta_1+2\beta_2,\beta_2,0,z)+\left[\frac{1}{3}+\beta_2^2\right]u(\alpha+2,\beta_1-2\beta_2,\beta_2,0,z)\right.$$
$$-\left[\frac{1}{3}\right]u(\alpha+4,\beta_1-2\beta_2,\beta_2,0,z)-\left[\frac{1}{3}+\beta_2^2\right]u(\alpha+2,\beta_1+2\beta_2,\beta_2,0,z)$$
$$\left.-4\beta_2^5\frac{\partial}{\partial\beta_1}u(\alpha,\beta_1,\beta_2,0,z)\right\}+\cdots. \tag{A15}$$

Writing

$$u_0 \equiv u(\alpha,\beta_1,\beta_2,0,z) = \int_C (s-\beta_2)^{\alpha}(s+\beta_2)^{\alpha+}e^{-zs}\,ds\,, \tag{A16}$$

we can then express (A15) as

$$u(\alpha,\beta_1,\beta_2,\epsilon^2,z) = u(\alpha,\beta_1,\beta_2,0,z)+\epsilon^2(\partial_{\epsilon^2}u)|_{\epsilon^2=0}+\cdots = u_0-\epsilon^2 h(u_0)/4\beta_2+\cdots, \tag{A17}$$

where $h(u_0)$ is a functional of u_0 and denotes the term in braces in (A15). That is, once the functions u_0 (which are independent of the magnitude of ϵ^2) are known, the behavior of $u(\alpha,\beta_1,\beta_2,\epsilon^2,z)$ can, at least in principle, be determined for any value of ϵ^2 provided the Taylor series expansion is valid.

For the Taylor series to be valid the functions u_0 must be analytic functions of ϵ^2 in the neighborhood of $\epsilon^2 = 0$. Furthermore, the coefficients in the Taylor series must be bounded for all values of z. The functions u_0 are analytic functions of ϵ^2 only for the contours corresponding to the solutions B_0, B_j, and $\tilde{B}_j$ because these solutions become solutions of the second order differential equation when $\epsilon^2 = 0$. On the other hand, u_0 integrated along the contours for B_0, B_j, and $\tilde{B}_j$ can be singular for certain values of z and the coefficients of the Taylor series are not bounded as z approaches these points. In the problem treated here, for example, singularities in b_j and $\tilde{b}_j$; can occur at the point $z = 0$ for $\alpha \geqslant 1$. In this case the expansion in powers of ϵ^2 is valid only for z bounded away from the origin, i.e., $|z| > |z_0| > 0$, for a given z_0, and for arg z restricted to the appropriate sectors, as given in the text by (4.15).

[1]J. E. Maggs and G. J. Morales, "Electrostatic Modes in a Magnetized Plasma With a Longitudinal Density Gradient," J. Plasma Phys. **29**, 177 (1983).

[2]W. Wasow, "Asymptotic Solution of the Differential Equation Governing Hydrodynamic Stability in a Domain Containing a Transition Point," Ann. Math. **58**, 222 (1953).

[3]A. L. Rabenstein, "Asymptotic Solutions of $u^{iv}+\lambda^2(zu''+\alpha u'+\beta u)=0$ for Large $|\lambda|$," Arch. Rat. Mech. Anal. **1**, 418 (1958).

[4]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), Sec. 5.8.

[5]Alfredo Baños, Jr., *Dipole Radiation in the Presence of a Conducting Half-Space* (Pergamon, New York, 1966), Chap. 3.

[6]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1963), Sec. 2.53.

[7]Ref. 6, Sec. 3.33 and 3.7.

[8]See Ref. 3.

[9]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Trancendental Functions*, Bateman Manuscript Project (McGraw-Hill, New York, 1953), Vol. 1, Eq. 1.6.6.

[10]L. J. Slater, "Confluent Hypergeometric Functions," in *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (U.S. Government Printing Office, Washington, D.C., 1964).

[11]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. Tricomi, Ref. 9, Sec. 6.11.1.

[12]Ref. 9, Sec. 6.11.2.

[13]V. W. Hochstrasser, "Orthogonal Polynomials," in *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (U.S. Government Printing Office, Washington, D.C., 1964).

[14]See Ref. 10.

[15]G. J. Morales and J. E. Maggs, "Guided Modes Between Upper Hybrid and Plasma Resonance," J. Plasma Phys. **30**, 267 (1983).

Exact solution for the discrete end-to-end intrachain memory function of a reptating polymer chain

Douglas A. Bernard and Jaan Noolandi[a)]
Xerox Research Centre of Canada, 2660 Speakman Drive, Mississauga, Ontario, L5K 2L1, Canada

(Received 7 September 1983; accepted for publication 16 December 1983)

This article considers the dynamics of polymer chains which are subject to entanglement constraints. We give the exact solution for the end-to-end intrachain memory function of a finite, N-link bead-rod chain which diffuses in configuration space according to the reptation model described by the stochastic earthworm equations of Doi and Edwards. Defining the appropriate two-point spatial correlation function, we develop a recursion expression for its Fourier transform, the equal-time structure factor, and evaluate this expression using a generating function method. The nonuniversal, N-dependent form of the resulting memory function is illustrated and compared with the universal form achieved in the continuum limit for $N \to \infty$.

PACS numbers: 61.40.Km, 66.10.Cb, 82.35. + t, 02.30. + g

1. INTRODUCTION

The Brownian motion of a flexible polymer chain in a viscous medium constitutes a fundamental research area in polymer physics. Specific aspects of the dynamics of polymers are exhibited in the kinetics of reactions involving the contact of reactive groups attached at various positions along the flexible chain backbone. For example, when reactions occur between the two ends of a nonbranched chain,[1] the internal Brownian motion manifests itself in the frequency of end-to-end encounters. In recent papers[2,3] we defined the first-order rate function which characterizes such an ensemble of self-reacting polymer chains, and showed it to be directly related to the *intrachain memory function*, the subject of the present paper.

The end-to-end intrachain memory function, denoted $S(t)$, represents the probability that the ends of a chain are closer than an infinitesimal "capture" radius r_0 at time t, given they were in contact at an earlier time $t = 0$. $S(t)$ is determined by the Brownian diffusion of the chain in the absence of reactions, which in turn depends on the type of polymer system. In polymer melts of high molecular weight, different chains interpenetrate and entanglements occur, leading to relaxation phenomena completely different from those found[4] in dilute solutions. Building on Edwards' mean field view[5] that each chain experiences entanglement constraints in the form of a confining tube defined by the chain's average contour, the *reptation model* proposed by de Gennes[6] asserts that the diffusion of such chains is essentially one-dimensional. However, all three-dimensional random-walk conformations are accessible because the chain is not confined beyond its ends.

In this paper we calculate the end-to-end intrachain memory function in polymer melts using the reptation model. Because $S(t)$ is a conditional probability, it can be defined as a particular value of the propagator for the relative end-to-end vector of the polymer chain. This propagator is obtained from a distribution function defined on the configuration space of the polymer chain. The polymer dynamics enters when we consider the time evolution of the corresponding ensemble of polymer chains. The initial ensemble, whose chain conformations are consistent with the condition of chain closure at $t = 0$, relaxes to an equilibrium ensemble at $t \to \infty$.

In our calculation of $S(t)$ we assume that the polymer dynamics is represented at early times by the Rouse bead-spring model[4] discussed in Appendix B, which describes Brownian motion in the absence of entanglements, and at later times by the reptation model as given by the earthworm equations of Doi and Edwards.[7] The crossover time T_δ is the time required for chain ends initially in contact to separate by a distance equal to the average distance between entanglements. Thus for $0 < t < T_\delta$ we assert that the chain is unaware of the entanglement network. Since the reptation model in its present form only describes the time evolution in discrete time steps, this scheme has the advantage that the time evolution can be described continuously at short times.

2. THE MODEL

The end-to-end intrachain memory function $S(t)$ is given by

$$S(t) = \Omega \Gamma(\underline{r} = 0 | \underline{r}' = 0; t), \quad 0 \leqslant t \leqslant \infty, \tag{2.1}$$

where $\Omega = \frac{4}{3}\pi r_0^3$ is an infinitesimal "reaction" volume determined by a "capture radius" r_0. The propagator $\Gamma(\underline{r}|\underline{r}';t)$ is the conditional probability density that the flexible polymer chain has an end-to-end vector $\underline{r}$ at time $t > 0$, given a vector $\underline{r}'$ at $t = 0$. Because the polymer chain undergoes Brownian motion in a viscous medium, all probability distributions relax to their time-independent equilibrium forms. Consequently,

$$\lim_{t \to \infty} \Gamma(\underline{r}|\underline{r}';t) = P_{eq}(\underline{r}), \quad \forall \underline{r}', \tag{2.2}$$

where $P_{eq}(\underline{r})$ is the equilibrium function (a probability density in $\underline{r}$) for the end-to-end vector of the polymer chain (see Appendix B). Therefore, $S(t)$ has the general form

$$S(t) = \Omega P_{eq}(\underline{r} = 0)[\sigma(t) + 1], \tag{2.3}$$

where $\sigma(t)$ is independent of the reaction volume and is called the *transient part* of the intrachain memory function because Eq. (2.2) implies

a) To whom correspondence should be addressed.

 0022-2488/84/051619-12$02.50

$$\lim_{t\to\infty}\sigma(t)=0. \tag{2.4}$$

The steady state value of the memory function is $\Omega P_{eq}(\underline{r}=0)$, which is the probability that the ends of a chain in conformational equilibrium are closer than the capture radius.

Thus the intrachain memory function is determined by a particular value of the propagator $\Gamma(\underline{r}|\underline{r}';t)$, which is in turn derivable from a distribution function defined on the configuration space appropriate to a flexible polymer chain. Because it is not necessary to account for all degrees of freedom in modeling the dominant aspects of the polymer dynamics, the real chain is simulated by a *primitive chain* consisting of $N+1$ identical beads connected by N massless links at universal joints. The general configurational distribution function $\Psi(\{\underline{R}\};t)$, defined on the full primitive chain configuration space, then represents the probability density of finding a chain with bead coordinates $\{\underline{R}\}=\{\underline{R}_1,\underline{R}_2,...,\underline{R}_{N+1}\}$ at time t. Because the end-to-end vector is given in terms of $\{\underline{R}\}$ by $\underline{R}_{N+1}-\underline{R}_1$, the propagator is given by

$$\begin{aligned}\Gamma(\underline{r}|\underline{r}';t)&=\langle\delta(\underline{r}-[\underline{R}_{N+1}-\underline{R}_1])\rangle_{\Psi(\{\underline{R}\}|\underline{r}';t)}\\&=\int d\{\underline{R}\}\,\delta(\underline{r}-[\underline{R}_{N+1}-\underline{R}_1])\Psi(\{\underline{R}\}|\underline{r}';t),\end{aligned} \tag{2.5}$$

where we use $\Psi(\{\underline{R}\}|\underline{r}';t)$ to denote the conditional configurational distribution function determined by the requirement of an end-to-end vector $\underline{r}'$ at time $t=0$.

In our subsequent calculations we work with the Fourier transform of the propagator, denoted $g_{1,N+1}(\underline{k}|\underline{r}';t)$, and defined by

$$\Gamma(\underline{r}|\underline{r}';t)=\frac{1}{(2\pi)^3}\int dk\,\exp(i\underline{k}\cdot\underline{r})g_{1,N+1}(\underline{k}|\underline{r}';t). \tag{2.6}$$

This is also called the *end-to-end equal time structure factor*, which is the $(1,N+1)$ component of the *equal-time structure factor*

$$g_{n,m}(\underline{k}|\underline{r}';t)=\langle\exp(i\underline{k}\cdot[\underline{R}_n-\underline{R}_m])\rangle_{\Psi(\{\underline{R}\}|\underline{r}';t)}, \tag{2.7}$$

defined for $n,m=1,2,...,N+1$. Combining Eqs. (2.3), (2.1) and (2.6), the intrachain memory function is related to the end-to-end equal-time structure factor by

$$\begin{aligned}\frac{S(t)}{\Omega P_{eq}(\underline{r}=0)}&=\sigma(t)+1\\&=\frac{1}{P_{eq}(\underline{r}=0)}\frac{1}{(2\pi)^3}\int d\underline{k}\,g_{1,N+1}(\underline{k}|\underline{r}'=0;t).\end{aligned} \tag{2.8}$$

Thus the intrachain memory function is ultimately determined by the polymer dynamics via the conditional configurational distribution function $\Psi(\{\underline{R}\}|\underline{r}';t)$. In polymer melts, the dynamics of the primitive chain is described by the earthworm equations of Doi and Edwards,[7] which state that the primitive chain undergoes a random walk in which it slides along its contour, jumping forward or backward with equal likelihood, with only the leading bead free to choose its direction. We express this concept in terms of the general configurational distribution function of the primitive chain by writing the time evolution as

$$\begin{aligned}\Psi(\underline{R}_1,\underline{R}_2,...,\underline{R}_{N+1};t+\Delta t)&=\tfrac{1}{2}\nu_\beta(\underline{R}_2-\underline{R}_1)\int d\underline{R}'_{N+1}\,\Psi(\underline{R}_2,...,\underline{R}_{N+1},\underline{R}'_{N+1};t)\\&+\nu_\beta(\underline{R}_{N+1}-\underline{R}_N)\int d\underline{R}'_1\,\Psi(\underline{R}'_1,\underline{R}_1,...,\underline{R}_N;t),\end{aligned} \tag{2.9}$$

where Δt is a microscopic hopping time interval introduced in the reptation model. The randomization of the direction of the leading bead is provided by the normal distribution in three dimensions,

$$\nu_\beta(\underline{u})=(\beta/\pi)^{3/2}\exp(-\beta u^2), \tag{2.10}$$

which maintains a fixed mean square interbead separation $\langle u^2\rangle_{\nu_\beta(u)}=3/2\beta$, also denoted a^2, during the hopping process. Thus Eq. (2.9) gives the distribution function at discrete times

$$t_\alpha=\alpha\Delta t+t_0,\quad \alpha=0,1,2,..., \tag{2.11}$$

given its form at a time t_0. The choice of t_0 and the form of the distribution function for earlier times $0\leqslant t\leqslant t_0$ will be discussed in Sec. 5A and Appendix B.

Equation (2.9) is physically equivalent to the original earthworm equations of Doi and Edwards, who treat the bead positions as stochastic functions $\underline{R}_n(t)$. Those equations are

$$\begin{aligned}\underline{R}_n(t+\Delta t)&=\tfrac{1}{2}[1+\eta(t)]R_{n+1}(t)\\&+\tfrac{1}{2}[1-\eta(t)]R_{n-1}(t),n=1,...,N+1,\end{aligned} \tag{2.12a}$$

with the auxiliary definitions

$$\underline{R}_0(t)=\underline{R}_1(t)+\underline{a}(t), \tag{2.12b}$$

$$\underline{R}_{N+2}(t)=\underline{R}_{N+1}(t)+\underline{a}'(t). \tag{2.12c}$$

Here, $\eta(t)$ is a stochastic function which has the values $+1$ or -1 with equal likelihood, and determines whether the primitive chain jumps forwards or backwards: $\underline{a}(t)$ and $\underline{a}'(t)$ are stochastic vector functions of magnitude a, but with arbitrary direction.

3. RECURSION FOR THE EQUAL-TIME STRUCTURE FACTOR

Our calculation of the intrachain memory function in the reptation model is based on recursion to past values of the equal-time structure factor defined in Eq. (2.7). Adopting the notation $g^{(\alpha)}_{n,m}=g_{n,m}(\underline{k}|\underline{r}';t_\alpha)$, we follow the approach of Doi and Edwards,[7] who considered the dynamic structure factor $\langle\exp(i\underline{k}\cdot[\underline{R}_n(t)-\underline{R}_m(0)])\rangle$, and use Eq. (2.9) to express $g^{(\alpha)}_{n,m}$ in terms of the earlier values $\{g^{(\alpha-1)}_{n',m'}\}_{n',m'\in[1,N+1]}$ of the structure factor. We find

$$g^{(\alpha)}_{n,n}=g^{(\alpha-1)}_{n,n}=1,\quad n\in[1,N+1], \tag{3.1a}$$

$$g^{(\alpha)}_{n,m} = \tfrac{1}{2}\left[g^{(\alpha-1)}_{n-1,m-1} + g^{(\alpha-1)}_{n+1,m+1}\right], \quad n\in[2,N],\ m\in[2,N], \tag{3.1b}$$

$$g^{(\alpha)}_{1,m} = \tfrac{1}{2}\left[C(\underline{k})g^{(\alpha-1)}_{1,m-1} + g^{(\alpha-1)}_{2,m+1}\right], \quad m\in[2,N], \tag{3.1c}$$

$$g^{(\alpha)}_{n,N+1} = \tfrac{1}{2}\left[g^{(\alpha-1)}_{n-1,N} + C(\underline{k})g^{(\alpha-1)}_{n+1,N+1}\right], \quad n\in[2,N], \tag{3.1d}$$

$$g^{(\alpha)}_{1,N+1} = \tfrac{1}{2}\left[C(\underline{k})g^{(\alpha-1)}_{1,N} + C(\underline{k})g^{(\alpha-1)}_{2,N+1}\right], \tag{3.1e}$$

with

$$g_{n,m}(\underline{k}\,|\underline{r}';t) = g_{n,m}(-\underline{k}\,|\underline{r}';t). \tag{3.2}$$

Randomization of the direction of the leading bead introduces the factor

$$C(\underline{k}) = \langle \exp(i\underline{k}\cdot\underline{u})\rangle_{\rho(\underline{u})} = \exp(-k^2/4\beta). \tag{3.3}$$

This factor appears in the above equations whenever n or m is at an end of the chain. Equations (3.1) are depicted in Fig. 1 for $m > n$, which shows the $g^{(\alpha-1)}_{n',m'}$ that determine $g^{(\alpha)}_{n,m}$ in a single recursion step $\alpha\to\alpha-1$. It also indicates that the factor $C(\underline{k})$ only appears when $g^{(\alpha)}_{n,m}$ is related to $g^{(\alpha-1)}_{n',m'}$ on a different row, r, where rows are defined by constant $n-m$:

$$r = N + 1 - |n-m|. \tag{3.4}$$

We now focus on the end-to-end equal-time structure factor $g^{(\alpha)}_{1,N+1}$, which determines the propagator and the memory function, and apply Eqs. (3.1) recursively. For example, after three steps we find

$$\begin{aligned} g^{(\alpha)}_{1,N+1} = {} & C(\underline{k})\{(1/2^3)g^{(\alpha-3)}_{1,N} + (1/2^3)g^{(\alpha-3)}_{2,N+1}\} && \text{(row } r=2\text{)} \\ & + C(\underline{k})^2\{(1/2^3)g^{(\alpha-3)}_{1,N-1} + 2(1/2^3)g^{(\alpha-3)}_{2,N} + (1/2^3)g^{(\alpha-3)}_{3,N+1}\} && \text{(row } r=3\text{)} \\ & + C(\underline{k})^3\{(1/2^3)g^{(\alpha-3)}_{1,N-2} + (1/2^3)g^{(\alpha-3)}_{4,N+1}\}. && \text{(row } r=4\text{)} \end{aligned} \tag{3.5}$$

After α' steps, we define $b^{(\alpha')}_{r,j}$ to be the jth numerical coefficient along row r of $g^{(\alpha-\alpha')}_{n,m}$ in the expansion of $g^{(\alpha)}_{1,N+1}$, where $n=j$ and $m=j+N+1-r$ (see Fig. 2). In particular, after α steps we have

$$\begin{aligned} g^{(\alpha)}_{1,N+1} = {} & \sum_{r=1}^{N} C(\underline{k})^{r-1}\left\{\sum_{j=1}^{r} b^{(\alpha)}_{r,j} g^{(0)}_{j,j+N+1-r}\right\} \\ & + C(\underline{k})^N\{b^{(\alpha)}_{N+1,1} + b^{(\alpha)}_{N+1,N+1}\}. \end{aligned} \tag{3.6}$$

Note that for $\alpha \geqslant N$ the recursive process reaches the last row $r = N+1$ and terminates at the end element $b^{(\alpha)}_{N+1,1} = b^{(\alpha)}_{N+1,N+1}$ because of Eq. (3.1a).

Therefore $g^{(\alpha)}_{1,N+1}$ is expressed in terms of the structure factors at time t_0, $\{g_{n,m}{}^{(\alpha=0)}\}_{n,m\in[1,N+1]}$, which are determined by the configurational distribution at time t_0. This initial condition will be discussed in Sec. 5A.

The row elements $b^{(\alpha)}_{r,j}$ for $\alpha = 1,2,\ldots$ obey the recursion relations

$$b^{(\alpha)}_{1,1} = 0, \tag{3.7a}$$

$$\left.\begin{aligned} b^{(\alpha)}_{r,1} = b^{(\alpha)}_{r,r} &= \tfrac{1}{2}\left[b^{(\alpha-1)}_{r-1,1} + b^{(\alpha-1)}_{r,2}\right], \quad r\in[2,N] \\ b^{(\alpha)}_{r,j} &= \tfrac{1}{2}\left[b^{(\alpha-1)}_{r,j-1} + b^{(\alpha-1)}_{r,j+1}\right], \quad j\in[2,r-1] \end{aligned}\right\}, \tag{3.7b}$$

and

$$\left.\begin{aligned} b^{(\alpha)}_{N+1,1} = b^{(\alpha)}_{N+1,N+1} &= \tfrac{1}{2}b^{(\alpha-1)}_{N,1} + b^{(\alpha-1)}_{N+1,1}, \quad r = N+1 \\ b^{(\alpha)}_{N+1,j} &= 0, \quad j\in[2,N] \end{aligned}\right\}, \tag{3.8}$$

with the initial conditions

$$b^{(\alpha=0)}_{1,1} = 1 \tag{3.9a}$$

and

$$b^{(\alpha=0)}_{r,j} = 0, \quad r>1. \tag{3.9b}$$

Note that the $b^{(\alpha)}_{r,j}$ do not depend on the number of links, N in the primitive chain, unless $r = N+1$. The evolution of the $b^{(\alpha)}_{r,j}$ is depicted in Fig. 3, where for $\alpha = 0$ an initial value $b^{(0)}_{1,1} = 1$ at the apex of the triangle diffuses downwards and across the rows, until for $\alpha\to\infty$ it ends up at the bottom corners, in the last row $r = N+1$, with a final value $2b^{(\infty)}_{N+1,1} = 1$.

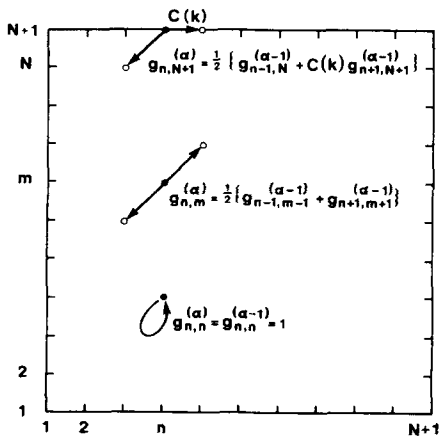


FIG. 1. This illustrates the recursion relations, Eqs. (3.1) for the equal time structure factor $g^{(\alpha)}_{n,m} = g_{n,m}(\underline{k}\,|\underline{r}';t_\alpha)$ defined in Eq. (2.7). It shows the $g^{(\alpha-1)}_{n',m'}$ that determine $g^{(\alpha)}_{n,m}$ in a single recursion step $\alpha\to\alpha-1$.

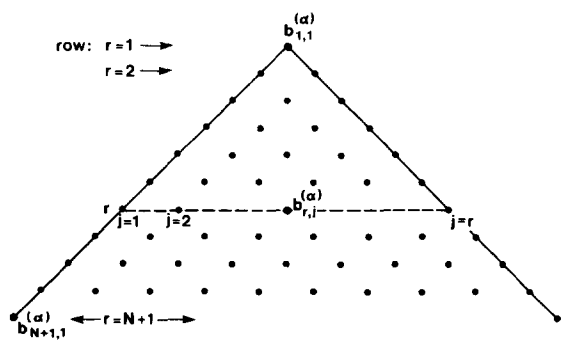


FIG. 2. In the expansion of the end-to-end equal time structure factor $g^{(\alpha)}_{1,N+1}$ using the recursion illustrated in Fig. 1, $b^{(\alpha')}_{r,j}$ is the jth numerical coefficient along row r of $g^{(\alpha-\alpha')}_{n,m}$ where $n=j$ and $m=j+N+1-r$. See also Eq. (3.6).

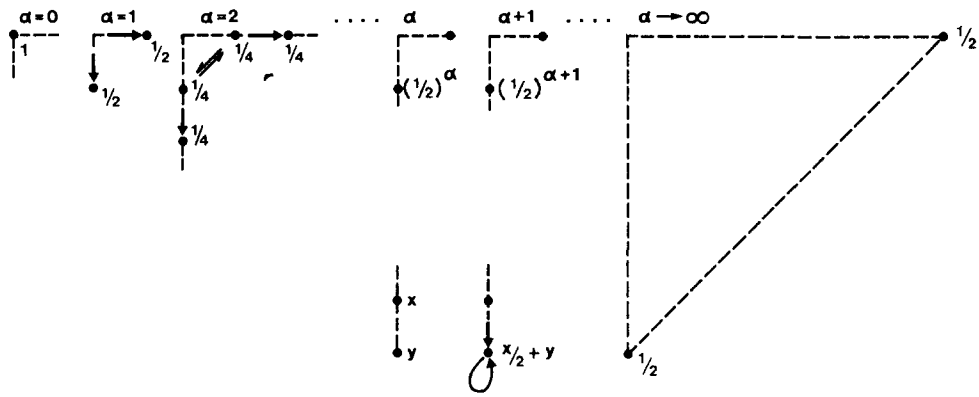


FIG. 3. The evolution of the row elements $b_{r,j}^{(\alpha)}$ for $\alpha = 1,2,\ldots,\infty$. For $\alpha = 0$, an initial value $b_{1,1}^{(0)} = 1$ at the apex of the triangle (see Fig. 2) diffuses downwards and across the rows, until it ends up at the bottom corners of the last row $r = N+1$, with $b_{N+1,1}^{(\infty)} = b_{N+1,N+1}^{(\infty)} = \frac{1}{2}$.

4. EXPRESSIONS FOR ROW ELEMENTS

The row elements $b_{r,j}^{(\alpha)}$, which determine the end-to-end equal-time structure factor in Eq. (3.6), can be computed by applying a generating function method based on the recursion relations, Eqs. (3.7) and (3.8). Since these relations differ in structure for $r = 1, 2,\ldots, N$ and $r = N+1$, the analysis is given in two parts.

A. The rows $r = 1,2,\ldots, N$

If we denote the end element of any row r by

$$a_r^{(\alpha)} = b_{r,1}^{(\alpha)} = b_{r,r}^{(\alpha)}, \tag{4.1}$$

then for $r = 1,2,\ldots, N$ the recursion relations in Eqs. (3.7) can be rewritten

$$b_{r,i}^{(\alpha+1)} = \sum_{j=1}^{r} U_{i,j} b_{r,j}^{(\alpha)} + \tfrac{1}{2} a_{r-1}^{(\alpha)} (\delta_{i,1} + \delta_{i,r}), \quad \alpha \geqslant 0,\ 1 \leqslant r \leqslant N, \tag{4.2}$$

with the initial condition of Eq. (3.9), and the formal definition $a_0^{(\alpha)} = 0\ \forall \alpha$. For fixed r, the $U_{i,j}$ form an $r \times r$ matrix $\mathbf{U}$ given by

$$U_{i,j} = \tfrac{1}{2}(\delta_{i,j-1} + \delta_{i,j+1}), \quad i,j \in [1,r]. \tag{4.3}$$

Defining the column vector $\mathbf{b}_r^{(\alpha)} = (b_{r,1}^{(\alpha)}, b_{r,2}^{(\alpha)},\ldots, b_{r,r}^{(\alpha)})^{\dagger}$ and the column vector $\mathbf{d} = (1,0,\ldots,0,1)^{\dagger}$, Eq. (4.2) becomes

$$\mathbf{b}_r^{(\alpha+1)} = \mathbf{U} \cdot \mathbf{b}_r^{(\alpha)} + \tfrac{1}{2} a_{r-1}^{(\alpha)} \mathbf{d}. \tag{4.4}$$

Applying this successively, we obtain an expression involving only the end elements:

$$\mathbf{b}_r^{(\alpha+1)} = \tfrac{1}{2}[a_{r-1}^{(\alpha)} \mathbf{U}^0 + a_{r-1}^{(\alpha-1)} \mathbf{U}^1 + \ldots + a_{r-1}^{(0)} \mathbf{U}^{\alpha}] \cdot \mathbf{d}, \tag{4.5}$$

where $\mathbf{U}^s$ is the sth power of $\mathbf{U}$. Now the $r \times r$ matrix $\mathbf{P}$ with elements

$$P_{i,j} = \sin[ij\pi/(r+1)], \quad i,j \in [1,r], \tag{4.6}$$

diagonalizes $\mathbf{U}$ in the form[8]

$$\mathbf{U} = [2/(r+1)] \mathbf{P} \cdot \mathbf{\Lambda} \cdot \mathbf{P}, \tag{4.7}$$

with

$$\Lambda_{i,j} = \delta_{i,j} \lambda_{r,i}, \quad i,j \in [1,r], \tag{4.8a}$$

and

$$\lambda_{r,i} = \cos[i\pi/(r+1)], \quad i \in [1,r]. \tag{4.8b}$$

Therefore,

$$\mathbf{b}_r^{(\alpha+1)} = [1/(r+1)] \mathbf{P} \cdot [a_{r-1}^{(\alpha)} \Lambda^0 + a_{r-1}^{(\alpha-1)} \Lambda^1 + \cdots + a_{r-1}^{(0)} \Lambda^{\alpha}] \cdot \mathbf{P} \cdot \mathbf{d}, \tag{4.9}$$

yielding the row elements

$$b_{r,j}^{(\alpha+1)} = \frac{1}{r+1} \sum_{k=1}^{r} \eta_{r,jk} [a_{r-1}^{(\alpha)} \lambda_{r,k}^{0} + a_{r-1}^{(\alpha-1)} \lambda_{r,k}^{1} + \cdots + a_{r-1}^{(0)} \lambda_{r,k}^{\alpha}], \tag{4.10}$$

where $\eta_{r,jk}$ is

$$\eta_{r,jk} = \sin[jk\pi/(r+1)]\{\sin[k\pi/(r+1)] + \sin[rk\pi/(r+1)]\}. \tag{4.11}$$

A generating function method is used to evaluate Eq. (4.10). For a general sequence $\{a^{(\alpha)}\}_{\alpha = 0,1,\ldots}$, we define the associated generating function $A(z)$ by

$$A(z) = \sum_{\alpha=0}^{\infty} a^{(\alpha)} z^{\alpha}, \tag{4.12}$$

where z is complex. For future reference, the inversion formula is

$$a^{(\alpha)} = \frac{1}{2\pi i} \oint_{|z| = \mu} \frac{A(z)}{z^{\alpha+1}} dz, \quad \alpha = 0,1,\cdots, \tag{4.13}$$

where we assume the radius of the contour circle to be $\mu < 1$. Defining $B_{r,j}(z)$ as the generating function for the sequence $\{b_{r,j}^{(\alpha)}\}_{\alpha = 0,1,\ldots}$ for an element of row r, $A_r(z) = B_{r,1}(z)$ as the generating function for the end element, and $\Gamma_{r,k}(z)$ as the generating function for the sequence $\{\lambda_{r,k}^{(\alpha)}\}_{\alpha = 0,1,\ldots}$, where $\lambda_{r,k}^{(\alpha)} = \lambda_{r,k}^{\alpha}$, Eq. (B13) becomes

$$B_{r,j}(z) - b_{r,j}^{(\alpha=0)} = \frac{z}{r+1} \sum_{k=1}^{r} \eta_{r,jk} [\Gamma_{r,k}(z) A_{r-1}(z)], \quad 1 \leqslant r \leqslant N, \tag{4.14}$$

or

$$B_{r,j}(z) = \Phi_{r,j}(z) A_{r-1}(z), \quad 2 \leqslant r \leqslant N, \tag{4.15}$$

where

$$\Phi_{r,j}(z) = \frac{z}{r+1} \sum_{k=1}^{r} \eta_{r,jk} \Gamma_{r,k}(z). \tag{4.16}$$

In particular, the end element has the relation

$$A_r(z) = \Phi_{r,1}(z) A_{r-1}(z), \quad 2 \leqslant r \leqslant N. \tag{4.17}$$

Note that the condition $r \geqslant 2$ is required in Eqs. (4.15) and (4.17) because $b_{r,j}^{(\alpha=0)}$ vanishes only for $r \geqslant 2$.

Equation (4.17) yields an explicit expression for $A_r(z)$ because Eq. (4.16) for $\Phi_{r,j}(z)$ can be reduced to (see Appendix A)

$$\Phi_{r,j}(z) = w^j \frac{(1+w^{r+1-2j})}{(1+w^{r+1})}, \quad r \geqslant 1, \tag{4.18}$$

where w is defined by

$$w = \frac{1-(1-z^2)^{1/2}}{z}. \tag{4.19}$$

Moreover, because Eqs. (3.7a) and (3.9a) imply $a_1^{(\alpha=0)} = 1$ and $a_1^{(\alpha)} = 0$ for $\alpha = 1,2,\cdots$, the generating function for the first row is

$$A_1(z) = 1. \tag{4.20}$$

When combined with Eq. (4.17), this yields

$$A_r(z) = \prod_{s=2}^{r} \Phi_{s,1}(z), \quad 2 \leqslant r \leqslant N. \tag{4.21}$$

Inserting Eq. (4.18), we obtain the compact result

$$A_r(z) = w^{r-1} \frac{(1+w)(1+w^2)}{(1+w^r)(1+w^{r+1})}, \quad 1 \leqslant r \leqslant N. \tag{4.22}$$

[Comparison with Eq. (4.20) shows Eq. (4.22) is valid down to $r = 1$.] This relation can be quickly verified for $r = 2$ because from Eq. (3.7b) or Fig. 3, the elements of the second row evolve geometrically, with $a_2^{(\alpha=0)} = 0$ and $a_2^{(\alpha)} = 2^{-\alpha}$ for $\alpha = 1,2,\cdots$, so that $A_2(z) = z/(2-z)$.

We can now obtain the row elements $b_{r,j}^{(\alpha)}$ for $2 \leqslant r \leqslant N$ from Eq. (4.15) and the inversion formula (4.13). The integral form in the variable w [see Eq. (4.19)] is

$$b_{r,j}^{(\alpha)} = \frac{1}{2\pi i} \times \oint_{C'} \frac{(1-w^2)(1+w^2)^{\alpha}(1+w)w^j(1+w^{r+1-2j})}{2^{\alpha} w^{\alpha-r+3}(1+w^{r-1})(1+w^r)(1+w^{r+1})}\, dw, \tag{4.23}$$

where C' follows a counterclockwise ellipse enclosing the origin and contained in the unit circle. The substitution $w = 1/y$ shows that C' can be replaced by C'', a clockwise ellipse enclosing the unit circle. Since the integrand has simple poles on the unit circle, it follows by combination and deformation of these contours that the integral is the sum of the residues of the integrand multiplied by $-\frac{1}{2}$. Evaluating the residues at the three sets of roots of $1 + w^{r-1} = 1 + w^r = 1 + w^{r-1} = 0$ using the identity

$$\frac{1}{(1+w^{r-1})(1+w^r)(1+w^{r+1})} = \frac{1}{(1-w^2)(1-w)} \times \left\{ \frac{1}{1+w^{r-1}} - \frac{w(1+w)}{1+w^r} + \frac{w^3}{1+w^{r+1}} \right\}, \tag{4.24}$$

we find that the row elements for $2 \leqslant r \leqslant N$ are exactly

$$b_{r,j}^{(\alpha)} = h_{r,j,-}^{(\alpha)} - 2h_{r,j}^{(\alpha)} + h_{r,j,+}^{(\alpha)}, \quad \alpha \geqslant 0, \ 2 \leqslant r \leqslant N, \tag{4.25}$$

where

$$h_{r,j,-}^{(\alpha)} = \frac{1}{r-1} \sum_{k=0}^{r-2} \cos^{\alpha}\varphi_k \cot\left(\frac{\varphi_k}{2}\right) \times \sin([j-1]\varphi_k), \quad \varphi_k = \frac{(2k+1)\pi}{r-1}, \tag{4.26a}$$

$$h_{r,j}^{(\alpha)} = \frac{1}{r} \sum_{k=0}^{r-1} \cos^{\alpha}\psi_k \cot\left(\frac{\psi_k}{2}\right) \tfrac{1}{2}\{\sin(j\psi_k) + \sin([j-1]\psi_k\}, \quad \psi_k = \frac{(2k+1)\pi}{r}, \tag{4.26b}$$

and

$$h_{r,j,+}^{(\alpha)} = \frac{1}{r+1} \sum_{k=0}^{r} \cos^{\alpha}\theta_k \cot\left(\frac{\theta_k}{2}\right) \sin(j\theta_k), \quad \theta_k = \frac{(2k+1)\pi}{r+1}. \tag{4.26c}$$

We note that Eq. (4.25) is also valid for $r = 1$ provided the first term is dropped. This follows because Eq. (4.15) turns out to be correct for $r = 1$ when coupled with the substitution of Eqs. (4.18) and (4.22). More directly, Fig. 3 shows that the first row element is

$$b_{1,1}^{(\alpha)} = \begin{cases} 1, & \alpha = 0, \\ 0, & \alpha = 1,2,\cdots. \end{cases} \tag{4.27}$$

B. The row $r = N + 1$

The end element $b_{N+1,1}^{(\alpha)} = a_{N+1} = b_{N+1,N+1}^{(\alpha)}$ for the last row requires a separate analysis because its recursion formula is different. Rewriting Eq. (3.8) and using the notation of Eq. (4.1), we have

$$a_{N+1}^{(\alpha+1)} = a_{N+1}^{(\alpha)} + \tfrac{1}{2}a_N^{(\alpha)}, \quad \alpha \geqslant 0. \tag{4.28}$$

Therefore,

$$2a_{N+1}^{(\alpha+1)} = \sum_{\alpha'=0}^{\alpha} a_N^{(\alpha')}. \tag{4.29}$$

In terms of generating functions,

$$2A_{N+1}(z) = A_N(z)z/(1-z), \tag{4.30}$$

where $A_N(z)$ is given by Eq. (4.22) and $A_{N+1}(z)$ is the generating function corresponding to the sequence $\{a_{N+1}^{(\alpha)}\}_{\alpha=0,1,\cdots}$. Because Fig. 3 shows that $2a_{N+1}^{(\alpha)}$ grows to unity as $\alpha \to \infty$, we remove the background level by defining the decaying sequence

$$d_{N+1}^{(\alpha)} = 1 - 2a_{N+1}^{(\alpha)}, \quad \alpha \geqslant 0. \tag{4.31}$$

The corresponding generating function is

$$D_{N+1}(z) = 1/(1-z) - 2A_{N+1}(z). \tag{4.32}$$

Using Eq. (4.22) and working in the variable w defined in Eq. (4.19), we obtain

$$D_{N+1}(z) = \frac{(1+w^2)(1-w^N)(1-w^{N+1})}{(1-w)^2(1+w^N)(1+w^{N+1})}. \tag{4.33}$$

The inversion formula Eq. (4.13) then yields

$$d_{N+1}^{(\alpha)} = \frac{1}{2\pi i} \oint_{C' \text{ or } C''} \frac{(1-w^2)(1-w^N)(1-w^{N+1})(1+w^2)^{\alpha}}{2^{\alpha} w^{\alpha+1}(1-w)^2(1+w^N)(1+w^{N+1})}\, dw. \tag{4.34}$$

As discussed for Eq. (4.23), the location of the contours C' and C'' reduces the integral to

$$d_{N+1}^{(\alpha)} = -\frac{1}{2} \sum_{\substack{\text{residues} \\ \text{on } |w|=1}} (\text{integrand}). \tag{4.35}$$

This gives the final result

$$1 - 2b^{(\alpha)}_{N+1,1} = 1 - 2a^{(\alpha)}_{N+1} = d^{(\alpha)}_{N+1} = h^{(\alpha)}_{N+1} - h^{(\alpha)}_{N}, \quad \alpha \geqslant 0, \tag{4.36}$$

where

$$h^{(\alpha)}_{r} = -\frac{1}{2^{\alpha} r} \sum_{w^{r+1}=0} \frac{(1+w)^2}{(1-w)^2} \frac{(1+w^2)^{\alpha}}{w^{\alpha}} = \begin{cases} \dfrac{1}{r} \displaystyle\sum_{k=0}^{r-1} \cot^2\left[\frac{(2k+1)\pi}{2r}\right] \cos^{\alpha}\left[\frac{(2k+1)\pi}{r}\right], & \alpha \geqslant 1, \\ r-1, & \alpha = 0. \end{cases} \tag{4.37}$$

C. Sum rules for the row elements

According to Eq. (4.12), the sum of the sequence $\{a^{(\alpha)}\}_{\alpha=0,1,\cdots}$ is given by the value at $z=1$ of the associated generating function $A(z)$. Using the generating functions obtained above for the row elements, we find

$$\sum_{\alpha=0}^{\infty} b^{(\alpha)}_{rj} = 1, \quad 1 \leqslant r \leqslant N,\ j = 1,2,\ldots,r, \tag{4.38}$$

and

$$\sum_{\alpha=0}^{\infty} \{1 - 2b^{(\alpha)}_{N+1,1}\} = \tfrac{1}{2}N(N+1). \tag{4.39}$$

5. THE DISCRETE INTRACHAIN MEMORY FUNCTION

In the preceding sections we defined the intrachain memory function, $S(t)$ in terms of the propagator $\Gamma(\underline{r}|\underline{r}';t)$ for the end-to-end vector $\underline{r}$. We then reexpressed $S(t)$ in terms of the Fourier transform of the propagator, the end-to-end time structure factor $g_{1,N+1}(\underline{k}|\underline{r}';t)$ [see Eq. (2.7)]. These were then written as expectation values using the conditional distribution function $\Psi(\{\underline{R}\}|\underline{r}'=0;t)$, defined on the full configuration space of the primitive chain, and determined by the polymer dynamics.

In the case of reptating polymer chains, our recursion solution for the strucure factor has the form shown in Eq. (3.6). Insertion of this in Eq. (2.8) gives the *discrete* intrachain memory function

$$\frac{S(t_\alpha)}{\Omega P_{\rm eq}(\underline{r}=0)} = \sigma(t) + 1 = \sum_{r=1}^{N} \left\{ \sum_{j=1}^{r} b^{(\alpha)}_{rj} K_{rj}(N,t_0) \right\} + 2b^{(\alpha)}_{N+1,1} K_{N+1,1}(N,t_0), \tag{5.1}$$

which is defined at the hopping times $t_\alpha = \alpha\Delta t + t_0$, $\alpha = 0,1,2,\cdots$. The row elements $b^{(\alpha)}_{rj}$ are determined by the dynamics of reptation and were computed in Sec. 4. The initial conditions for the recursion solution are contained in the $\underline{k}$-space integrals

$$K_{rj}(N,t_0) = \frac{1}{P_{\rm eq}(\underline{r}=0)} \frac{1}{(2\pi)^3} \times \int d\underline{k}\, C(\underline{k})^{r-1} g_{jj+N+1-r}(\underline{k}|\underline{r}'=0;t_0), \tag{5.2}$$

where $C(\underline{k})$ is given in Eq. (3.3). For the row $r = N+1$, the $\underline{k}$-space-integral contains $g_{1,1}$, which is unity [see Eq. (3.1a)] for all t, so that for the last term in Eq. (5.1) we immediately have

$$K_{N+1,1}(N,t_0) = 1. \tag{5.3}$$

The K_{rj} for $r \leqslant N$ are discussed in the next subsection.

By "discrete" we refer to the discrete nature of the primitive chain in the reptation model, which consists of N links, and which diffuses according to the hopping model for chain dynamics, Eq. (2.9). The discrete aspect of the problem will be discussed further in Sec. 5B, where we compare the actual forms of the memory function for different N with the limiting, continuum form for large N. Before this is done, we must calculate the initial condition.

A. The initial condition at $t = T_\delta$ for the recursion solution

The $\underline{k}$-space integrals in Eq. (5.2) are the initial condition for the recursion solution, and are determined by the full set of structure factors $\{g_{n,m}{}^{(\alpha=0)}\}_{n,m\in[1,N+1]}$ evaluated using the distribution function $\Psi(\{\underline{R}\}|\underline{r}'=0;t_0)$ at some starting time $t = t_0$. Since this distribution function is known at time $t = 0$ [see Eq. (B19)], it would seem natural to set $t_0 = 0$. However, physical arguments suggest[3] that the polymer chain undergoes Brownian motion in an isotropic medium for $0 \leqslant t \leqslant T_\delta$, and set $t_0 = T_\delta$. The time for crossover to reptational dynamics, T_δ, is defined by

$$\langle r^2 \rangle_{\Gamma(\underline{r}|\underline{r}'=0;t=T_\delta)} = a^2 \tag{5.4}$$

and represents the time for the ends of the chain to move apart the average distance, a, between entanglements.

We describe the polymer dynamics in the absence of entanglements using the Rouse model, which is described in Appendix B. We express $\Psi(\{\underline{R}\}|\underline{r}'=0;t)$ in terms of the known $\Psi(\{\underline{R}\}|\underline{r}'=0;t=0)$, using the Green's function in Eq. (B16) for the diffusion equation of a Rouse chain, Eq. (B4). Since this gives the time evolution of the distribution function, we can calculate the structure factors, the propagator and the intrachain memory function for $0 \leqslant t \leqslant T_\delta$ (see Appendix B).

In particular, the set of equal-time structure factors $\{g_{n,m}{}^{(\alpha=0)}\}_{n,m\in[1,N+1]}$ which define the initial condition for the recursion solution at $t = t_0 = T_\delta$ are given in Eq. (B23). Inserting this result into Eq. (5.2) and using Eq. (B32), we find that the $\underline{k}$-space integrals for $r \leqslant N$ are

$$K_{rj}(N,t_0) = [1 - (1-(r-1)/N)^2 \rho^2_{jj+N+1-r}(t_0)]^{-3/2}, \quad r \leqslant N,\ j = 1,2,\ldots r, \tag{5.5}$$

where $\rho_{n,m}(t)$ is the normalized correlation function

$$\rho_{n,m}(t) = \frac{\langle [\underline{R}_m(0) - \underline{R}_n(0)] \cdot \underline{r}(t) \rangle}{|n-m|a^2}, \quad n < m \tag{5.6}$$

of a free Rouse chain, given in Eq. (B26).

The crossover time T_δ where these initial conditions are computed is determined from Eq. (5.4) using the propagator $\Gamma(\underline{r}|\underline{r}';t)_{\rm Rouse}$ from the Rouse model. From Eqs. (B29) and (B30) we obtain

$$1 - \rho^2_{1,N+1}(T_\delta) = 1/N. \tag{5.7}$$

Using the continuum limit of the Rouse model and the resulting small t form of $1 - \rho_{1,N+1}(t)$ in Eq. (B35b), we conclude that

$$T_\delta/\tau_1 \simeq 1/N^2, \tag{5.8}$$

where τ_1 is the fundamental relaxation time of a Rouse chain, given in Eq. (B33).

As discussed in subsection 3 of Appendix B, it is necessary to consider the continuum limit of the Rouse model in order to obtain a memory function which is integrable at $t = 0$. In this limit, the number of links in the primitive chain approaches infinity, and the mean square interbead separation approaches zero, so that the primitive chain becomes continuously flexible while remaining finite in overall size. In the reptation model it is impossible to consider the continuum limit in this manner because the mean square interbead separation, a^2, is bounded below by (and is generally identified with) the average square separation between entanglement junctions along the real chain backbone. Since this is a fundamental constant (independent of molecular weight), the number of links, N, in the reptation model is fixed by the molecular weight of the polymer chain being simulated. Only when the molecular weight of the actual polymer becomes very large does the reptation model reach its continuum limit (see Sec. 5B).

Thus the primitive chain used when the polymer undergoes reptation differs from the one which applies when the Rouse model is in effect. However, because the Rouse model determines the initial conditions of the reptation model, we must make the two primitive chains commensurate at the bead positions $\underline{R}_1, \underline{R}_2, \ldots, \underline{R}_{N+1}$ of the discrete, reptating primitive chain. Thus the Rouse primitive chain is chosen to have the number of links $N'(s,N) = sN$, where $s = 1,2,\ldots$, is the multiplicity over the reptating primitive chain. The nth bead in the reptating chain then corresponds to the Rouse bead with index $n'(s,n) = 1 + s(n-1)$. Denoting the continuum limits of the free Rouse chain correlation function at its commensurate points by

$$\rho^*_{n,m}(t) = \lim_{s\to\infty} \rho_{n'(s,n),m'(s,m)}(t), \tag{5.9a}$$

we find, using Eq. (B26),

$$\rho^*_{n,m}(t) = \frac{1}{1-\mu}\frac{4}{\pi^2}\sum_{q=1(2)}^{\infty}\{\cos \nu q\pi + \cos(\nu-\mu)q\pi\} \times \frac{1}{q^2}\exp\left(\frac{-q^2 t}{\tau_1}\right), \tag{5.9b}$$

where $q = 1(2)$ denotes $q = 1,3,\cdots$,

$$\nu = (n-1)/N \tag{5.9c}$$

and

$$\mu = 1 - (m-n)/N. \tag{5.9d}$$

This is the form of the correlation function that we use in computing our initial condition, Eq. (5.5), at $t = t_0 = T_\delta$.

B. The continuum limit $N\to\infty$

We designate our above solution the *discrete* intrachain memory function because it reflects both the N link structure of the primitive chain, and the hopping model, Eq. (2.9), for the dynamics of reptation. The discrete character of the solution disappears in the limit of a continuously flexible primitive chain, attained by considering polymers of large molecular weight. When $N\to\infty$, we then find that the intrachain memory function acquires a universal form.

To obtain the *continuum* intrachain memory function we consider the various ingredients of the discrete solution for large N, beginning with the initial condition, Eq. (5.5). When measured in units of the fundamental Rouse relaxation time τ_1, the interval $(0,T_\delta)$ for Rouse dynamics becomes negligible as $N\to\infty$ [see Eq. (5.8)]. Consequently, the $g_{n,m}(\underline{k}\,|\underline{r}' = 0;T_\delta)$ entering Eq. (5.2) must reduce to $g_{n,m}(\underline{k}\,|\underline{r}' = 0;t = 0)$, which are the structure factors for a ring Gaussian chain, i.e., a free Gaussian chain subject to the condition of end-to-end contact. Under this condition, the j dependence must go away, so that

$$K_{r,j}(N,T_\delta)\to J_{r,N}, \quad T_\delta\to 0. \tag{5.10}$$

Given a Gaussian ring, the $J_{r,N}$ can be computed directly. Alternately, we just use our general formula (5.5). Because the correlation functions are normalized, i.e.,

$$\rho_{n,m}(0) = 1, \tag{5.11}$$

we immediately obtain

$$J_{r,N} = [\,[(r-1)/N\,](2-(r-1)/N\,)\,]^{-3/2}, \quad 2\leqslant r\leqslant N. \tag{5.12}$$

Consequently, Eq. (5.1) becomes

$$\frac{S(t_\alpha)}{\Omega P_{\rm eq}(\underline{r}=0)} = \sigma(t_\alpha) + 1 = \sum_{r=2}^{N} f_r^{(\alpha)} J_{r,N} + 2b^{(\alpha)}_{N+1,1}, \tag{5.13}$$

where $\alpha = 1,2,\cdots$, and

$$f_r^{(\alpha)} = \sum_{j=1}^{r} b^{(\alpha)}_{r,j}. \tag{5.14}$$

In Eq. (5.13) we have restricted t_α to positive definite values to avoid the singularity of $S(t)$ at $t = 0$.

Equation (4.25) states $b^{(\alpha)}_{r,j} = h^{(\alpha)}_{r,j,-} - 2h^{(\alpha)}_{r,j} + h^{(\alpha)}_{r,j,+}$, where the $h^{(\alpha)}_{r,\cdots}$ are given in Eq. (4.26). It is possible to show that

$$\sum_{j=1}^{r} h^{(\alpha)}_{r,j,-} = h^{(\alpha)}_{r-1}, \tag{5.15a}$$

$$\sum_{j=1}^{r} h^{(\alpha)}_{r,j} = h^{(\alpha)}_{r}, \tag{5.15b}$$

and

$$\sum_{j=1}^{r} h^{(\alpha)}_{r,j,+} = h^{(\alpha)}_{r+1}, \tag{5.15c}$$

where $h^{(\alpha)}_r$ is defined in Eq. (4.37). Therefore, $h^{(\alpha)}_r$ enters the difference formulas for both the intermediate rows and the last row, given by

$$f_r^{(\alpha)} = h^{(\alpha)}_{r-1} - 2h^{(\alpha)}_r + h^{(\alpha)}_{r+1} \tag{5.16a}$$

and

$$2b^{(\alpha)}_{N+1,1} = 1 + h^{(\alpha)}_N - h^{(\alpha)}_{N+1}, \tag{5.16b}$$

where Eq. (5.16b) is just Eq. (4.36) restated.

We next consider the behavior of Eqs. (5.16) for large N. The summand in Eq. (4.37) is $\cot^2(x_k/2)\cdot\cos^\alpha x_k$, where $x_k = (2k+1)\pi/r \in (0,2\pi)$. For $x \ll 1$ we have

$$\cos^\alpha x = e^{\alpha \log\cos x} = e^{-\alpha x^2(1 + x^2/6 + 2x^4/45 + \cdots)/2}. \tag{5.17}$$

Therefore, the width of $\cos^{\alpha} x$ about $x = 0$ is $\Delta x \equiv (2/\alpha)^{1/2}$. If $\alpha \to \infty$, this width becomes very small, so for $0 < x < \Delta x$, we have

$$\cot^2(x/2)\cos^{\alpha} x \simeq (4/x^2)e^{-\alpha x^2/2}. \tag{5.18}$$

If at least one term in the summation of Eq. (4.37) falls in the range $(0,\Delta x)$, as expressed by the condition $x_{k=0} \in (0,\Delta x)$ or

$$\pi^2\alpha/2r^2 \leqslant 1, \tag{5.19}$$

then Eq. (5.18) can be used to reduce Eq. (4.37) to

$$h_r^{(\alpha)} \simeq \frac{8}{\pi^2} r \sum_{p=1(2)}^{\infty} p^{-2} e^{-p^2\pi^2\alpha/2r^2}, \tag{5.20}$$

where $p = 1(2)$ denotes $p = 1,3,\cdots$. Now $\Delta t = a^2/2D$, where D is the tube diffusion coefficient[6] and $\tau_{rep} = L^2/D\pi^2$, where τ_{rep} is the tube disengagement time[6] for an N-link chain of contour length $L = Na$. Therefore, $\pi^2\alpha/2N^2 = t_\alpha/\tau_{rep}$, which expresses time in units of the fundamental relaxation time of the reptation model. Defining

$$\tilde{t} = t/\tau_{rep}, \tag{5.21}$$

and $\mu = r/N$, we rewrite $h_r^{(\alpha)} = N\mu h(\tilde{t}/\mu^2)$, where

$$h(\tilde{t}) = \frac{8}{\pi^2} \sum_{p=1(2)}^{\infty} p^{-2} e^{-p^2\tilde{t}}. \tag{5.22}$$

Converting the differences in Eq. (5.16) to derivatives, we find the row coefficients reduce to

$$f_r^{(\alpha)} \simeq \frac{\partial^2}{\partial r^2} h_r^{(\alpha)} = N^{-1} \frac{\partial^2}{\partial \mu^2}\left[\mu h\left(\frac{\tilde{t}}{\mu^2}\right)\right], \tag{5.23a}$$

and

$$1 - 2b_{N+1,1}^{(\alpha)} \simeq \frac{\partial}{\partial r} h_r^{(\alpha)}\Big|_{r=N} = \frac{\partial}{\partial \mu}\left[\mu h\left(\frac{\tilde{t}}{\mu^2}\right)\right]\Big|_{\mu=1}. \tag{5.23b}$$

Additionally, we use the fact that N is large to reduce $J_{r,N}$ to $[\mu(2-\mu)]^{-3/2}$ [see Eq. (5.12b)]. Converting the sum in Eq. (5.13) to an integral and using the above equations, we obtain the continuum form of the intrachain memory function. In terms of the transient part, $\sigma(t)$ defined in Eq. (2.3), this is

$$\lim_{N\to\infty} \sigma(t) = \int_0^1 \frac{d\mu}{[\mu(2-\mu)]^{3/2}} \frac{\partial^2}{\partial \mu^2}\left[\mu h\left(\frac{\tilde{t}}{\mu^2}\right)\right] - \frac{\partial}{\partial \mu}\left[\mu h\left(\frac{\tilde{t}}{\mu^2}\right)\right]\Big|_{\mu=1}. \tag{5.24}$$

The integral in Eq. (5.24) converges because as $\mu \to 0$, $(\partial^2/\partial\mu^2)[\mu h(\tilde{t}/\mu^2)] \propto \exp(-\tilde{t}/\mu^2)$.

This result can also be obtained using a diffusion equation for the equal time structure factor obtained via the continuum limit of the recursion relations in Sec. 3. The diffusion equation is

$$\frac{\partial g_{n,m}^{(\alpha)}}{\partial \alpha} = \frac{1}{2}\left(\frac{\partial}{\partial n} + \frac{\partial}{\partial m}\right)^2 g_{n,m}^{(\alpha)}, \tag{5.25a}$$

where $\alpha \geqslant 0$, $n < m \in [0,N]$, and with boundary conditions

$$\frac{\partial g_{n,m}^{(\alpha)}}{\partial n} = \kappa g_{n,m} \quad \text{at } n = 0, \tag{5.25b}$$

$$\frac{\partial g_{n,m}^{(\alpha)}}{\partial m} = -\kappa g_{n,m}^{(\alpha)} \quad \text{at } m = N, \tag{5.25c}$$

where $\kappa = \frac{1}{6}k^2a^2$. The initial condition is

$$g_{n,m}^{(\alpha=0)} = \exp[-\kappa|n-m|(1-|n-m|/N)], \tag{5.25d}$$

which can be inferred from Eqs. (B23), (B24), and (5.11).

The continuum limit of the intrachain memory function has been given earlier in Fig. 1 of Ref. 2 and is shown again in Fig. 4. It is possible to show that the short time form is

$$\frac{S(t)}{\Omega P_{eq}(\underline{r}=0)} = \sigma(t) + 1 = 0.676\tilde{t}^{-3/4}[1 + 0.477\tilde{t}^{1/2} + O(\tilde{t})], \quad N\to\infty, \quad \tilde{t} \ll 1, \tag{5.26a}$$

so that the memory function is integrable at $t = 0$ in its continuum limit. The asymptotic expansion of Eq. (5.24) at long times is

$$\sigma(t) = \frac{8}{\pi^2} \sum_{p=1(2)}^{\infty} \frac{1}{p^2} e^{-p^2\tilde{t}}\left\{\frac{3}{2}\frac{1}{p^2\tilde{t}} - \frac{3}{2}\frac{1}{(p^2\tilde{t})^2} + O(t^{-3})\right\} \simeq 1.216\frac{e^{-\tilde{t}}}{\tilde{t}}, \quad N\to\infty, \quad \tilde{t} \gg 1. \tag{5.26b}$$

By contrast, the discrete form of the memory function, for fixed N, has the long-term form

$$\sigma(t) \simeq 1.216e^{-\tilde{t}}/N, \quad t \gg 1, \tag{5.27}$$

as can be shown from Eqs. (5.13) and (4.37).

Figures 4 and 5 show the continuum and discrete forms of the intrachain memory function. Figure 4, for $10^{-1.5} < \tilde{t} < 10^{1.0}$, illustrates the nonuniversal form of the discrete solution for various values of N, even when time variable is scaled in terms of the N-dependent reptational time scale τ_{rep}. Figure 5 compares the short time behavior of the continuum solution with that of the discrete solution, as determined by its initial value, $S_{Rouse}(t = T_\delta)$.

The above discussion suggests conditions for the continuum form of $\sigma(t)$ in Eq. (5.24) to be a good approximation to the discrete form in Eq. (5.13) [obtained from Eq. (5.1) via

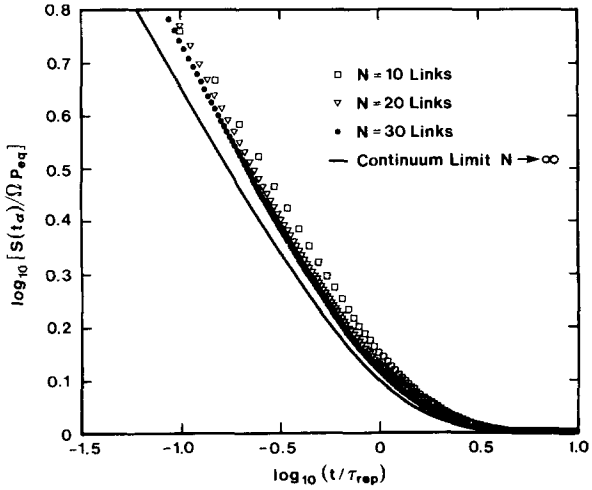


FIG. 4. The intrachain memory function $S(t_\alpha)$ for the N-link bead-rod chain obeying the reptation model. Normalized [see Eq. (2.3)] and defined at discrete times t_α [see Eq. (2.11)], it is shown for a chain with 10, 20, and 30 links. Comparison with the continuum limit for $N \to \infty$ (solid line) shows the nonuniversal character of the discrete N solution. We have used the fundamental reptation time τ_{rep} to remove the N dependence of the time scale.

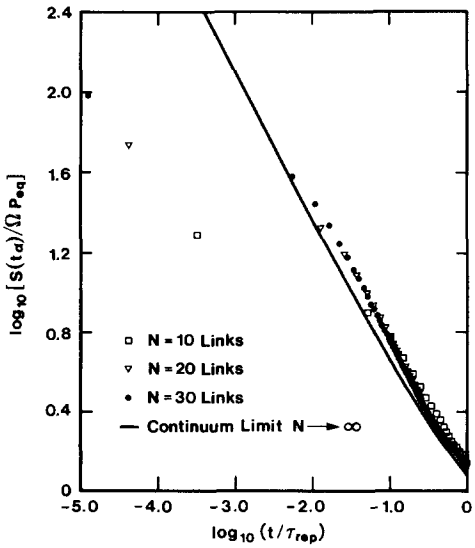


FIG. 5. This extension of Fig. 4 illustrates the short-time behavior of the intrachain memory function at $\alpha = 0,1,2,\cdots$ for a reptating chain with 10, 20, and 30 links. The initial value $S(t_{\alpha=0})$ is the intrachain memory function of a Rouse chain computed at the crossover time T_δ (see Sec. 5A). The solid line shows the $(t/\tau_{rep})^{-3/4}$ form of the continuum limit for $N\to\infty$, $t\to 0$.

$t_0 = T_\delta \to 0$]. To achieve Eq. (5.18), the width $\Delta x = (2/\alpha)^{1/2}$ must be small, say less than 0.1. Using $\pi^2\alpha/2N^2 = \tilde{t}$ this can be expressed by

$$\tilde{t} > \tilde{t}_{\min} = (10\pi/N)^2. \tag{5.28}$$

An additional constraint is Eq. (5.19), which can be rewritten in terms of $\mu = r/N$ as $\mu \gtrsim \tilde{t}^{1/2}$. Although $r\in[2,N]$ implies $\mu\in[0,1]$, so that this relation seems impossible to satisfy for nonzero $\tilde{t}$, examination of the integrand in the continuum form, Eq. (5.24) shows that the dominant contribution in the sum over rows (or integral over μ) occurs at $\mu_0 \simeq \tilde{t}^{1/2}$. This suggests that if μ_0 lies within the domain of integration [0,1], i.e., if

$$\tilde{t} < 1, \tag{5.29}$$

then Eq. (5.19) will be satisfied for the dominant terms. Combining Eqs. (5.28) and (5.29), we see that the continuum and discrete forms of $\sigma(t)$ should agree with times $\tilde{t}$ such that

$$\tilde{t}_{\min} < \tilde{t} < 1. \tag{5.30}$$

This relation can only be satisfied when $\tilde{t}_{\min} < 1$, which by Eq. (5.28) is equivalent to a lower limit on the number of chain links:

$$N > 30 \tag{5.31}$$

in order for the discrete form of $\sigma(t)$ to begin to approach the continuum form.

ACKNOWLEDGMENTS

The authors would like to thank Mr. L. M. Marks for showing us the generating function method. We are also grateful to Dr. K. M. Hong for details of the continuum solution, as well as the evaluation of the contour integrals in this paper.

One of us (D. A. B.) thanks the Natural Sciences and Engineering Research Council of Canada for an Industrial Research Fellowship.

APPENDIX A: THE GENERATING FUNCTION $\Phi_{r,j}(z)$

Our earlier analysis shows that the generating function $B_{r,j}(z)$ for the row element $b_{r,j}^{(\alpha)}$ is given by Eq. (4.15) in terms of the generating function $\Phi_{r,j}(z)$ defined in Eq. (4.16). We first require the generating function $\Gamma_{r,k}(z)$, corresponding to the sequence $\{\lambda_{r,k}^{(\alpha)}\}_{\alpha=0,1,\cdots}$. Because $\lambda_{r,k}^{(\alpha)} = \lambda_{r,k}^{\alpha}$, we have a geometric series:

$$\Gamma_{r,k}(z) = \sum_{\alpha=0}^{\infty} \lambda_{r,k}^{\alpha} z^\alpha = \frac{1}{1-\lambda_{r,k}z}. \tag{A1}$$

Therefore, Eq. (4.16) becomes

$$\Phi_{r,j}(z) = \frac{z}{r+1} \sum_{k=1}^{r} \frac{\eta_{r,jk}}{1-\lambda_{r,k}z}. \tag{A2}$$

The coefficient $\eta_{r,jk}$ is given in Eq. (4.11), which can be rewritten

$$\eta_{r,jk} = \begin{cases} 0, & k = 0,2,\cdots, \\ \cos[(j-1)k\pi/(r+1)] - \cos[(j+1)k\pi/(r+1)], & k = 1,3,\cdots. \end{cases} \tag{A3}$$

Using Eq. (4.8b) and defining $R = r+1$, Eq. (A2) becomes

$$\frac{\Phi_{r,j}(z)}{z} = \frac{1}{R} \sum_{k=1(2)}^{R-1} \frac{\cos[(j-1)k\pi/(r+1)] - \cos[(j+1)k\pi/(r+1)]}{1 - z\cos(\pi k/R)}, \tag{A4}$$

where $k = 1(2)$ denotes $k = 1,3,\cdots$. The sum in Eq. (A4) can be done explicitly using the following identity due to Montroll[9]:

$$\frac{1}{R} \sum_{k=0}^{R-1} \frac{\exp(2lk\pi i/R)}{1 - z\cos(2\pi k/R)} = P_R(l,z), \tag{A5}$$

where

$$P_R(l,z) = (w^l + w^{R-l})/(1-w^R)(1-z^2)^{1/2}, \tag{A6}$$

and

$$w = [1 - (1-z^2)^{1/2}]/z. \tag{A7}$$

Equation (A5) implies

$$\frac{1}{R} \sum_{k=0}^{2R-1} \frac{\exp(lk\pi i/R)}{1 - z\cos(k\pi/R)} = 2P_{2R}(l,z) \tag{A8}$$

and

$$\frac{1}{R} \sum_{k=0(2)}^{2R-2} \frac{\exp(lk\pi i/R)}{1 - z\cos(k\pi/R)} = P_R(l,z). \tag{A9}$$

Subtracting, we find

$$\frac{1}{R} \sum_{k=1(2)}^{2R-1} \frac{\exp(lk\pi i/R)}{1 - z\cos(k\pi/R)} = 2P_{2R}(l,z) - P_R(l,z), \tag{A10}$$

or

$$\frac{2}{R} \sum_{k=1(2)}^{R-1} \frac{\cos(lk\pi/R)}{1 - z\cos(k\pi/R)} + \frac{\epsilon_R^l \exp(l\pi i)}{R(1+z)} = 2P_{2R}(l,z) - P_R(l,z), \tag{A11}$$

where $\epsilon_R^l = 1$ when R is odd and 0 otherwise. Using $l = j-1$ and $l = j+1$ in Eq. (A4), we obtain

$$\frac{\Phi_{r,j}(z)}{z} = P_{2R}(j-1,z) - P_{2R}(j+1,z) - \tfrac{1}{2}[P_R(j-1,z) - P_R(j+1,z)]. \quad \text{(A12)}$$

After some algebra, Eqs. (A6) and (A7) yield the key result

$$\Phi_{r,j}(z) = w^j(1 + w^{r+1-2j})/(1 + w^{r+1}), \quad \text{(A13)}$$

where $w = [1 - (1 - z^2)^{1/2}]/z$.

APPENDIX B: DYNAMICS OF A ROUSE CHAIN

In order to obtain the initial condition at $t = T_\delta$ for the recursion solution (see Sec. 5A), we describe the polymer dynamics for $0 \leqslant t \leqslant T_\delta$ using the Rouse model,[4] which defines a diffusion equation, in the absence of entanglements, for the configurational distribution function $\Psi(\{\underline{R}\};t)$ in the full configuration space of the primitive chain (see Sec. 2). This enables us to calculate the equal-time structure factor, the propagator and the intrachain memory function (see subsections 1–3 of this appendix). First, we review the Rouse model.

In the Rouse model, a purely entropic free energy is introduced by comparing the distribution function with its equilibrium value $\Psi_{eq}(\{\underline{R}\})$. The free energy is

$$\mathcal{F}(\{\underline{R}\};t) = k_B T \ln[\Psi(\{\underline{R}\};t)/\Psi_{eq}(\{\underline{R}\})], \quad \text{(B1)}$$

where k_B is Boltzmann's constant and T is the absolute temperature. The equilibrium distribution function for the N-link primitive chain is a product of Gaussians and derives from the assumption that the intervening monomers in each section or "blob"[6] of the real chain between points $\underline{R}_i$ and $\underline{R}_{i+1}$ obey random walk statistics for any given link vector $\underline{u}_i = \underline{R}_{i+1} - \underline{R}_i$. It is

$$\Psi_{eq}(\{\underline{R}\}) = V^{-1} \prod_{j=1}^{N} \varkappa_\beta(\underline{u}_j), \quad \text{(B2a)}$$

where V is the volume and the normal distribution in three dimensions is

$$\varkappa_\beta(u) = (\beta/\pi)^{3/2} \exp(-\beta u^2). \quad \text{(B2b)}$$

The mean squared blob size is $\langle u^2 \rangle_{\varkappa_\beta(\underline{u})} = 3/2\beta$, also denoted a^2. The free energy of Eq. (B1) implies a Brownian force on the ith bead, which represents the driving impetus towards equilibrium. The force is $\underline{F}_i^B = -\nabla_{\underline{R}_i} \mathcal{F}(\{\underline{R}\};t)$, and is balanced during the relaxation towards equilibrium by a Stokes' drag force. For a viscous background medium which is at rest, the force is $\underline{F}_i^S = -\zeta \langle \underline{\dot{R}}_i \rangle$, where ζ is the net friction coefficient of the group of monomers being simulated by the bead. Neglecting transient effects by dropping the inertial term from Newton's second law, leaving a force balance condition $\underline{F}_i^B + \underline{F}_i^S = 0$, we can solve for the mean bead velocity to give $\langle \underline{\dot{R}}_i \rangle = -\zeta^{-1} \nabla_{\underline{R}_i} \mathcal{F}(\{\underline{R}\};t)$. This result enables us to substitute for $\langle \underline{\dot{R}}_i \rangle$ in the equation of continuity in configuration space,

$$\partial_t \Psi(\{\underline{R}\};t) + \sum_{i=1}^{N+1} \nabla_{\underline{R}_i} \cdot (\langle \underline{\dot{R}}_i \rangle \Psi(\{\underline{R}\};t)) = 0, \quad \text{(B3)}$$

to obtain a diffusion equation for Ψ. In relative coordinates $\{\underline{u}\}$ this is[10]

$$\partial_t \Psi - \sum_{j=1}^{N} \nabla_{\underline{u}_j} \cdot \left(\zeta^{-1} k_B T \Psi \sum_{k=1}^{N} A_{j,k}^R \nabla_{\underline{u}_k} \ln\left[\frac{\Psi}{\Psi_{eq}}\right] \right) = 0, \quad \text{(B4)}$$

where $\mathbf{A}^R$ is the $N \times N$ Rouse matrix with components

$$A_{j,k}^R = 2\delta_{j,k} - \delta_{j,k-1} - \delta_{j,k+1}. \quad \text{(B5)}$$

The orthogonal matrix Ω with components

$$\Omega_{i,j} = [2/(N+1)]^{1/2} \sin[ij\pi/(N+1)], \quad \text{(B6)}$$

diagonalizes the Rouse matrix in the form

$$\Xi = \Omega \cdot \mathbf{A}^R \cdot \Omega, \quad \text{(B7)}$$

where the diagonal elements of Ξ are

$$\xi_j = 4 \sin^2[j\pi/2(N+1)], \quad j = 1,2,...,N. \quad \text{(B8)}$$

The corresponding coordinate transformation of the link vectors is

$$\underline{v}_j = \sum_{k=1}^{N} \Omega_{j,k} \underline{u}_k, \quad \text{(B9)}$$

which reduces the diffusion equation in the Rouse model to

$$\partial_t \varphi_j - \zeta^{-1} k_B T \xi_j \nabla_{\underline{v}_j} \cdot (\varphi_j \nabla_{\underline{v}_j} \ln[\varphi_j/\varkappa_\beta(\underline{v}_j)]) = 0, \quad \text{(B10)}$$

for $j = 1,2,...,N$. In achieving Eq. (B10), the distribution function has the assumed form

$$\Psi(\{\underline{R}\};t) = V^{-1} \prod_{j=1}^{N} \varphi_j(\underline{v}_j). \quad \text{(B11)}$$

Wilemski and Fixman[11] have shown that the diffusion equation (B10) has the Green's function

$$G_j(\underline{v}_j | \underline{v}_j^0;t) = \varkappa_{\beta_j}(\underline{v}_j - \underline{v}_j^0 \gamma_j(t)), \quad \text{(B12)}$$

where

$$\gamma_j(t) = \exp(-t/\tau_j) \quad \text{(B13)}$$

and

$$\beta_j(t) = \beta/[1 - \gamma_j(t)^2] \quad \text{(B14)}$$

are determined by a relaxation time τ_j such that

$$\tau_j^{-1} = (3k_B T/\zeta a^2)\xi_j. \quad \text{(B15)}$$

Consequently, the full Green's function for the Rouse model is

$$G_{\text{Rouse}}(\{\underline{u}\}|\{\underline{u}^0\};t) = \prod_{j=1}^{N} G_j(\underline{v}_j|\underline{v}_j^0;t). \quad \text{(B16)}$$

Therefore, the distribution function $\Psi(\{\underline{R}\}|\underline{r}';t)$ which evolves from the initial condition $\Psi(\{\underline{R}\}|\underline{r}';t=0)$ is

$$\Psi(\{\underline{R}\}|\underline{r}';t) = \int d\{\underline{u}^0\} \times G_{\text{Rouse}}(\{\underline{u}\}|\{\underline{u}^0\};t)\Psi(\{\underline{R}^0\}|\underline{r}';t=0). \quad \text{(B17)}$$

The equal-time structure factor, defined in Eq. (2.7), for a chain subject to Rouse dynamics, is then given by

$$g_{n,m}(\underline{k}|\underline{r}';t)_{\text{Rouse}} = V \int d\{\underline{u}\} \int d\{\underline{u}^0\} \exp(i\underline{k} \cdot [\underline{u}_n + \cdots + \underline{u}_{m-1}]) \times G_{\text{Rouse}}(\{\underline{u}\}|\{\underline{u}^0\};t)\Psi(\{\underline{R}^0\}|\underline{r}';t=0). \quad \text{(B18)}$$

To compute the equal-time structure factor $g_{n,m}(\underline{k}|\underline{r}'=0;t)_{\text{Rouse}}$ based on the condition of chain closure at $t = 0$, we need the initial distribution $\Psi(\{\underline{R}\}|\underline{r}'=0;t=0)$. Because the chain is Gaussian subject only to the require-

ment $\underline{r}' = 0$, or $\underline{R}_{N+1} - \underline{R}_1 = 0$ at $t = 0$, we have[12]

$$\Psi(\{\underline{R}\}|\underline{r}' = 0;t = 0) = \frac{\Psi_{eq}(\{\underline{R}\})\delta(\underline{r}' - [\underline{R}_{N+1} - \underline{R}_1])}{P_{eq}(\underline{r}')}\Big|_{\underline{r}'=0}, \tag{B19}$$

where $\Psi_{eq}(\{\underline{R}\})$ is given by Eq. (B2a) and $P_{eq}(\underline{r})$ is the equilibrium distribution of the end-to-end vector

$$P_{eq}(\underline{r}) = \langle\delta(\underline{r} - [\underline{R}_{N+1} - \underline{R}_1])\rangle_{\psi_{eq}(\{\underline{R}\})} = \wp_{\beta/N}(\underline{r}). \tag{B20}$$

Using Eq. (B9) and the relation $\underline{R}_{N+1} - \underline{R}_1 = \underline{u}_1 + \underline{u}_2 + \cdots \underline{u}_n$, the initial condition can be rewritten

$$\Psi(\{\underline{R}\}|\underline{r}' = 0;t = 0) = V^{-1}\left(\frac{\beta}{\pi N}\right)^{-3/2}\frac{1}{(2\pi)^3} \times \int d\underline{k}\prod_{j=1}^{N}\wp_\beta(\underline{v}_j)\exp(i\underline{k}\cdot\underline{\mu}_j\underline{v}_j), \tag{B21}$$

where

$$\mu_j = \sum_{l=1}^{N}\Omega_{j,l}. \tag{B22}$$

1. The equal-time structure factor of a Rouse chain

Inserting Eq. (B21) in Eq. (B18) and performing the integrations then yields the desired equal-time structure factor

$$g_{n,m}(\underline{k}\,|\underline{r}' = 0;t)_{\text{Rouse}} = \exp(-N\nu_{n,m}(t)k^2/4\beta), \tag{B23}$$

where

$$\nu_{n,m}(t) = \begin{cases} 0, & n = m, \\ (|n-m|/N)\{1 - (|n-m|/N)\rho^2_{n,m}(t)\}, & n \neq m, \end{cases} \tag{B24}$$

which is defined in terms of a normalized correlation function of a *free* Rouse chain

$$\rho_{n,m}(t) = \frac{\langle[\underline{R}_m(0) - \underline{R}_n(0)]\cdot\underline{r}(t)\rangle}{|n-m|a^2}, \quad n < m, \tag{B25}$$

given by

$$\rho_{n,m}(t) = \frac{1}{|n-m|}\sum_{q=1}^{N}\mu_q^{(n,m)}\mu_q\gamma_q(t), \tag{B26}$$

where

$$\mu_q^{(n,m)} = \sum_{l=n}^{m-1}\Omega_{q,l}. \tag{B27}$$

Using Eq. (B6), we calculate the coefficients $\mu_q^{(n,m)}$ to be

$$\mu_q^{(n,m)} = \left(\frac{2}{N+1}\right)^{1/2} \times \frac{\sin(nq\pi/(N+1)) - \sin(mq\pi/(N+1)) - \sin([n-1]q\pi/(N+1)) + \sin([m-1]q\pi/(N+1))}{2(1-\cos[q\pi/(N+1)])}, \tag{B28a}$$

with

$$\mu_q = \mu_q^{(1,N+1)} = \begin{cases} (2/(N+1))^{1/2}\dfrac{\sin(q\pi/N+1)}{1-\cos(q\pi/N+1)}, & q \text{ odd}, \\ 0, & q \text{ even}. \end{cases} \tag{B28b}$$

2. The propagator and memory function of a Rouse chain

We can now compute the propagator $\Gamma(\underline{r}|\underline{r}' = 0;t)$. Inserting the end-to-end equal time structure factor $g_{1,N+1}(\underline{k}\,|\underline{r}' = 0;t)$ from Eq. (B23) into Eq. (2.7), we find

$$\Gamma(\underline{r}|\underline{r}' = 0;t)_{\text{Rouse}} = \wp_{\beta_\Gamma(t)}(\underline{r}), \tag{B29}$$

where

$$\beta_\Gamma(t) = \beta/N\nu_{1,N+1}(t) = \beta/N\{1 - \rho^2_{1,N+1}(t)\}. \tag{B30}$$

The intrachain memory function is then given immediately by Eqs. (2.1) and (B2b) as

$$S(t)_{\text{Rouse}} = \Omega(\beta_\Gamma(t)/\pi)^{3/2} = \Omega P_{eq}(\underline{r} = 0)\{1 - \rho^2_{1,N+1}(t)\}^{-3/2}. \tag{B31}$$

Here we have used

$$P_{eq}(\underline{r} = 0) = (\beta/N\pi)^{3/2}, \tag{B32}$$

which follows from Eq. (B20).

3. The continuum limit of the Rouse model

As in the reptation model, the primitive chain in the Rouse model consists of $N + 1$ identical beads connected by N links at universal joints. The bead positions correspond to positions at uniform intervals of N_b monomers along the real chain backbone, and, if the overall degree of polymerization or monomer number is N_u, then $N = N_u/N_b$.

In contrast with the reptation model, the choice of primitive chain to simulate a given real chain is somewhat arbitrary. For fixed N_u, the longest Rouse relaxation time, τ_1, defined by Eq. (B15), is independent of the "blobsize," N_b, or the number of links, N, in the limit of a continuously flexible primitive chain. In this limit, $N_b \to 0$ and $N \to \infty$, and we have

$$\tau_1 = \zeta_{\text{tot}}\langle r^2\rangle_{eq}/3\pi^2 k_B T, \tag{B33}$$

where $\zeta_{\text{tot}} = N\zeta$ is the net friction coefficient of the real chain and $\langle r^2\rangle_{eq} = Na^2$ is the mean-square end-to-end separation for conformational equilibrium.

Before taking the continuum limit, we consider the short time behavior of the propagator and memory function of a discrete Rouse chain with large but finite N. From Eqs. (B26) and (B33), we find

$$1 - \rho_{1,N+1}(t) \simeq \left\{\frac{8}{\pi^2}\frac{N}{N+1}\sum_{q=1(2)}^{N}\cos^2\frac{q\pi}{2(N+1)}\right\}\frac{t}{\tau_1}, \quad t \to 0, \tag{B34a}$$

where $q = 1(2)$ denotes $q = 1,3,\cdots$. Consequently, for fixed N, using Eq. (B31),

$$S(t)_{\text{Rouse}} = \Omega P_{\text{eq}}(\underline{r}=0) A_N (t/\tau_1)^{-3/2}, \quad t \to 0, \tag{B34b}$$

where A_N is an N-dependent coefficient.

Physical considerations show[3] that $S(t)$ should be integrable at $t = 0$. However, $S(t)_{\text{Rouse}}$ in Eq. (B34b) is not integrable for fixed N. To eliminate this artifact and obtain dynamical functions which are valid for $t \to 0$, we consider the $N \to \infty$ limit before the $t \to 0$ limit and find

$$\lim_{N\to\infty} \rho_{1,N+1}(t) \simeq \frac{8}{\pi^2} \sum_{q=1(2)}^{\infty} \frac{1}{q^2} \exp\left(-q^2 \frac{t}{\tau_1}\right), \tag{B35a}$$

which has[13] the following form for small t

$$\lim_{N\to\infty} \{1 - \rho_{1,N+1}(t)\} \simeq \frac{4}{\pi^{3/2}} \left(\frac{t}{\tau_1}\right)^{1/2}, \quad t \to 0, \tag{B35b}$$

so that $S(t)$ is now integrable for $t \to 0$:

$$\lim_{N\to\infty} S(t)_{\text{Rouse}} \simeq \Omega P_{\text{eq}}(\underline{r}=0) \left(\frac{8}{\pi^{3/2}}\right)^{-3/2} \left(\frac{t}{\tau_1}\right)^{-3/4} \tag{B35c}$$

[1] M. A. Winnik, T. Redpath, and D. H. Richards, Macromolecules **13**, 328 (1980).
[2] D. A. Bernard and J. Noolandi, Phys. Rev. Lett. **50**, 253 (1983).
[3] D. A. Bernard and J. Noolandi, J. Chem. Phys., to be submitted.
[4] P. E. Rouse, J. Chem. Phys. **21**, 1272 (1953).
[5] S. F. Edwards, Proc. Phys. **92**, 9 (1967).
[6] P. G. de Gennes, J. Chem. Phys. **55**, 572 (1971).
[7] M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2 **74**, 1789,1802,1818 (1978); **75**, 38 (1979).
[8] J. L. Synge and B. A. Griffith, *Principles of Mechanics* (McGraw-Hill, New York, 1949), pp. 498ff.
[9] E. W. Montroll, Proc. Sym. Appl. Math. **16**, 193 (1964).
[10] R. B. Bird, O. Hassager, R. C. Armstrong, and C. F. Curtis, *Dynamics of Polymeric Liquids* (Wiley, New York, 1977), Vol. 2, pp. 585ff.
[11] G. Wilemski and M. Fixman, J. Chem. Phys. **60**, 866 (1974).
[12] H. Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971), p. 51.
[13] M. Doi, Chem. Phys. **9**, 455 (1975).