Transition probability spaces

Johan G. F. Belinfante

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 8 September 1975)

The set of pure quantum states is described as an abstract space with a geometry determined by transition probabilities. We describe all possible structures for three-dimensional transition probability spaces with less than ten states, as well as some even larger spaces of a certain symmetric type. It is shown that the orthoclosed subspaces of a transition probability space form an atomistic orthomodular poset.

1. INTRODUCTION

In an axiomatic study of quantum mechanics, Mielnik introduced the concept of transition probability space.[1] The set of all pure states of a quantum mechanical system is viewed as an abstract space with a geometry determined by transition probabilities. In general, the states of a transition probability space need not be realizable in a Hilbert space. The existence of a representation of pure states by unit vectors in a Hilbert space restricts the geometric structure of the state space. Mielnik provides numerical criteria under which a transition probability space can be embedded in a Hilbert space. His condition for embedding involves only two-dimensional subspaces holding up to ten states. As Mielnik points out, the two-dimensional subspaces are especially important due to their relation to the superposition principle. Nevertheless, it is also of interest to study the structure of three-dimensional transition probability spaces. We were led to study finite three-dimensional transition probability spaces in our effort to gain insight into the structure defined by the axioms. There are many more structures possible for a three-dimensional transition probability space than for a two-dimensional one. In the present paper, we describe all possible structures for three-dimensional transition probability spaces holding less than ten states. We also prove a number of general theorems about transition probability spaces, using the three-dimensional examples as illustrations.

The axiomatic concept of transition probability is abstracted from the Hilbert space model of quantum theory. In this model, the transition probability between two pure states represented by unit vectors φ and ψ in Hilbert space is $|(\varphi,\psi)|^2$.

There is a one-to-one correspondence between pure states and projection operators with one-dimensional range. We may therefore identify the set of pure states with the set S of all one-dimensional projection operators. If x is the projection operator whose range is the one-dimensional subspace spanned by the unit vector ψ, then $x\varphi=(\psi,\varphi)\psi$. The absolute square of the inner product of two unit vectors can be rewritten as the trace of the product of the corresponding projection operators. Thus, the transition probability $p(x,y)$ between the pure states x and y is given by $p(x,y)=\mathrm{tr}\,xy$. This function satisfies $0\leq p(x,y)\leq 1$ and $p(x,y)=p(y,x)$. Furthermore, $p(x,y)=1$ if and only if $x=y$. Another less trivial property is described in Sec. 2, and is used to define a transition probability space.

The plan of the remainder of the paper is as follows. Section 3 contains some general theorems needed for the construction of the examples as well as for the analysis of the subspace structure. Section 4 is devoted to a detailed analysis of all transition probability spaces of dimension three with less than ten states. In Sec. 5, we use these examples in the course of our analysis of the structure of the partially ordered set of all subspaces of a transition probability space. If we restrict our attention to the orthoclosed subspaces, we obtain an atomistic orthomodular poset. In the final section, we analyze the structure of a certain class of symmetric three-dimensional transition probability spaces.

The axioms for transition probability do not by themselves constitute a complete framework for quantum mechanics. Rather they represent a common core of a number of different axiomatic structures. By studying these axioms in isolation from others, we obtain a better idea of what they imply.

2. AXIOMS

In the sequel, we consider a mapping $p: S\times S\to[0,1]$, where S is an abstract set whose elements are called (pure) states. A subset B of S is a *basis* if every state x in S satisfies

$$\sum_{y\in B} p(x,y)=1.$$

When the basis is an infinite set, the sum over all states in the basis here is to be understood as the least upper bound for sums over all finite subsets of the basis. For short, we often write $p(x,B)$ for the sum of $p(x,y)$ as y ranges over B. Heuristically, we may regard a basis as a sample space for an experiment, the states in a basis being the results or outcomes possible for that experiment.

The mapping p is *symmetric* if $p(x,y)=p(y,x)$ for all states x and y in S. If p is symmetric, then any two bases B_1 and B_2 have the same cardinality, $\#(B_1)=\#(B_2)$. We can define the *dimension* of S to be the cardinality of any basis, provided that at least one basis exists. That all bases hold the same number of states was known to Landsberg[2] in 1947, and was reproved by Mielnik.[1] When S is a finite set, the proof is just that

$$\#(B_1)=\sum_{x\in B_1}\sum_{y\in B_2} p(x,y)=\#(B_2).$$

States x and y are *orthogonal* if $p(x,y)=0$. A set of states is *pairwise orthogonal* if each distinct pair of

states in the set are orthogonal. A pairwise orthogonal set of states is *maximal* if it is not a proper subset of any larger pairwise orthogonal set of states. By Zorn's lemma, every pairwise orthogonal set of states can be extended to a maximally pairwise orthogonal set of states.

The mapping p is said to satisfy the *separation axiom* if we have $p(x,y)=1$ if and only if $x=y$. If the separation axiom holds, then any basis for S must be a maximally pairwise orthogonal set of states. If p is symmetric and satisfies the separation axiom, then orthogonality is a symmetric and irreflexive relation.

A symmetric function p satisfying the separation axiom is a *transition probability* if every maximally pairwise orthogonal subset of S is a basis. A *transition probability space* (S,p) consists of a set S and a transition probability p on S.

It is an immediate consequence of the definition that each state of a transition probability space belongs to at least one basis. We obtain a trivial example of a transition probability space by taking S to be any set, and setting $p(x,y)$ equal to unity for $x=y$, and zero for $x\neq y$. In this case, the whole space is the one and only basis.

The concept of transition probability can be used to characterize completeness for inner product spaces. An inner product space V is complete if and only if $p(x,y)=\mathrm{tr}\,xy$ is a transition probability on the set S of all one-dimensional projection operators on V. This follows immediately from Gudder's theorem[3] that an inner product space is complete if every maximal orthonormal set satisfies the Parseval identity.

In any transition probability space one can introduce a natural topology by defining a metric $d(x,y)$ equal to the least upper bound for $|p(x,z)-p(y,z)|$ as z varies over all states. By the separation axiom, we always have $1-p(x,y)\leq d(x,y)$. The transition probability p is jointly continuous in this metric space topology. In general, a transition probability space need not be a complete metric space.

In the Hilbert space model of quantum mechanics, we have[4]

$$d(x,y)=\operatorname*{lub}_{(z)}\,\mathrm{tr}(x-y)z=\|x-y\|.$$

Hence $d(x,y)$ is the largest eigenvalue of the operator $x-y$. Since $xyx=p(x,y)x$, we have $(x-y)^3=[1-p(x,y)](x-y)$, so this largest eigenvalue is $d(x,y)=\sqrt{1-p(x,y)}$. This formula implies that the quantity $\sqrt{1-p(x,y)}$ satisfies a triangle inequality in the Hilbert space model of quantum mechanics. This triangle inequality is one of Mielnik's embedding criteria. We may also describe $d(x,y)$ as the minimum distance between unit vectors in the ranges of the projection operators x and y.

3. GENERAL CONSIDERATIONS

A subset T of a transition probability space (S,p) is a *subspace* if we obtain a transition probability on T by restricting p to $T\times T$. Symmetry and separation are automatically true in T. Thus, a subset T is a subspace if and only if every maximally pairwise orthogonal subset of T is a basis for T.

Any pairwise orthogonal subset of a transition probability space is a subspace. In particular, the empty set and all singletons are subspaces.

The *orthocomplement* $A^\perp$ of a subset A of a transition probability space is the set of all states which are orthogonal to every state in A. In general, A and $A^\perp$ are disjoint. The orthocomplement of the union of two sets is the intersection of their orthocomplements. If $A\subset B$, then $B^\perp\subset A^\perp$.

In general, the orthocomplement of a subset need not be a subspace. The simplest example of this phenomenon occurs in a three-dimensional space with eight states to be discussed later.

Nevertheless, the orthocomplement of a subspace is a subspace. To see this, we consider first the special case of a pairwise orthogonal set A. If B is a maximally pairwise orthogonal subset in $A^\perp$, then $A\cup B$ is a basis for the whole space. If x is a state in $A^\perp$, then $1=p(x,A\cup B)=p(x,B)$. Thus B is a basis for $A^\perp$, and hence $A^\perp$ is a subspace. To prove more generally that the orthocomplement of any subspace T is a subspace, we need only show that $B^\perp=T^\perp$ for any basis B of T. For this, we note that if y is orthogonal to B, then $B\cup\{y\}$ is pairwise orthogonal, so $p(x,B)+p(x,y)\leq 1$ for any state x. If x belongs to T, then $p(x,B)=1$, so $p(x,y)=0$ and $y\in T^\perp$. Thus $B^\perp\subset T^\perp$. The opposite inclusion also holds, because $B\subset T$.

Any subset A of a transition probability space satisfies $A\subset A^{\perp\perp}$. A subset A is *orthoclosed* if $A=A^{\perp\perp}$. The orthorcomplement of any subset is orthoclosed. The empty set and the whole space are orthoclosed subspaces. In general, we can easily find the orthoclosed subspaces since every orthoclosed subspace is the orthocomplement of a pairwise orthogonal subset. In fact, if T is an orthoclosed subspace, then T is the orthocomplement of any basis for $T^\perp$.

If B is a pairwise orthogonal subset of a transition probability space, then $B^{\perp\perp}$ is the greatest subspace having B as a basis. Indeed, $B^{\perp\perp}$ is a subspace, and B is a basis for $B^{\perp\perp}$ since $B^\perp\cap B^{\perp\perp}=\phi$. If B is a basis for another subspace T, then $B^\perp=T^\perp$, and $T\subset T^{\perp\perp}=B^{\perp\perp}$.

As an application of this, we note that each singleton is an orthoclosed subspace: Any singleton $\{x\}$ is pairwise orthogonal, and since $\{x\}$ is a basis for $\{x\}^{\perp\perp}$, we have $p(x,y)=1$, and hence $x=y$, for all y in $\{x\}^{\perp\perp}$. Thus $\{x\}=\{x\}^{\perp\perp}$.

To determine whether a given subset of a transition probability space is a subspace, it suffices to know the orthogonality relation on the transition probability space. A subset T of a transition probability space is a subspace if and only if $B^\perp=T^\perp$ for every maximally pairwise orthogonal subset B in T. We have already proved this condition is necessary. It is also sufficient. If $B^\perp=T^\perp$ for each maximally pairwise orthogonal subset B in T, then $T\subset B^{\perp\perp}$. Since B is a basis for $B^{\perp\perp}$, we have $p(x,B)=1$ for every state x in T, that is, B is also a basis for T. Hence T is a subspace.

Each subspace T of a transition probability space is itself a transition probability space, and thus has a definite dimension, and $\dim T + \dim T^{\perp} = \dim S$. In fact, any basis for T and any basis for $T^{\perp}$ are disjoint, and their union is a basis for S. Any basis for a subspace T is also a basis for $T^{\perp\perp}$, so T and $T^{\perp\perp}$ have the same dimension. If orthoclosed subspaces T_1 and T_2 have the same finite dimension and satisfy $T_1 \subset T_2$, then $T_1 = T_2$. Here any basis B for T_1 is also a basis for T_2, so $T_1^{\perp} = B^{\perp} = T_2^{\perp}$, and $T_1 = T_2$.

If (S_1, p_1) and (S_2, p_2) are transition probability spaces, then we can define a transition probability p on the disjoint union of S_1 and S_2 by setting $p(x,y) = 0$ if $x \in S_1$ and $y \in S_2$, or $x \in S_2$ and $y \in S_1$, and $p(x,y) = p_i(x,y)$ if both x and y belong to S_i where $i = 1, 2$.

A transition probability space is *irreducible* if it is not the union of two nonempty orthogonal subsets. If a transition probability space is the union of two orthogonal subsets T_1 and T_2, then T_1 and T_2 are orthoclosed subspaces, and $T_1 = T_2^{\perp}$. Therefore, it does not matter whether we frame the definition of irreducibility in terms of orthogonal subsets or in terms of orthogonal subspaces.

The study of transition probability spaces reduces to the study of the irreducible ones. Every transition probability space is the union of a set of mutually orthogonal irreducible subspaces.

We have described above a method for constructing transition probability spaces as disjoint unions. A modification of this method can be used to construct irreducible transition probability spaces. If (S_1, p_1) and (S_2, p_2) are transition probability spaces of the same dimension $d \geq 2$, then the disjoint union of S_1 and S_2 can be made into an irreducible d-dimensional transition probability space by setting $p(x,y) = 1/d$ when $x \in S_1$ and $y \in S_2$, or $x \in S_2$ and $y \in S_1$, and $p = p_i$ when x and y belong to the same S_i, where $i = 1$ or 2. For this construction, it does not matter if the spaces (S_1, p_1) and (S_2, p_2) fail to be irreducible. Thus, for example, we could take them to be pairwise orthogonal. By repeating this construction over and over, we can construct arbitrarily large irreducible transition probability spaces of any finite dimension $d \geq 2$.

While there is no upper limit on the size of an irreducible transition probability space of a given dimension, there does exist a lower limit. There are at least $2d$ states in any irreducible transition probability space of dimension $d \geq 2$. In proving this, we may assume without loss of generality that the number of states is finite, so there are a finite number of bases, $B_1, B_2, \ldots, B_b$. For each integer $n = 1, \ldots, b$, let $I_n = B_1 \cap \cdots \cap B_n$ and $U_n = B_1 \cup \cdots \cup B_n$. Each state x in $I_n \sim I_{n+1}$ is orthogonal to every state in U_n because both belong to at least one of the bases $B_1, \ldots, B_n$. Hence $1 = p(x, B_{n+1}) = p(x, U_{n+1} \sim U_n)$, and summing over all such states x, we find $\#(I_n \sim I_{n+1})$ is equal to $p(I_n \sim I_{n+1}, U_{n+1} \sim U_n)$. On the other hand, since $I_n \sim I_{n+1}$ is pairwise orthogonal, $p(I_n \sim I_{n+1}, y) \leq 1$ for all y, and summing over all states y in $U_{n+1} \sim U_n$, we find the inequality $\#(I_n \sim I_{n+1}) \leq \#(U_{n+1} \sim U_n)$, that is,

$$\#(I_n) + \#(U_n) \leq \#(I_{n+1}) + \#(U_{n+1}).$$

For $n = 1$, we have $I_1 = U_1 = B$, so $\#(I_1) + \#(U_1) = 2d$. Hence $2d \leq \#(I_n) + \#(U_n)$ holds for all n, and in particular for $n = b$. Since the whole space S is irreducible, and since $d \geq 2$, no state can belong to every basis, that is, $I_b = \phi$. On the other hand, every state belongs to some basis, so $U_b = S$. Thus $2d \leq \#(S)$.

A mapping $h: S_1 \to S_2$ from a transition probability space (S_1, p_1) to a transition probability space (S_2, p_2) is a *homomorphism* if $p_1(x,y) = p_2(h(x), h(y))$ for all states x and y in S_1. By the separation axiom, it follows that every homomorphism is one-to-one. The image of any subspace T of S_1 under h is a subspace $h[T]$ of S_2 with the same dimension as T. In particular, $h[S_1]$ is a subspace of S_2 and therefore $\dim S_1 \leq \dim S_2$. In general, the image of an orthoclosed subspace need not be orthoclosed. If h is an onto homomorphism, then h^{-1} is also a homomorphism. We may therefore define an *isomorphism* between transition probability spaces as an onto homomorphism. Isomorphisms do preserve orthoclosed subspaces. A *symmetry* of a transition probability space is an isomorphism of the space with itself.

There can be only one state in a one-dimensional transition probability space, and since $p(x,x) = 1$, all one-dimensional transition probability spaces are isomorphic.

Two-dimensional transition probability spaces were studied by Mielnik.[1] Each state x in a two-dimensional transition probability space is orthogonal to exactly one other state x' because $\{x\}^{\perp}$ is one-dimensional. Thus, each state belongs to exactly one basis, and distinct bases are disjoint. It follows that a two-dimensional transition probability space must have an even number of states. If there are only two states, the space is pairwise orthogonal. If there are four or more states, then the space is irreducible. The mapping which takes x into x' is a symmetry, and $x'' = x$.

Transition probability spaces (S_1, p_1) and (S_2, p_2) are *similar* if there is an invertible mapping $f: S_1 \to S_2$ such that states x and y in S_1 are orthogonal if and only if their images $f(x)$ and $f(y)$ are orthogonal states in S_2. Two-dimensional transition probability spaces with the same number of states need not be isomorphic, but they are similar.

Since the number of states in an irreducible two-dimensional transition probability space is an even number greater than or equal to four, and since an irreducible three-dimensional transition probability space must have at least six states, it follows that there are no irreducible transition probability spaces with 2, 3, or 5 states.

4. THREE-DIMENSIONAL SPACES

Two states in a transition probability space are *adjacent* if they are neither equal nor orthogonal. The *valence* of a state x is the number of states adjacent to x.

In an irreducible transition probability space, no state can have valence zero, unless it is the only state in the space. On account of the separation axiom, no state in any transition probability space can have valence one.

Thus, in an irreducible transition probability of dimension $d \geq 2$, all states have valence two or more.

In a transition probability space holding a total of N states, there are exactly $N-v-1$ states orthogonal to a state x of valence v. Since each state belongs to at least one basis, no state can have valence exceeding $N-d$. In a three-dimensional transition probability space, $\{x\}^{\perp}$ is two-dimensional, so $N-v-1$ must be even. That is, in a three-dimensional space with an even number of states, each state has odd valence, and vice versa. Moreover, there is a one-to-one correspondence between bases B of the whole space holding x and bases for $\{x\}^{\perp}$, because the other two states in B form one of the $\frac{1}{2}(N-v-1)$ bases for $\{x\}^{\perp}$. Therefore, a state of valence v in a three-dimensional space belongs to exactly $\frac{1}{2}(N-v-1)$ bases.

The minimum number of states in an irreducible three-dimensional transition probability space is six. In such a space with six states, each state has valence three, and thus belongs to a single basis. The whole space is therefore the union of two disjoint bases. All such spaces are similar. The transition probabilities can be represented by a 6×6 matrix,

$$\begin{bmatrix} 1 & 0 & 0 & a-b & 1-c & b+c-a \\ 0 & 1 & 0 & 1-a & c-d & a+d-c \\ 0 & 0 & 1 & b & d & 1-b-d \\ a-b & 1-a & b & 1 & 0 & 0 \\ 1-c & c-d & d & 0 & 1 & 0 \\ b+c-a & a+d-c & 1-b-d & 0 & 0 & 1 \end{bmatrix}$$

where a, b, c, d must satisfy $0<b<a<1$, $0<d<c<1$, $b+d<1$, $a<b+c$, and $c<a+d$. These inequalities can be satisfied, for example, by setting $b=d=\frac{1}{3}$ and $a=c=\frac{2}{3}$.

If x is a state in a transition probability space, and if B is a basis not holding x, then x is adjacent to at least two distinct states in B. Indeed, if only one state y in B were adjacent to x, we would have $1=p(x,B)=p(x,y)$, so that $x=y$, which is absurd.

Given two distinct bases B_1 and B_2, there is at least one state in B_1 not in B_2, and this state must be adjacent to at least two states in B_2. So there are at least two states in B_2 not in B_1. For three-dimensional transition probability spaces, this implies that two distinct bases can have at most one state in common.

The union of three distinct bases B_1, B_2 and B_3 in a transition probability space of dimension d holds at least $d+4$ states. Indeed, if all states common to B_2 and B_3 belong to B_1, then since B_2 and B_3 each hold two states not in B_1, we are done in this case. On the other hand, if there is a state x common to B_2 and B_3 not in B_1, then B_1 holds at least two states adjacent to x, and these states are not in the union of B_2 and B_3. Since the union of B_2 and B_3 itself holds at least $d+2$ states, we are done again.

Three bases form a *triangle* if no state is common to all three, yet each pair of them holds a state in common. No triangles can exist in a three-dimensional transition probability space because the union of three bases in such a space must hold at least seven states.

Three bases are *linked* if two are disjoint, and the third holds one state in common with each of the first two. In a three-dimensional transition probability space, no fourth basis is contained in the union of three linked bases because it would form a triangle with at least two of the three linked bases.

In a seven-state three-dimensional transition probability space, there must be at least three bases. Their union holds all seven states. If the space is irreducible, no state can be common to all three bases. Hence the three bases are linked, and there can be no fourth basis. It follows that all seven-state three-dimensional irreducible transition probability spaces are similar. We can number the states so that the three bases are $\{1,2,3\}$, $\{3,4,5\}$, and $\{5,6,7\}$.

The transition probabilities for a seven-state three-dimensional irreducible transition probability space can be represented by a 7×7 matrix,

$$\begin{bmatrix} 1 & 0 & 0 & a & 1-a & c-b & a+b-c \\ 0 & 1 & 0 & 1-a & a & 1-c & c-a \\ 0 & 0 & 1 & 0 & 0 & b & 1-b \\ a & 1-a & 0 & 1 & 0 & 1-b & b \\ 1-a & a & 0 & 0 & 1 & 0 & 0 \\ c-b & 1-c & b & 1-b & 0 & 1 & 0 \\ a+b-c & c-a & 1-b & b & 0 & 0 & 1 \end{bmatrix},$$

where a, b, and c are positive numbers less than 1 satisfying $a<c$, $b<c$, and $c<a+b$.

The states of a seven-state three-dimensional irreducible transition probability space do not all have the same valence. Two of them have valence two, and five have valence four.

A *bivalent state* is a state of valence two. A *square subspace* is a two-dimensional subspace holding four states.

A bivalent state x in a transition probability space belongs to exactly one square subspace. Moreover, this subspace is orthoclosed. Indeed, if u is one of the states adjacent to x, and if B is a basis holding u, then x is not in B; so B must also hold the other state v adjacent to x. Thus u and v are orthogonal. Since the state x is orthogonal to all states except u and v, it is orthogonal to $\{u,v\}^{\perp}$, that is, x belongs to the two-dimensional orthoclosed subspace $Q_1=\{u,v\}^{\perp\perp}$. In a two-dimensional space, each state is adjacent to all other states, save one. But x is adjacent only to u and v, so Q_1 is a square. The states u and v belong to any square subspace Q_2 holding x, so $Q_1\subset Q_2^{\perp\perp}$. Since the orthoclosed subspaces Q_1 and $Q_2^{\perp\perp}$ are both two-dimensional, $Q_1=Q_2^{\perp\perp}$ and so $Q_2\subset Q_1$. Finally, since Q_1 and Q_2 both hold four states, $Q_1=Q_2$.

The only irreducible transition probability spaces holding two adjacent bivalent states are squares. In fact,

if x and y are adjacent bivalent states, then both belong to an orthoclosed square subspace Q. If there were a state z not in Q, then z is orthogonal to x, so there is a basis holding both x and z. Since y is not in this basis, it is adjacent to two states in this basis. Since y is bivalent, these two states must be x and the state x' in Q orthogonal to x. Since z and x' both belong to this basis, they are orthogonal. Similarly, z is orthogonal to the state y' in Q orthogonal to y. Hence z is orthogonal to every state in Q. That is, the whole space is the union of Q and $Q^{\perp}$. Since the whole space is irreducible, then $Q^{\perp}=\phi$.

As a corollary, it follows that the set of bivalent states in an irreducible transition probability space of dimension $d \geq 3$ is pairwise orthogonal. So there can be at most d bivalent states in such a space. If there actually are d bivalent states, they form a basis B. Suppose there are N states in all. Each of the $N-d$ states not in B is adjacent to at least two states in B. Hence the sum of the valences of the states in B is at least $2(N-d)$. Since this sum is exactly $2d$, then $N \leq 2d$. But an irreducible space must hold at least $2d$ states, so $N=2d$. For $d=3$, this would imply $N=6$, but for such a space we know all states have valence three. The upshot is that a three-dimensional irreducible transition probability space can hold at most two bivalent states.

The seven-state three-dimensional irreducible spaces described above serve to illustrate these results. Here the two bivalent states 3 and 5 belong to the orthoclosed square subspaces $\{3,4,6,7\}$ and $\{1,2,4,5\}$ respectively. We may picture the whole space as two squares joined at a corner.

There are two dissimilar structures possible for a three-dimensional irreducible transition probability space with eight states. The states in such a space can only have valence three or five, so no state belongs to three distinct bases. There must be at least three bases, and at least two, say B_1 and B_2, must overlap. If the three states not in $B_1 \cup B_2$ lie on a third basis, then there is no fourth basis, because triangles are prohibited. The only other possibility is that a third basis has a state in common with B_1 or B_2. Since triangles are prohibited and no state belongs to three bases, this third basis intersects only one of the first two, say B_2. These three bases account for only seven states. There must be a fourth basis to hold the remaining state. The other two states in the fourth basis must be held in common with the first and third bases, respectively. The ban on triangles rules out any further bases. We are left with four bases, each one intersecting two others, and disjoint from the remaining one.

Both these types of space exist. For the first type, the most general transition probability matrix is

$$\begin{bmatrix}
1 & 0 & 0 & a & 1-a & b-c & 1-d & c+d-b \\
0 & 1 & 0 & 1-a & a & 1-b & d-e & b+e-d \\
0 & 0 & 1 & 0 & 0 & c & e & 1-c-e \\
a & 1-a & 0 & 1 & 0 & f-c & 1-g & c+g-f \\
1-a & a & 0 & 0 & 1 & 1-f & g-e & e+f-g \\
b-c & 1-b & c & f-c & 1-f & 1 & 0 & 0 \\
1-d & d-e & e & 1-g & g-e & 0 & 1 & 0 \\
c+d-b & b+e-d & 1-c-e & c+g-f & e+f-g & 0 & 0 & 1
\end{bmatrix}.$$

For the second type, the most general transition probability matrix is

$$\begin{bmatrix}
1 & 0 & 0 & a & 1-a & a & 0 & 0 \\
0 & 1 & 0 & 1-a & a & 1-a-b & b & 1-b \\
0 & 0 & 1 & 0 & 0 & b & 1-b & b \\
a & 1-a & 0 & 1 & 0 & 1-b & b & 1-a-b \\
1-a & a & 0 & 0 & 1 & b & 0 & a \\
a & 1-a-b & b & 1-b & 0 & 0 & 0 & 1-a \\
0 & b & 1-b & b & 0 & 0 & 1 & 0 \\
0 & 1-b & b & 1-a-b & a & 1-a & 0 & 1
\end{bmatrix}.$$

In the first type of space, one state has valence three, and seven have valence five, while in the second type, four states have valence three and four have valence five.

There are five dissimilar types of three-dimensional irreducible transition probability spaces with nine states. We shall describe them only briefly. The simplest type has exactly three bases, which are mutually

disjoint, say $\{1,2,3\}$, $\{4,5,6\}$, and $\{7,8,9\}$. All nine states in this case have valence six.

In the remaining types, each basis intersects at least one other basis. The states may have valence 2, 4, or 6. At most one bivalent state can occur, and it occurs in only one type of space. This type has four bases, three of which meet in the one bivalent state, and a fourth basis which intersects just one of the first three bases. For example, we may take these bases to be $\{1,2,5\}$, $\{3,4,5\}$, $\{5,6,7\}$, and $\{7,8,9\}$. In addition to the one bivalent state, there is one state with valence four, and seven with valence six.

In analyzing the remaining three types, the concept of linked bases is useful. Given any three linked bases in a nine-state three-dimensional transition probability space, the two states not in their union are orthogonal, as one can easily show by an explicit calculation.

One finds there are two more types of irreducible spaces with four bases. For the one type, we can take these bases to be $\{1,2,3\}$, $\{3,4,5\}$, $\{5,6,7\}$, and $\{7,8,9\}$. In the other type, we may take them as $\{1,2,3\}$, $\{4,5,6\}$, $\{7,8,9\}$, and $\{3,5,7\}$. In each of these types of space, there are three states with valence four, and six with valence six.

Finally, there is one type of space in which all nine states have valence four. Each state therefore belongs to two distinct bases. This type of space has six bases, which can be taken as $\{1,2,3\}$, $\{4,5,6\}$, $\{7,8,9\}$, $\{1,4,7\}$, $\{2,5,8\}$, and $\{3,6,9\}$.

5. ORTHOCLOSED SUBSPACES

The subspaces of a transition probability space are partially ordered by inclusion. In general, neither the union nor the intersection of two subspaces is a subspace. Moreover, even if the subspaces are orthoclosed, their union and intersection need not be subspaces. Since singletons are orthoclosed subspaces, any pair of adjacent states provides an example of this phenomenon in the case of unions. For intersections, the simplest counterexample is provided by the eight-state three-dimensional irreducible space with four bases. If we number the states as in the preceding section, the four bases are $\{1,2,3\}$, $\{3,4,5\}$, $\{5,6,7\}$, and $\{7,8,1\}$. The intersection of the orthoclosed square subspaces $\{1\}^\perp = \{2,3,7,8\}$ and $\{5\}^\perp = \{3,4,6,7\}$, for example, is the set $\{3,7\}$, which is orthoclosed, but not a subspace. As a matter of fact, the partially ordered set of all orthoclosed subspaces in this case is not a lattice. The two square subspaces $\{2,3,7,8\}$ and $\{3,4,6,7\}$, for example, have no greatest lower bound since they both cover the two singletons $\{3\}$ and $\{7\}$.

Nevertheless, if two subspaces T_1 and T_2 in a transition probability space are orthogonal, then their union is a subspace. In fact, if B is a maximally pairwise orthogonal subset of $T_1 \cup T_2$, then $B \cap T_1$ and $B \cap T_2$ are bases for T_1 and T_2, respectively, so that

$$B^\perp = (B \cap T_1)^\perp \cap (B \cap T_2)^\perp = T_1^\perp \cap T_2^\perp = (T_1 \cup T_2)^\perp.$$

Hence $T_1 \cup T_2$ is a subspace.

If a subspace T_1 in a transition probability space is contained in another subspace T_2, then $T_1^\perp \cap T_2$ is a subspace. For, if B is a maximally pairwise orthogonal subset of $T_1^\perp \cap T_2$, and if B_1 is a basis for T_1, then

$$\phi = B^\perp \cap (T_1^\perp \cap T_2) = B^\perp \cap B_1^\perp \cap T_2 = (B \cup B_1)^\perp \cap T_2,$$

and hence $B \cup B_1$ is a basis for T_2. If x is a state in $T_1^\perp \cap T_2$, then $1 = p(x, B \cup B_1) = p(x, B)$, so B is a basis for $T_1^\perp \cap T_2$.

If subspaces T_1 and T_2 of a transition probability space satisfy $T_1 \subset T_2$ and $T_1^\perp \cap T_2 = \phi$, then $T_1^\perp = T_2^\perp$. To see this, we note that if B is a basis for T_1, then $B^\perp = T_1^\perp$, and therefore $B^\perp \cap T_2 = \phi$. Then B is a basis for T_2, and $T_1^\perp = B^\perp = T_2^\perp$.

The above results can be simplified if we consider orthoclosed subspaces. The union of two orthogonal orthoclosed subspaces T_1 and T_2 is a subspace, but need not be orthoclosed. However, $(T_1 \cup T_2)^{\perp\perp}$ is an orthoclosed subspace, and it is the least upper bound for T_1 and T_2 in the poset of orthoclosed subspaces. If $T_1 \subset T_2$ and $T_1^\perp \cap T_2 = \phi$ for orthoclosed subspaces, then $T_1 = T_2$.

An *orthocomplemented poset* is a partially ordered set $(P, \leq)$ with a least element 0 and a greatest element 1, and equipped with an operation taking each element $a \in P$ into another element $a^\perp \in P$, such that 0 is the only lower bound for a and $a^\perp$, every element $a \in P$ satisfies $a^{\perp\perp} = a$, and if $a \leq b$, then $b^\perp \leq a^\perp$.

Elements a and b in an orthocomplemented poset are *orthogonal* if $a \leq b^\perp$. An orthocomplemented poset if *orthomodular* if every pair of orthogonal elements has a least upper bound, and if $a \leq b$ and $a^\perp \wedge b = 0$ imply $a = b$.

We may summarize the above results by saying that the set of all orthoclosed subspaces of a transition probability space form an orthomodular poset. That the transition probability axioms imply orthomodularity was discovered another way by K. Bugajska.[5]

An element b in a poset *covers* an element a if $a \leq b$ and there is no element between a and b. In a poset with a least element, an *atom* is an element covering this least element. A poset is *atomistic* if every element b is the least upper bound of the set of atoms a such that $a \leq b$.

The poset of orthoclosed subspaces of a transition probability space is atomistic since every subspace is the union of all singletons contained in it.

Since orthomodularity is just one of several generalizations of the modular law in lattice theory, it is natural to ask if any other generalization of modularity holds for the orthoclosed subspace poset of a transition probability space. A poset is *upper semimodular* if whenever two distinct elements both cover some common element, then some element covers both of them.[6] Lower semimodularity is defined dually. Neither upper nor lower semimodularity need hold in the poset of orthoclosed subspaces of a transition probability space. The simplest counterexample is provided by the six-state three-dimensional irreducible transition probability space. This space has two disjoint bases. If we pick one state in each basis, then their singletons are two atoms not covered by any common orthoclosed subspace, so upper semimodularity fails. Lower semimodularity also fails

here. Even the MacLane exchange axiom,[7] known to quantum theorists as the Jauch–Piron covering axiom,[8] fails here.

There does exist another covering axiom which holds in any atomistic orthomodular poset. If an element c covers an element b in an atomistic orthomodular poset, then there exists an atom $a \leq b^{\perp}$ such that $c = a \vee b$ and $b = a^{\perp} \wedge c$.

6. SPACES WITH SYMMETRY

The most interesting examples of transition probability spaces are those which possess a high degree of symmetry. Various symmetry requirements could be imposed. Von Neumann proposed three such axioms in his 1937 manuscript on continuous geometries with a transition probability.[9] While von Neumann's system differs from the axioms considered here, it is not hard to translate his requirements into reasonable equivalents for transition probability spaces. Recall that a symmetry of a transition probability space is an isomorphism of the space with itself. Each such symmetry induces an automorphism of the poset of orthoclosed subspaces. Von Neumann's first axiom is that all automorphisms of the orthoclosed subspace poset are induced by symmetries of the transition probability space. His second axiom says that if T_1 and T_2 are orthoclosed subspaces satisfying $\dim T_1 \leq \dim T_2$, then there is a symmetry h such that $h(T_1) \subset T_2$. His final axiom says that in the special case that $\dim T_1 = \dim T_2$, then h can be chosen so that $h(T) = T$ for every orthoclosed subspace T orthogonal to both T_1 and T_2.

A somewhat weaker requirement was proposed by Mielnik[1] under the name "superposition principle." Mielnik's requirement is that all orthoclosed two-dimensional subspaces are isomorphic.

Here we consider an even weaker requirement, namely that all states have the same valence v. In the case of a three-dimensional irreducible transition probability space with N states, this requirement implies that each state belongs to exactly $m = \frac{1}{2}(N - v - 1)$ bases. The total number of bases times the number of states on each basis equals the total number of states times the number of bases to which each state belongs. Hence the total number of bases in such a space is $b = \frac{1}{3}Nm$. Since b is an integer, either N is divisible by 3, or else m is divisible by 3. Each basis holds three states. The $N-3$ states not on a given basis are each adjacent to at least two states in that basis. Hence the sum of the valences of the three states in the basis is at least $2(N-3)$. Since this sum is in fact $3v$, then $2(N-3) \leq 3v$, or $m \leq \frac{1}{6}(N+3)$. In particular, if N is not divisible by three, then $m \geq 3$, and the space must hold at least 16 states.

Thus the simplest examples of irreducible three-dimensional transition probability spaces with states of equal valence are those described by the parameters (N, m) below:

$N=6$, $m=1$,

$N=9$, $m=1$ or 2,

$N=12$, $m=1$ or 2,

$N=15$, $m=1,2$, or 3.

The spaces with $N=6$ and $N=9$ were already described in Sec. 4. The cases with $m=1$ are spaces with $N/3$ disjoint bases. They can be constructed using the disjoint union technique described in Sec. 3. In fact the disjoint union technique applied to spaces of type (N_1, m_1) and (N_2, m_2) yields another space of type (N, m) provided $m_1 = m_2 = m$ and $N = N_1 + N_2$.

It is not difficult to construct an example of an irreducible space with $N=12$ and $m=2$. Its eight bases fall into two groups of four each. Each group of four bases are mutually disjoint, and each basis in the one group intersects all but one of the other group. We can number the states from 1 to 12 so that these bases are $\{1,2,3\}$, $\{4,5,6\}$, $\{7,8,9\}$, and $\{10,11,12\}$ for the one group, and $\{1,4,7\}$, $\{2,5,10\}$, $\{3,8,11\}$, and $\{6,9,12\}$ for the other group. The transition probability matrix for this space contains three arbitrary parameters.

ACKNOWLEDGMENTS

The author thanks Professor Walter Noll for drawing his attention to Mielnik's work. He also thanks his former students, Ray Artz and Richard Pelletier, for interesting discussions about transition probability spaces.

[1] B. Mielnik, Commun. Math. Phys. 9, 55 (1968).
[2] P. T. Landsberg, Philos. Mag. 38, 824 (1947).
[3] S. P. Gudder, Am. Math. Mon. **81**, 29 (1974); Errata: *ibid.* **82**, 251 (1975), and *ibid.* **82**, 818 (1975).
[4] V. Bargmann, Ann. Math. 59 (2), 1 (1954).
[5] K. Bugajska, Int. J. Theoret. Phys. 9, 93 (1974).
[6] S. P. Gudder and L. Haskins, Proc. Am. Math. Soc. 28, 395 (1971).
[7] S. MacLane, Duke Math. J. 4, 455 (1938).
[8] J. M. Jauch and C. Piron, Helv. Phys. Acta 42, 842 (1969).
[9] J. von Neumann, *Collected Works* (Pergamon, England, 1961), Vol. 4, pp. 191–194.

Einstein–Maxwell metrics admitting a dual interpretation

N. Tariq

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

B. O. J. Tupper

Department of Mathematics, University of New Brunswick, Fredericton, N. B., Canada
(Received 12 August 1975)

Conditions are given under which the metric part of a solution of the source-free Einstein–Maxwell equations may be interpreted as the metric part of a solution with sources. Examples are given of space-times which admit this dual interpretation and also of space-times admitting one interpretation only.

1. INTRODUCTION

The field equations of Einstein—Maxwell theory in the presence of sources may, with a suitable choice of units, be written in the form[1]

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -E_{\mu\nu} - \rho_0 u_\mu u_\nu, \tag{1.1}$$

$$E_{\mu\nu} = F_{\mu\alpha}F^{\alpha}{}_{\nu} + \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}, \tag{1.2}$$

$$F_{\mu\nu;\sigma} + F_{\nu\sigma;\mu} + F_{\sigma\mu;\nu} = 0, \tag{1.3}$$

$$F^{\mu\nu}{}_{;\nu} = J^\mu, \tag{1.4}$$

where $E_{\mu\nu}$ is the electromagnetic energy tensor, $F_{\mu\nu}$ the electromagnetic field tensor, J^μ the four-current, ρ_0 the invariant rest-mass density, and u_μ is the four-velocity. The presence of the term $\rho_0 u_\mu u_\nu$ in Eq. (1.1) ensures that the right hand side of the equation has vanishing divergence in agreement with the left hand side of the equation. Few attempts to solve this set of equations have been made and in some investigations[2,3] into the inclusion of a source term in Einstein—Maxwell theory the equations are simplified by the assumption $\rho_0 = 0$, so that Eq. (1.1) takes the form

$$R_{\mu\nu} = -E_{\mu\nu}, \tag{1.5}$$

since $E_{\mu\nu}$ is trace-free. This assumption implies not only that the Ricci scalar is zero, but also that $E_{\mu\nu}$ has vanishing divergence, as in the source-free case. Since $E^{\mu\nu}{}_{;\nu} = F^{\mu}{}_{\nu}J^\nu$, the physical consequence of this assumption is that the Lorentz force vanishes, i.e.,

$$F^{\mu}{}_{\nu}J^\nu = 0. \tag{1.6}$$

In the special theory of relativity, this condition implies that the electric and magnetic fields are mutually perpendicular.

In this investigation, we also adopt as the field equations of Einstein—Maxwell theory in the presence of sources, Eqs. (1.2)—(1.5), which imply condition (1.6). We show that it is possible for some space-times to satisfy the source-free Einstein—Maxwell equations, i.e., Eqs. (1.2)—(1.5) with $J^\mu = 0$, and also satisfy the Einstein—Maxwell equations (1.2)—(1.5) for fields with sources under condition (1.6). In other words, while the Ricci tensor uniquely determines the electromagnetic energy tensor $E_{\mu\nu}$, the latter does not necessarily uniquely determine whether the field tensor $F_{\mu\nu}$ satisfies Eqs. (1.4) with $J^\mu = 0$ or Eqs. (1.4) and (1.6) with $J^\mu \neq 0$. Examples are given of space—times which satisfy both sets of equations and also of space—times which can satisfy one set of equations but not the other.

2. NOTATION AND EQUATIONS

Our approach is via the spin-coefficient formalism of Newman and Penrose.[4] A tetrad system of null vectors, $(l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$, where l^μ, n^μ are real and $m^\mu, \bar{m}^\mu$ are conjugate complex vectors, is defined by the relations

$$l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1,$$

with all other contractions zero. The spin coefficients are defined by

$$\kappa = l^\mu m^\nu l_{\nu;\mu}, \quad \sigma = m^\mu m^\nu l_{\nu;\mu}, \quad \rho = \bar{m}^\mu m^\nu l_{\nu;\mu}, \quad \tau = n^\mu m^\nu l_{\nu;\mu},$$
$$\alpha = \tfrac{1}{2}(\bar{m}^\mu n^\nu l_{\nu;\mu} - \bar{m}^\mu \bar{m}^\nu m_{\nu;\mu}), \quad \beta = \tfrac{1}{2}(m^\mu n^\nu l_{\nu;\mu} - m^\mu \bar{m}^\nu m_{\nu;\mu}),$$
$$\gamma = \tfrac{1}{2}(n^\mu n^\nu l_{\nu;\mu} - n^\mu \bar{m}^\nu m_{\nu;\mu}), \quad \epsilon = \tfrac{1}{2}(l^\mu n^\nu l_{\nu;\mu} - l^\mu \bar{m}^\nu m_{\nu;\mu}),$$
$$\nu = n^\mu n^\nu \bar{m}_{\nu;\mu}, \quad \lambda = \bar{m}^\mu n^\nu \bar{m}_{\nu;\mu}, \quad \mu = m^\mu n^\nu \bar{m}_{\nu;\mu}, \quad \pi = l^\mu n^\nu \bar{m}_{\nu;\mu}.$$

Four intrinsic derivatives are defined by

$$D\phi = \phi_{;\mu}l^\mu, \quad \Delta\phi = \phi_{;\mu}n^\mu, \quad \delta\phi = \phi_{;\mu}m^\mu, \quad \bar{\delta}\phi = \phi_{;\mu}\bar{m}^\mu,$$

and give the following commutation relations (integrability conditions):

$$\begin{aligned}
(\Delta D - D\Delta)\phi &= (\gamma + \bar{\gamma})D\phi + (\epsilon + \bar{\epsilon})\Delta\phi - (\tau + \bar{\pi})\bar{\delta}\phi - (\bar{\tau} + \pi)\delta\phi,\\
(\delta D - D\delta)\phi &= (\bar{\alpha} + \beta - \bar{\pi})D\phi + \kappa\Delta\phi - \sigma\bar{\delta}\phi - (\bar{\rho} + \epsilon - \bar{\epsilon})\delta\phi,\\
(\delta\Delta - \Delta\delta)\phi &= -\bar{\nu}D\phi + (\tau - \bar{\alpha} - \beta)\Delta\phi + \bar{\lambda}\bar{\delta}\phi + (\mu - \gamma + \bar{\gamma})\delta\phi,\\
(\delta\bar{\delta} - \bar{\delta}\delta)\phi &= (\mu - \bar{\mu})D\phi + (\rho - \bar{\rho})\Delta\phi + (\bar{\alpha} - \beta)\bar{\delta}\phi - (\alpha - \bar{\beta})\delta\phi,
\end{aligned} \tag{2.1}$$

The three Maxwell scalars are defined by

$$\phi_0 \equiv F_{\mu\nu}l^\mu m^\nu, \quad \phi_2 \equiv F_{\mu\nu}\bar{m}^\mu n^\nu, \tag{2.2}$$
$$\phi_1 \equiv \tfrac{1}{2}F_{\mu\nu}(l^\mu n^\nu + \bar{m}^\mu m^\nu),$$

and Eqs. (1.2) and (1.5) may be written in the form

$$\Phi_{AB} = \phi_A\bar{\phi}_B, \tag{2.3}$$

where Φ_{AB} are the complex tetrad components of the Ricci tensor and A, B take the values 0,1,2.

The Maxwell Eqs. (1.3) and (1.4) in tetrad form are

$$\begin{aligned}
D\phi_1 - \bar{\delta}\phi_0 &= (\pi - 2\alpha)\phi_0 + 2\rho\phi_1 - \kappa\phi_2 + \tfrac{1}{2}I_0,\\
\delta\phi_2 - \Delta\phi_1 &= -\nu\phi_0 + 2\mu\phi_1 + (\tau - 2\beta)\phi_2 + \tfrac{1}{2}I_2,\\
\delta\phi_1 - \Delta\phi_0 &= (\mu - 2\gamma)\phi_0 + 2\tau\phi_1 - \sigma\phi_2 + \tfrac{1}{2}I_1,\\
D\phi_2 - \bar{\delta}\phi_1 &= -\lambda\phi_0 + 2\pi\phi_1 + (\rho - 2\epsilon)\phi_2 + \tfrac{1}{2}\bar{I}_1,
\end{aligned} \tag{2.4}$$

where the I_A are the current scalars defined by

$$I_0=J^\mu l_\mu,\quad I_1=J^\mu m_\mu,\quad \bar{I}_1=J^\mu \bar{m}_\mu,\quad I_2=J^\mu n_\mu. \tag{2.5}$$

Note that I_0, I_2 are real, whereas I_1 is complex.

Condition (1.6) leads to the following equations,

$$\begin{aligned}&(\phi_1+\bar{\phi}_1)I_2-\phi_2 I_1-\bar{\phi}_2\bar{I}_1=0,\\&(\phi_1+\bar{\phi}_1)I_0-\phi_0\bar{I}_1-\bar{\phi}_0 I_1=0,\\&(\phi_1-\bar{\phi}_1)\bar{I}_1+\phi_2 I_0-\bar{\phi}_0 I_2=0,\end{aligned} \tag{2.6}$$

together with the complex conjugate of the third equation. Equations (2.4) and (2.6) have been given previously by Tariq[5] and Zund.[3]

3. THE NONNULL CASE

For a nonnull electromagnetic field the tetrad can be chosen such that $\phi_0=\phi_2=0$, $\phi_1\equiv\phi\neq 0$. In this case l^μ, n^μ are the principal null vectors of the electromagnetic field. Equations (2.3), (2.4) and (2.6) become

$$\Phi_{11}=\phi\bar{\phi}, \tag{3.1}$$

$$\begin{aligned}&D\phi=2\rho\phi+\tfrac{1}{2}I_0,\quad \Delta\phi=-2\mu\phi-\tfrac{1}{2}I_2,\\&\delta\phi=2\tau\phi+\tfrac{1}{2}I_1,\quad \bar{\delta}\phi=-2\pi\phi-\tfrac{1}{2}\bar{I}_1,\end{aligned} \tag{3.2}$$

and

$$(\phi+\bar{\phi})I_0=0,\quad (\phi+\bar{\phi})I_2=0,\quad (\phi-\bar{\phi})I_1=0. \tag{3.3}$$

The invariant $J_\mu J^\mu$ expressed in tetrad components is

$$J_\mu J^\mu=2(I_0 I_2-I_1\bar{I}_1), \tag{3.4}$$

and Eqs. (3.3) imply that

(i) if ϕ is real then $I_0=I_2=0$, $I_1\neq 0$ so that J^μ is spacelike,

(ii) if ϕ is imaginary then $I_1=0$ and at least one of I_0, I_2 is nonzero. In this case J^μ can be timelike ($I_0 I_2>0$), spacelike ($I_0 I_2<0$), or null.

(iii) if ϕ is a complex then no current exists.

When ϕ is real, Eqs. (3.2) take the form

$$\begin{aligned}&D\phi=2\rho\phi,\quad \Delta\phi=-2\mu\phi,\\&\delta\phi=(\tau-\bar{\pi})\phi,\quad \bar{\delta}\phi=(\bar{\tau}-\pi)\phi,\end{aligned} \tag{3.5}$$

with

$$\rho=\bar{\rho},\quad \mu=\bar{\mu}, \tag{3.6}$$

and the tetrad components I_A are given by

$$I_0=I_2=0,\quad I_1=-2(\tau+\bar{\pi})\phi, \tag{3.7}$$

so that $\tau+\bar{\pi}\neq 0$ for nonzero current.

When ϕ is imaginary, Eqs. (3.2) take the form

$$\begin{aligned}&D\phi=(\rho+\bar{\rho})\phi,\quad \Delta\phi=-(\mu+\bar{\mu})\phi,\\&\delta\phi=2\tau\phi,\quad \bar{\delta}\phi=2\bar{\tau}\phi,\end{aligned} \tag{3.8}$$

with

$$\tau+\bar{\pi}=0, \tag{3.9}$$

and the tetrad components I_A are given by

$$I_0=-2(\rho-\bar{\rho})\phi,\quad I_2=-2(\mu-\bar{\mu})\phi,\quad I_1=0. \tag{3.10}$$

In this case, for a nonnull current, $\rho\neq\bar{\rho}$ and $\mu\neq\bar{\mu}$, whereas, for a null current, only one of $\rho\neq\bar{\rho}$ and $\mu\neq\bar{\mu}$ must hold.

Now suppose we have a solution of the Einstein–Maxwell equations for a source-free electromagnetic field, i.e., Eqs. (3.1) and (3.2) with $I_A=0$ in the latter equation. Is it possible that the same space–time is a solution of Eqs. (3.1) and (3.2) with $I_A\neq 0$? The value of $\phi\bar{\phi}$ and the values of the spin coefficients will be the same in each case, so the question becomes: Is it possible to find two different functions ϕ, each with the same modulus, one satisfying (3.2) with $I_A=0$ and the other satisfying (3.2) with $I_A\neq 0$? The Eqs. (3.5)–(3.10) give a prescription for finding such solutions. If a solution of the source-free nonnull Einstein-Maxwell equations is such that both the principal null congruences of the electromagnetic field have zero twist, i.e., Eq. (3.6) holds, and also $\tau+\bar{\pi}\neq 0$, then the space-time is also a solution of the Eqs. (1.2)–(1.6) provided that Eqs. (3.5) satisfy the integrability conditions (2.1). In this case the current J^μ is spacelike, its tetrad components being given by Eq. (3.7). Similarly, if $\tau+\bar{\pi}=0$ and at least one of ρ, μ are not real, then Eqs. (1.2)–(1.6) are satisfied by the space–time, provided that Eq. (3.8) satisfy the integrability conditions. In this case the current can be timelike spacelike or null.

Note that if a solution of the Einstein–Maxwell equations for nonnull electromagnetic fields admits this dual interpretation, then, in its source-free interpretation, the Maxwell scalar ϕ must be strictly complex, i.e., both its real and imaginary parts are nonzero. If this were not so, Eq. (3.2) with $I_A=0$ would impose upon the spin coefficients precisely those conditions which would render $I_A=0$ in the dual interpretation. Hence we have the following theorem.

Theorem 1: A necessary condition for a solution of the source-free Einstein–Maxwell equations for nonnull electromagnetic fields to admit interpretation as a solution of Eqs. (1.2)–(1.6) is that, in the source-free interpretation, the Maxwell scalar ϕ be strictly complex whereas, in the source-present interpretation, the corresponding scalar ϕ be either real or pure imaginary.

In applying the prescription to known solutions of the source-free nonnull Einstein–Maxwell equations we find that many cannot admit the dual interpretation since they have zero twist, i.e., Eq. (3.6) holds, but they also have $\tau=\bar{\pi}=0$, so that Eq. (3.7) gives zero current. Among the solutions in this category are

(i) the Reissner-Nordstrom solution,

(ii) the twist-free Petrov type I solution found by Tariq and Tupper,[6,7] namely,

$$ds^2=2\,du\,dr-u^{-2n}r^{-2m}\,dy^2-u^{-2m}r^{-2n}\,dz^2,$$

where $m=\frac{1}{4}(\sqrt{3}-1)$, $n=-\frac{1}{4}(\sqrt{3}+1)$,

(iii) the conformally flat Bertrotti–Robinson[8] solution,

(iv) the nonnull Petrov type N solution found by McLenaghan and Leroy,[9]

(v) the Petrov type D solution with metric

$$ds^2=r^{-2}\,du^2+2\,du\,dr-r^2(dy^2+dz^2).$$

In fact, we have been unable to find any solution satisfying Eqs. (3.5)–(3.7), i.e., one which gives rise to a spacelike current in its dual interpretation. However,

we now give an example of a source-free solution which can be also interpreted as a solution with timelike current. This solution, which is of Petrov type I, was found by Tariq and Tupper[7] in a special form and later generalized[10,11]; the metric is

$$ds^2 = (dt - 2z\,d\phi)^2 - r^2\,d\phi^2 - \tfrac{1}{2}a^{-2}r^{-2}(dr^2 + dz^2), \tag{3.11}$$

where a is a constant parameter. The principal null congruences l^μ, n^μ of the electromagnetic field are not aligned with those of the gravitational field. The nonzero spin coefficients are

$$\rho = a^2 i, \quad \mu = i, \quad \sigma = a^2 \exp(-2\sqrt{2}iat),$$
$$\lambda = \exp(2\sqrt{2}iat), \quad \alpha = -\beta = az r^{-1} \exp(\sqrt{2}iat). \tag{3.12}$$

The squared modulus of the Maxwell scalar is

$$\phi\bar{\phi} = 2a^2. \tag{3.13}$$

Considered as a source-free solution, we find that

$$\phi = \sqrt{2}\,a \exp[2i(\ln 2r - b)], \tag{3.14}$$

where b is an arbitrary constant, and the nonzero components of the Maxwell tensor are

$$F_{12} = 2r^{-1}\cos 2(\ln 2r - b),$$
$$F_{24} = 4ayr^{-1}\cos 2(\ln 2r - b), \tag{3.15}$$
$$F_{34} = 2\sin 2(\ln 2r - b).$$

From the expressions (3.12) we see that Eq. (3.9) is satisfied and Eq. (3.10) gives a nonzero timelike current. Equations (3.8) become

$$D\phi = \Delta\phi = \delta\phi = \bar{\delta}\phi = 0, \tag{3.16}$$

and so the integrability conditions are satisfied trivially. For the source-present solution, Eqs. (3.13) and (3.16) give

$$\phi = \pm\sqrt{2}ai, \tag{3.17}$$

since ϕ is imaginary. It follows that the only nonzero component of $F_{\mu\nu}$ is

$$F_{34} = \mp 4, \tag{3.18}$$

i.e., a constant magnetic field in the r direction. The tetrad components I_A are

$$I_0 = \pm 4\sqrt{2}a^3, \quad I_1 = 0, \quad I_2 = \pm 4\sqrt{2}a,$$

which leads to

$$J^\mu = \pm 8a^2\delta_1{}^\mu. \tag{3.19}$$

Thus the space–time with metric (3.11), which satisfies the source-free Einstein–Maxwell equations with electromagnetic field given (3.15), may also be interpreted as a solution of the Einstein–Maxwell equations with sources with electromagnetic field and four-current given by Eqs. (3.18) and (3.19), i.e., the expressions (3.15) and (3.18) each give rise to the same electromagnetic energy tensor.

The Kerr–Newman solution[12] appears to satisfy the necessary conditions for the dual interpretation as a solution with time-like current, since, with respect to the tetrad usually employed,[12,13] the spin coefficients satisfy $\bar{\tau} + \pi = 0$, $\rho \neq \bar{\rho}$, $\mu \neq \bar{\mu}$. However, in this case, only l^μ is a principal null vector of the electromagnetic field; the Maxwell scalars ϕ_1, ϕ_2 are both nonzero. We can set ϕ_2 to zero by a tetrad rotation of the form

$$l^\mu \to l^\mu, \quad m^\mu \to m^\mu + Tl^\mu,$$
$$n^\mu \to n^\mu + \bar{T}m^\mu + T\bar{m}^\mu + T\bar{T}l^\mu,$$

and, as a result, the new spin coefficients τ and π no longer satisfy the necessary condition $\bar{\tau} + \pi = 0$. Hence, the Kerr–Newman solution does not admit the dual interpretation.

4. THE NULL CASE

For a null electromagnetic field we may choose l^μ as the repeated principal null direction of the electromagnetic field. In this case, we have $\phi_0 = \phi_1 = 0$, $\phi_2 \equiv \phi \neq 0$ so that Eq. (2.3) becomes

$$\Phi_{22} = \phi\bar{\phi}, \tag{4.1}$$

and Eqs. (2.4) and (2.6) lead to

$$\kappa = I_0 = 0, \quad I_1 = 2\sigma\phi, \quad \sigma\phi^2 + \bar{\sigma}\bar{\phi}^2 = 0,$$
$$\delta\phi = (\tau - 2\beta)\phi + \tfrac{1}{2}I_2, \quad D\phi = (\rho - 2\epsilon)\phi + \tfrac{1}{2}\bar{I}_1. \tag{4.2}$$

From these equations we see that the four-current can be null or spacelike only. Note that the current is null, i.e., $I_1 = 0$, if and only if $\sigma = 0$, which is the generalized Mariot–Robinson theorem due to Zund.[3] As a consequence of this theorem we have the following theorem.

Theorem 2: If any solution of the source-free Einstein–Maxwell equations for null electromagnetic fields can also be interpreted as a solution of Eqs. (1.2)–(1.6), then the four-current arising in the second interpretation is necessarily null.

We now have the following prescription for testing whether or not a solution of the source-free Einstein–Maxwell equations for null electromagnetic fields can also be interpreted as a solution of Eqs. (1.2)–(1.6) with a null four-current. The quantity $\phi\bar{\phi}$ and the spin coefficients are known. Look for a new function ϕ, with the same modulus as the original ϕ, which again satisfies the equation $D\phi = (\rho - 2\epsilon)\phi$, but for which the expression $\delta\phi - (\tau - 2\beta)\phi$ is a nonzero real function. If such a ϕ can be found, then the solution admits the dual interpretation.

Some specific examples of space-times admitting this dual interpretation will now be given. First, we consider the Petrov type N plane-wave solutions of the Einstein–Maxwell equations for null electromagnetic fields. The general metric of these solutions can be written in the form[14]

$$ds^2 = 2(V\bar{V} + W + \bar{W})\,du^2 + 2\,du\,dr - 2\,dz\,d\bar{z}, \tag{4.3}$$

where $V = V(u, z)$, $W = W(u, z)$ are arbitrary functions. With respect to the null tetrad,

$$l^\mu = (0, 1, 0, 0), \quad n^\mu = (1, -V\bar{V} - W - \bar{W}, 0, 0),$$
$$m^\mu = (0, 0, 0, -1), \quad \bar{m}^\mu = (0, 0, -1, 0),$$

the only spin coefficient is $\nu = -V_{,z}\bar{V} - W_{,z}$, and we find

$$\phi\bar{\phi} = V_{,z}\bar{V}_{,\bar{z}}. \tag{4.4}$$

Since $\delta\phi = 0$ we can take $\phi = V_{,z}\exp[if(u)]$, where $f(u)$ is

an arbitrary real function of u. We require a new ϕ, satisfying (4.4) and $D\phi=0$, such that $\delta\phi=-\phi_{,\bar{z}}=\frac{1}{2}I_2$, a real function. We have been unable to prove or disprove the existence of a suitable ϕ for the dual interpretation in the case of arbitrary $V(u,z)$, but we have found a ϕ with the necessary properties in some special cases. In particular, when

$$V(u,z)=q(u)z^{2n}, \tag{4.5}$$

where $q(u)$ is an arbitrary function and n an arbitrary real number, a suitable ϕ does exist. In this case,

$$\phi\bar{\phi}=4n^2q^2(u)z^{2n-1}\bar{z}^{2n-1},$$

and the required new expression for ϕ is

$$\phi=2n\bar{z}^n z^{-n}V_{,z},$$

i.e.,

$$\phi=4n^2q(u)\bar{z}^n z^{n-1}. \tag{4.6}$$

This gives the following expression for I_2,

$$I_2=-2\phi_{,\bar{z}}=-8n^3q(u)\,|z|^{2(n-1)} \tag{4.7}$$

which is real, as required. The four-current has the form

$$J^\mu=-8n^3q(u)\,|z|^{2(n-1)}\delta_2{}^\mu.$$

In both interpretations the nonzero components of the electromagnetic field are given by

$$F_{13}=\phi,\quad F_{14}=\bar{\phi}.$$

Note that the conformally flat solution of the Einstein–Maxwell equations for null electromagnetic fields is a special case of the type N plane-wave solutions discussed above. The metric has the form[14,15]

$$ds^2=2q^2(u)z\bar{z}\,du^2+2\,du\,dr-2\,dz\,d\bar{z},$$

which corresponds to the metric (4.3) with $V=q(u)z$ and $W=0$. Thus, the conformally flat solution is a member of the class of solutions characterized by (4.5) with $n=\frac{1}{2}$ and so admits the dual interpretation. In this case, the original Maxwell scalar, which is a function of u only, is replaced by

$$\phi=q(u)\bar{z}^{1/2}z^{-1/2},$$

and the resulting four-current is

$$J^\mu=-q(u)\,|z|^{-1}\delta_2^\mu.$$

As a second example, we consider the Petrov type D solution with metric

$$ds^2=[2m(u)/r]\,du^2-2\,du\,dr-2r^2\,dz\,d\bar{z}. \tag{4.8}$$

With respect to the null tetrad,

$$l^\mu=(0,1,0,0),\qquad n^\mu=(1,-m/r,0,0),$$
$$m^\mu=(0,0,0,-1/r),\quad m^\mu=(0,0,-1/r,0),$$

the nonzero spin coefficients are

$$\rho=-1/r,\quad \mu=-m/r^2,\quad \gamma=-m/2r^2.$$

The nonzero Weyl scalar is $\psi_2=m/r^3$ and

$$\phi\bar{\phi}=\dot{m}/r^2, \tag{4.9}$$

which shows that $\dot{m}>0$; we put $\dot{m}=k^2(u)$. It follows from the Bianchi identities that, for the source-free solution, ϕ must be of the form

$$\phi=k(u)r^{-1}\exp[if(u)],$$

where $f(u)$ is an arbitrary function of u. We require a new ϕ satisfying (4.9) and $D\phi=-(1/r)\phi$ such that $\delta\phi=-(1/r)\phi_{,\bar{z}}=\frac{1}{2}I_2$, real. These conditions are satisfied by

$$\phi=k(u)r^{-1}\bar{z}^{1/2}z^{-1/2},$$

which gives $I_2=-\frac{1}{2}kr^{-2}\,|z|^{-1}$, so that the four-current is

$$J^\mu=-\tfrac{1}{2}kr^{-2}\,|z|^{-1}\delta_2{}^\mu.$$

Finally, we turn our attention to the inverse problem to that considered so far, namely, given a solution of Eqs. (1.2)–(1.6) with $J^\mu\neq0$, can the solution also be interpreted as a source-free solution. This amounts to finding a function ϕ satisfying the equation $\delta\phi-(\tau-2\beta)\phi=0$ and having the same modulus as the original ϕ which satisfied the equation $\delta\phi-(\tau-2\beta)\phi=\frac{1}{2}I_2$. We now give an example of a solution with sources which cannot have this dual interpretation.

Consider the solution of the Eqs. (1.2)–(1.6) for null electromagnetic field with null current found by Vaidya.[16] This has metric

$$ds^2=[1-2m(u)/r]\,du^2+2\,du\,dr-r^2(d\theta^2+\sin^2\theta\,d\psi^2). \tag{4.10}$$

With respect to the null tetrad

$$l^\mu=(0,\,1,\,0,\,0),\quad n^\mu=[1,\,-\tfrac{1}{2}(1-2m/r),\,0,\,0],$$
$$m^\mu=(0,\,0,\,1/\sqrt{2}r,\,i/\sqrt{2}r\sin\theta),$$
$$\bar{m}^\mu=(0,\,0,\,1/\sqrt{2}r,\,-i/\sqrt{2}r\sin\theta,$$

the nonzero spin coefficients are

$$\rho=-1/r,\quad \mu=-(1/2r)(1-2m/r),$$
$$\gamma=m/2r^2,\quad \alpha=-\beta=-(1/2\sqrt{2}r)\cot\theta,$$

and the nonzero Weyl scalar is $\psi_2=-m/r^3$, so that the solution is of Petrov type D. We find that

$$\phi\bar{\phi}=-\dot{m}/r^2, \tag{4.11}$$

so that $\dot{m}<0$; we write $\dot{m}=-k^2(u)$. The Maxwell scalar ϕ must satisfy (4.11) and also

$$D\phi=-(1/r)\phi,$$
$$\delta\phi+(1/\sqrt{r})\cot\theta\phi=\tfrac{1}{2}I_2. \tag{4.12}$$

This is satisfied by

$$\phi=k(u)r^{-1},$$

which gives

$$I_2=\sqrt{2}k(u)r^{-1}\cot\theta,$$

i.e.,

$$J^\mu=\sqrt{2}k(u)r^{-1}\cot\theta\delta_2{}^\mu.$$

Now we look for a new ϕ satisfying Eqs. (4.11), (4.12) and

$$\delta\phi+(1/\sqrt{2}r)\cot\theta\,\phi=0. \tag{4.13}$$

From Eqs. (4.11) and (4.12) we have

$$\phi = k(u) r^{-1} \exp[if(u, \theta, \psi)].$$

Substituting this into Eqs. (4.13) and equating real and imaginary parts we find

$$\frac{\partial f}{\partial \theta} = 0, \quad \frac{\partial f}{\partial \psi} = \cot\theta,$$

which are incompatible. Hence, no function ϕ satisfying Eqs. (4.11), (4.12) and (4.13) can be found, so that the metric (4.10) cannot be interpreted as a solution of the source-free Einstein–Maxwell equations.

5. KILLING VECTORS

We now turn to the question of whether or not the symmetries of the metric field of a dual-interpretation solution are inherited by the electromagnetic field, i.e., does

$$\underset{\xi_i}{\mathcal{L}} g_{\mu\nu} = 0 \Rightarrow \underset{\xi_i}{\mathcal{L}} F_{\mu\nu} = 0$$

in each of the two interpretations? The problem of the inheritance of symmetries for the case of nonnull electromagnetic fields has been discussed at length by Michalski and Wainwright.[17] Here we give only the results for the dual interpretation solutions found in Secs. 3 and 4.

The space-time (3.11) admits a four-parameter group of motions of Bianchi type I. The four Killing vectors are

$$\xi_1^\mu = (1,0,0,0), \quad \xi_2^\mu = (0,0,0,1),$$

$$\xi_3^\mu = (2\phi, 0, 1, 0), \quad \xi_4^\mu = (0, r, z, -\phi).$$

In the source-free interpretation, the first three Killing vectors, ξ_i^μ $(i = 1, 2, 3)$, but not ξ_4^μ, are symmetries of the electromagnetic field, i.e., for $F_{\mu\nu}$ given by Eq. (3.15) we have[10,17]

$$\underset{\xi_i}{\mathcal{L}} F_{\mu\nu} = 0, \quad \underset{\xi_4}{\mathcal{L}} F_{\mu\nu} \neq 0.$$

However in the source-present interpretation, with $F_{\mu\nu}$ given by Eq. (3.18), all four Killing vectors are symmetries of $F_{\mu\nu}$. In this case, the four-current J^μ, given by Eq. (3.19), is parallel to the time-like Killing vector ξ_1^μ.

In the case of the type N plane-wave solutions with metric (4.3), there are always at least two Killing vectors, namely

$$\xi_1^\mu = (0, 1, 0, 0), \quad \xi_2^\mu = (0, 0, z, -\bar{z}).$$

In the source-free case only the first of these is a symmetry of the electromagnetic field, whereas in the source-present case both ξ_1^μ and ξ_2^μ satisfy

$$\underset{\xi}{\mathcal{L}} F_{\mu\nu} = 0.$$

The four-current, which is null, is again parallel to one of the Killing vectors, namely ξ_1^μ.

The similarity in the behavior of the electromagnetic fields with respect to the Killing vectors for the two solutions given above does not occur in the case of solution(4.8). This solution admits three Killing vectors, namely,

$$\xi_1^\mu = (0, 0, 0, 1),$$

$$\xi_2^\mu = (0, 0, 1, 0),$$

$$\xi_3^\mu = (0, 0, z, -\bar{z}).$$

In the source-free interpretation $\xi_1{}^\mu$ and $\xi_2{}^\mu$ are symmetries of the electromagnetic field whereas $\xi_3{}^\mu$ is not. In the source-present interpretation, these properties are reversed, i.e., ξ_3^μ is a symmetry of $F_{\mu\nu}$ whereas $\xi_1{}^\mu$, $\xi_2{}^\mu$ are not. Furthermore, the four-current in this interpretation is not parallel to any of the Killing vectors.

ACKNOWLEDGMENTS

We wish to express our thanks to E.C. Ihrig and J. Wainwright for helpful discussions. This research was supported in part by the National Research Council of Canada through operating Grant A7589.

[1]V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon, London, 1964).
[2]P.A. Goodinson and R.A. Newing, J. Inst. Math. Applics. 6, 212 (1970).
[3]J.D. Zund, Tensor N.S. 28, 283 (1974).
[4]E. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[5]N. Tariq, Ph.D. Thesis, Univ. New Brunswick (1974).
[6]N. Tariq and B.O.J. Tupper, Tensor N.S. 28, 83 (1974).
[7]N. Tariq and B.O.J. Tupper, Gen. Rel. Grav. 6, 345 (1975).
[8]N. Tariq and B.O.J. Tupper, J. Math. Phys. 15, 2232 (1974).
[9]R.G. McLenaghan and J. Leroy, Proc. R. Soc. Lond. A 327, 229 (1972).
[10]R.G. McLenaghan and N. Tariq, J. Math. Phys. 16, 2306 (1975).
[11]B.O.J. Tupper, Gen. Rel. Grav. 7 (1976), to appear.
[12]E.T. Newman, E. Couch, K. Chinnapered, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. 6, 918 (1965).
[13]R.W. Lind, J. Math. Phys. 16, 39 (1975).
[14]M. Cahen and J. Leroy, Bull. Acad. R. Belg. 51, 1179 (1965).
[15]R.G. McLenaghan, N. Tariq, and B.O.J. Tupper, J. Math. Phys. 16, 829 (1975).
[16]P.C. Vaidya, Proc. Indian Acad. Sci. Sec. A 33, 264 (1951).
[17]H. Michalski and J. Wainwright, Gen. Rel. Grav. 6, 289 (1975).

An effective-potential approach to stationary scattering theory for long-range potentials

D. Masson* and E. Prugovečki*

Department of Mathematics, University of Toronto, Toronto M5S 1A1, Canada
(Received 18 August 1975)

It is shown that previously derived integral equations for two-body scattering with long-range potentials (equations which replace the Lippmann-Schwinger equations) can be reduced to a form which is solvable by iterative methods. The method is applicable to potentials $V(r)$ which behave asymptotically as $r^{-\alpha}$, $1/2 < \alpha \leq 1$, and in particular to Coulomb-like potentials.

1. INTRODUCTION

It has been recently pointed out[1,2] that although the standard approach to stationary nonrelativistic theory based on Lippmann—Schwinger equations breaks down when long-range forces are present, it can be generalized to include the long-range case. The idea behind the generalization relies on the possibility of introducing an operator Z which compensates for the anomalous asymptotic behavior of the interacting wavefunction in relation to that of the free wavefunction (an anomaly that is peculiar to the long-range case).

Since only asymptotic behavior is of importance in the construction of the "asymptotically compensating" operator Z, there is a whole family of operators which can play this role. In the present note we indicate how this nonuniqueness can be exploited to advantage to obtain integral equations for the scattering wavefunctions which can be solved by perturbational methods and yield the on-energy-shell T matrix. In order to simplify the discussion and notation, we limit ourselves to the case of two-body interactions with a spherically symmetric potential (cf. Theorem 1 for an exact statement of technical conditions), $V(r) \sim 0(r^{-\alpha})$, where $1/2 < \alpha \leq 1$. It is hoped, however, that since the underlying general theory[2] stays valid in the multichannel case, the present approach is also of relevance in that context.

In Sec. 4 we show that there are operators Z acting in $L^2(R^3)$ such that (note that we adopt the convention in which Ω_+ corresponds to $t \to +\infty$!),

$$\Omega_{\pm}^{*}\Psi_{\pm} = \operatorname*{s-lim}_{t \to \pm\infty} \exp(iH_0 t) Z \exp(-iHt)\Psi_{\pm}, \tag{1.1}$$

for a dense set of vectors $\Psi_{\pm}$ in the ranges $\mathcal{R}_{\pm}$ of $\Omega_{\pm}$. They act on the incoming and outgoing spherical waves $\Phi_{klm}^{\pm}$ at energy $E = k^2$ (we choose units of reduced mass M so that $2M = 1$) via an asymptotically compensating term $\phi_l(r,k)$,

$$(Z\Phi_{klm}^{\pm})(r;\mathbf{r}/r) = \Phi_{klm}^{\pm}(r + \phi_l(r,k), \mathbf{r}/r) \tag{1.2}$$

which makes the behavior of $Z\Phi_{klm}^{\pm}$ as $r \to \infty$ essentially the same as in the short range case, i.e., that of a free spherical wave

$$\Phi_{klm}(\mathbf{r}) = \sqrt{2/\pi}\, j_l(kr) Y_l^m(\mathbf{r}/r) \tag{1.3}$$

modified by a phase shift $\delta_l(k)$.

From (1.1) we derive[1] that

$$Z\Phi_{klm}^{\pm} = \Phi_{klm} + (H_0 - E \pm i0)^{-1}(H_0 - E) Z\Phi_{klm}^{\pm}, \tag{1.4}$$

where one should note that in contradistinction to previously considered[1-3] asymptotically compensating operators Z which modify the free wave Φ_{klm}, the one in (1.2) modifies the behavior of the full wavefunctions $\Phi_{klm}^{\pm}$. This choice was motivated by the desire to recast (1.4) in the form of a Lippmann—Schwinger equation

$$(Z\Phi_{klm}^{\pm})(\mathbf{r}) = \Phi_{klm}(\mathbf{r}) + \int_{\mathbb{R}^3} G_{0l}^{\pm}(\mathbf{r},\mathbf{r}',E) V_{l,\mathrm{eff}}(\mathbf{r}',k)(Z\Phi_{klm}^{\pm})(\mathbf{r}')\, d\mathbf{r}' \tag{1.5}$$

in terms of the advanced and retarded free Green functions $G_{0l}^{\pm}(\mathbf{r},\mathbf{r}',E)$ and an effective potential $V_{l,\mathrm{eff}}$ [cf. (3.8)], which would give rise to a convergent perturbation series. This goal is achieved in Sec. 3 by a judicious choice of $\phi_l(r,k)$ in relation to the long-range part of $V(r)$. (Our particular choice of $\phi_l(r,k)$ turns out to be independent of l.) This result is similar in nature to that recently obtained by different methods for the "solution-type" Lippmann—Schwinger equations in long-range scattering by Zorbas.[4] However, the Zorbas equations are impracticable since they are not soluable by iterative methods.

2. INTEGRAL EQUATIONS FOR THE DISTORTED WAVEFUNCTIONS

The free spherical waves can be written in the form

$$\Phi_{klm}(\mathbf{r}) = \frac{u_l(r,k)}{kr} Y_l^m(\mathbf{r}/r) \tag{2.1}$$

where $u_l(r,k)$ are solutions of the equation

$$\left(-\frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2}\right) u_l(r,k) = k^2 u_l(r,k), \tag{2.2}$$

which for $rk \to \infty$ behave asymptotically in the following manner:

$$u_l(r,k) = \sqrt{2/\pi}\sin(kr - l\pi/2) + O(1/kr). \tag{2.3}$$

Any wavepacket $\Psi(\mathbf{r}) \in L^2(\mathbb{R}^3)$ can be unambiguously represented by an element of $L^2([0,\infty))$

$$\tilde{f}_{lm}(k) = \text{l.i.m.} \int_0^{\infty} r u_l(r,k)\, dr \int_{r=1} Y_l^m(\omega)\Psi(\mathbf{r})\, d\omega, \tag{2.4}$$

where the ω-integration is carried out over the unit sphere $\{\mathbf{r} \mid r = 1\}$. In fact, from the element of $L^2([0,\infty))$ given by

$$f_{lm}(r) = \text{l.i.m.} \int_0^{\infty} u_l(r,k)\tilde{f}_{lm}(k)\, dk \tag{2.5}$$

one can recover the wavepacket,

$$\Psi(\mathbf{r}) = \text{l.i.m.} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{f_{lm}(r)}{r} Y_l^m(\mathbf{r}/r). \tag{2.6}$$

If we assume the interaction to be given by a piecewise continuous, spherically-symmetric potential

$$\begin{aligned} &V(r) = V_S(r) + V_L(r), \\ &V_S(r) = O(r^{-2+\epsilon}) \text{ as } r \to 0+, \\ &V_L(r) = O(r^{-\alpha}) \text{ as } r \to \infty, \end{aligned} \tag{2.7}$$

with $\epsilon > 0$ and $\frac{1}{2} < \alpha \leq 1$, then $\Psi_\pm(\mathbf{r}) = (\Omega_\pm \Psi)(\mathbf{r})$ can be represented by

$$\Psi_\pm(\mathbf{r}) = \text{l.i.m.} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{f_{lm}^\pm(r)}{r} Y_l^m(\mathbf{r}/r), \tag{2.8}$$

where

$$f_{lm}^\pm(r) = \text{l.i.m.} \int_0^\infty u_l^\pm(r,k) \tilde{f}_{lm}(k)\, dk,$$

$$\Phi_{klm}^\pm(\mathbf{r}) = \frac{u_l^\pm(r,k)}{kr} Y_l^m(\mathbf{r}/r) = (\Omega_\pm \Phi_{klm})(\mathbf{r}). \tag{2.9}$$

The operator Z can be defined in terms of the asymptotically compensating term $\phi_l(r,k)$ by

$$(Z\Psi_\pm)(\mathbf{r}) = \text{l.i.m.} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{(Z_l f_{lm}^\pm)(r)}{r} Y_l^m(\mathbf{r}/r), \tag{2.10}$$

$$(Z_l f_{lm}^\pm)(r) = \text{l.i.m.} \int_0^\infty u_l^\pm(r + \phi_l(r,k), k) \tilde{f}_{lm}(k)\, dk.$$

Naturally, the functions $\phi_l(r,k)$ must be chosen so that (1.1) holds or equivalently that

$$\lim_{t \to \pm\infty} \| (Z\Omega_\pm - I) \exp(-iH_0 t)\Psi \| = 0 \tag{2.11}$$

for a suitably chosen dense set D_0 of functions $\Psi \in L^2(\mathbb{R}^3)$.

By combining the fact that

$$\exp(-iH_0 t) f_{lm}\tilde{}\,(k) = \exp(-ik^2 t) \tilde{f}_{lm}(k) \tag{2.12}$$

with relations (2.6) and (2.9) we see that the existence of such a set D_0 would certainly be established if we show the existence of dense sets $\tilde{D}_{lm} \subset L^2([0,\infty))$ for all $l = 0, 1, \cdots$ and $m = -l, -l+1, \ldots, l$ such that for any $\tilde{f} \in \tilde{D}_{lm}$,

$$\text{s-lim}_{t \to \pm\infty} \int_0^\infty \{(Z_l u_l^\pm)(r,k) - u_l(r,k)\} \exp(-ik^2 t) \tilde{f}(k)\, dk = 0 \tag{2.13}$$

in the $L^2([0,\infty))$ topology. We shall discuss in Sec. 4 the feasibility of choices for the functions $\phi_l(r,k)$ in

$$(Z_l u_l^\pm)(r,k) = u_l^\pm(r + \phi_l(r,k), k), \tag{2.14}$$

for which dense sets $\tilde{D}_{lm}$ of functions $\tilde{f}(k)$ satisfying (2.13) exist. The argument is based on the work of Matveev and Skriganov.[3]

From (1.1) we infer (cf. Appendix) that

$$\Omega_\pm^* \Psi_\pm = \text{s-lim}_{\epsilon \to 0+} \int_{-\infty}^{\infty} \frac{\pm i\epsilon}{H_0 - \lambda \pm i\epsilon} Z\, d_\lambda E_\lambda^H \Psi_\pm. \tag{2.15}$$

Consequently, by noting that

$$\frac{\pm i\epsilon}{H_0 - \lambda \pm i\epsilon} = I - \frac{H_0 - \lambda}{H_0 - \lambda \pm i\epsilon} \tag{2.16}$$

and setting $\Psi = \Omega_\pm^* \Psi_\pm$, we arrive at

$$Z\Psi_\pm = \Psi + \text{s-lim}_{\epsilon \to 0+} \int_{-\infty}^{\infty} \frac{1}{H_0 - \lambda \pm i\epsilon} (H_0 - \lambda) Z\, d_\lambda E_\lambda^H \Psi_\pm. \tag{2.17}$$

The above equation in Hilbert space, in turn, implies (1.4) for the eigenfunction expansions of H_0 and H. In terms of $u_l(r,k)$ and $u_l^\pm(r,k)$ the relation (1.4) is equivalent to

$$\begin{aligned} (Z_l u_l^\pm)(r,k) = u_l(r,k) \\ + \int_0^\infty G_{0l}^\pm(r,r';E)(H_l^{(r')} - E)(Z_l u_l^\pm)(r',k)\, dr', \end{aligned} \tag{2.18}$$

where $H_l^{(r)}$ is the differential operator

$$H_l^{(r)} = \left(-\frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} \right). \tag{2.19}$$

The free Green's function is given by

$$G_{0l}^\pm(r,r';k^2) = u_l(r_<,k)[u_l(r_>,k) \mp i(-1)^l u_{-l-1}(r_>,k)], \tag{2.20}$$

where

$$\begin{aligned} &r_< = \min(r,r'), \quad r_> = \max(r,r'), \\ &u_l(r,k) = \sqrt{kr}\, J_{l+1/2}(kr), \end{aligned} \tag{2.21}$$

and $J_\nu(x)$ is the Bessel function of order ν.

The on-energy-shell T matrix also contains the Z operator.[1,2] In fact, by substituting in the formula

$$T = (1/2\pi i)(\Omega_+^* \Omega_- - I) = (1/2\pi i)(\Omega_+^* - \Omega_-^*)\Omega_- \tag{2.22}$$

for the T operator the expressions (2.15)—(2.16) we get

$$T = \frac{1}{\pi}\, \text{s-lim}_{\epsilon \to 0+} \int_{-\infty}^{\infty} \frac{\epsilon}{(H_0 - \lambda)^2 + \epsilon^2} (H_0 - \lambda) Z\Omega_-\, d_\lambda E_\lambda^H. \tag{2.23}$$

Expressing the integration with respect to the spectral functions E_λ^H of H in terms of the wavefunctions (2.9) and taking the $\epsilon \to 0+$ under the integral sign, one obtains for $\Psi', \Psi \in D_0$

$$\langle \Psi' \mid T\Psi \rangle = \sum_{l,m} \int_0^\infty dk\, \overline{\tilde{f}'_{lm}(k)}\, T(k,l,m) \tilde{f}_{lm}(k), \tag{2.24}$$

where

$$T(k,l,m) = \int_0^\infty \overline{u_l(r,k)}(H_l^{(r)} - E)(Zu_l^-)(r,k)\, dr. \tag{2.25}$$

The relation (2.24) shows that the $T(k,l,m)$ play the role of on-energy-shell T matrix components at energy k^2 and angular momentum l,m. The method presented in the next section provides an iterative procedure for the computation of the term

$$(H_l^{(r)} - E)(Zu_l^\pm)(r) \tag{2.26}$$

appearing in (2.25), and therefore also a method for the computation of (2.25).

3. DERIVATION OF THE EFFECTIVE SHORT RANGE POTENTIAL

The operator Z has been defined by its action on the partial-wave eigenfunctions of the Hamiltonian $H = -\Delta$

$+V(r)$ [where $V(r)=0(r^{-\alpha})$ as $r\to\infty$, $\frac{1}{2}<\alpha\leq 1$] in terms of the asymptotically compensating term $\phi_l(r,k)$. The term $\phi_l(r,k)$ can be chosen to satisfy two important criteria:

(1) The wave operators $\Omega_{\pm}^{*}$ can be obtained as strong limits of $\exp(iH_0t)Z\exp(-iHt)$ [Eq. (1.1)]

(2) The Lippmann–Schwinger equation (2.18) can be solved by iteration since a short range (but energy dependent) effective potential $V_{l,\mathrm{eff}}$ satisfying

$$(H_l^{(r)}-k^2)(Zu_l^{\pm})(r,k)=-V_{l,\mathrm{eff}}(Zu_l^{\pm})(r,k) \tag{3.1}$$

can be defined which yields an effective Hilbert-Schmidt kernel for the solution of (2.18).

In this section we justify the second assertion. The first is dealt with in Sec. 4.

To derive $V_{l,\mathrm{eff}}$, we simply calculate (2.26) and impose an obvious condition on $\phi_l(r,k)$.

Let $g(r)$ be a C^2 function satisfying

$$\begin{aligned} &g(r)=0, \quad r\leq R,\\ &g(r)=1, \quad r\geq 2R. \end{aligned} \tag{3.2}$$

The value of R will be chosen appropriately later. We define V_S and V_L appearing in (2.7) in terms of

$$\begin{aligned} &V_L(r)=g(r)V(r),\\ &V_S(r)=[1-g(r)]V(r), \end{aligned} \tag{3.3}$$

where V_L is long range and goes asymptotically like $\gamma r^{-\alpha}$, $\frac{1}{2}<\alpha\leq 1$; V_S is short range and less singular for $r\to 0+$ than r^{-2}, and $V=V_L+V_s$. One has that

$$\begin{aligned}(H_l^{(r)}-k^2)(Zu_l^{\pm})(r,k)&\\ =\left(-\frac{\partial^2}{\partial r^2}+\frac{l(l+1)}{r^2}-k^2\right)u_l^{\pm}(r+\phi_l(r,k),k).&\end{aligned} \tag{3.4}$$

By using the relation

$$\left(-\frac{\partial^2}{\partial x^2}+\frac{l(l+1)}{x^2}-k^2+V(x)\right)u_l^{\pm}(x,k)=0 \tag{3.5}$$

we can eliminate $(\partial^2u^{\pm}/\partial r^2)(x,k)$, with $x=r+\phi_l(r,k)$, from (3.4) and thus obtain

$$\begin{aligned}(H_l^{(r)}-k^2)(Zu_l^{\pm})(r,k)=&\Bigg[\{2\phi_l'(r,k)+[\phi_l'(r,k)]^2\}\\ &\times\left(k^2-\frac{l(l+1)}{[r+\phi_l(r,k)]^2}-V(r+\phi_l(r,k))\right)+l(l+1)\\ &\times\left(\frac{1}{r^2}-\frac{1}{[r+\phi_l(r,k)]^2}\right)-\phi_l''(r,k)[1+\phi_l'(r,k)]^{-1}\frac{\partial}{\partial r}\\ &-V(r+\phi_l(r,k))\Bigg](Zu_l^{\pm})(r,k),\end{aligned} \tag{3.6}$$

where $\phi_l'=\partial\phi_l/\partial r$ and $\phi_l''=\partial^2\phi_l/\partial r^2$.

If one chooses $\phi_l(r,k)$ to satisfy the first-order differential equation (other choices are possible, but this appears to be most convenient for our purposes)

$$\begin{aligned}&\{2\phi_l'(r,k)+[\phi_l'(r,k)]^2\}[k^2-V_L(r+\phi_l(r,k))]\\ &\quad=V_L(r+\phi_l(r,k)),\end{aligned} \tag{3.7}$$

then (3.6) reduces to (3.1) with

$$V_{l,\mathrm{eff}}=q_{l0}(r,k)+q_1(r,k)\frac{\partial}{\partial r} \tag{3.8}$$

where

$$\begin{aligned}q_{l0}(r,k)=&V_S(r+\phi(r,k))+l(l+1)\\ &\times[(r+\phi(r,k))^{-2}-r^{-2}]-\{2\phi'(r,k)\\ &+[\phi'(r,k)]^2\}\left(\frac{l(l+1)}{[r+\phi(r,k)]^2}+V_S(r+\phi(r,k))\right),\end{aligned} \tag{3.9}$$

and

$$q_1(r,k)=\phi''(r,k)[1+\phi'(r,k)]^{-1}. \tag{3.10}$$

Note that the subscript l has been dropped from ϕ since Eq. (3.7) is independent of l.

The above calculation is justified provided that we can show the existence of a twice continuously differentiable solution to the differential equation (3.7). This follows[6] from the theory of ordinary differential equations for a suitable choice of R.

We recast (3.7) in the form

$$\frac{dx}{dr}=[1+g(x,k)]^{1/2} \tag{3.11}$$

where

$$\begin{aligned}&x=\phi(r,k)+r,\\ &g(x,k)=V_L(x)[k^2-V_L(x)]^{-1}.\end{aligned} \tag{3.12}$$

A unique twice continuously differentiable solution to (3.11) satisfying $x(R)=R$ is given by

$$\int_0^x[k^2-V_L(t)]^{1/2}\,dt=kr \tag{3.13}$$

provided that $V_L(x)$ and $(dV_L/dx)(x)$ are continuous and $k^2-V_L(x)>0$ for $x\in[0,\infty)$. This is certainly true if R is chosen sufficiently large, $V(x)$ is continuously differentiable, and $k^2-V(x)>0$ for $x\in[R,\infty)$. The properties of $\phi(r,k)$ may be summarized as follows.

Theorem 1: Let $V(r)\in C^1([R,\infty))$ for R sufficiently large with $\lim_{r\to\infty}r^{\alpha}V(r)=\gamma\neq 0$ and $V'(r)=O(r^{-\alpha-1})$ for some $\frac{1}{2}<\alpha\leq 1$. Then for $k\geq k_0>0$ there exists a unique solution $\phi(r,k)$ to the differential equation (3.7) such that

(1) $\phi(r,k)$ is C^2 for $r\in[0,\infty)$,

(2) $\phi(r,k)=0$ for $r\leq R$,

(3) $\phi(r,k)-\frac{1}{2k^2}\int_R^r V_L(t)\,dt-K_R(k)$

$=O\left(\frac{1}{k^4}\right)\cdot O[p_{\alpha}(r)r^{-2\alpha+1}]$, where

$$K_r=\int_R^{\infty}\left[1-\frac{1}{2k^2}V_L(t)-\left(1-\frac{V_L(t)}{k^2}\right)^{1/2}\right]dt$$

with $p_{\alpha}(r)=1$ for $\alpha<1$ and $p_{\alpha}(r)=\log r$ for $\alpha=1$,

(4) $\frac{\partial^n \phi}{\partial r^n}(r,k) - \frac{\gamma}{2k^2} r^{-\alpha+1-n}$

$= O\left(\frac{1}{k^4}\right) \cdot O[p_\alpha(r) r^{1-2\alpha-n}], \quad n=1,2.$

The short range character of $V_{l,\text{eff}}$ follows from properties (3) and (4) since for large r and $k \neq 0$,

$$q_{l0} = O(r^{-3} \int_R^r V(r)\, dr),$$
$$q_1 = O(r^{-\alpha-1}). \tag{3.14}$$

Thus under the hypothesis of Theorem 1 the integral equation (2.18) becomes

$$(Zu_l^\pm)(r,k) = u_l(r,k) - \int_0^\infty G_{0l}^\pm(r,r',E) \times \left(q_{l0}(r',k) + q_1(r',k)\frac{\partial}{\partial r'}\right)(Zu_l^\pm)(r',k)\, dr'. \tag{3.15}$$

An integration by parts in the last term {justified if $V(r)$ is $C^2([R,\infty))$} yields

$$(Zu_l^\pm)(r,k) = u_l(r,k) - \int_0^\infty K^\pm(r,r',k)(Zu_l^\pm)(r',k)\, dr', \tag{3.16}$$

where

$$K^\pm(r,r',k) = q_{l2}(r',k) G_{0l}^\pm(r,r',k^2) - q_1(r',k)\frac{\partial G_{0l}^\pm}{\partial r'}(r,r',k^2), \tag{3.17}$$

$$q_{l2}(r',k) = q_{l0}(r',k) - \frac{\partial}{\partial r'} q_1(r',k).$$

Using a standard technique, one can multiply (3.16) by

$$h(r) = r^{(-3+6)/2}(1+r^{(6+\alpha)/2})^{-1}(1+r) \tag{3.18}$$

and define

$$v_l^\pm(r,k) = h(r)(Zu_l^\pm)(r,k) \tag{3.19}$$

and $v_l(r,k) = h(r)u_l(r,k)$ to obtain the integral equation

$$v_l^\pm(r,k) = v_l(r,k) - \int_0^\infty K_S^\pm(r,r',k) v_l^\pm(r',k)\, dr' \tag{3.20}$$

with the kernel

$$K_S^\pm(r,r',k) = h(r)K^\pm(r,r',k)[h(r')]^{-1}, \tag{3.21}$$

which is Hilbert–Schmidt.

Having solved (3.17) by standard methods, one can then recover $u_l^\pm(r,k)$ from (3.19). Hence we have the following theorem.

Theorem 2: With the hypothesis of Theorem 1 and the additional assumption that $V''(r)$ is continuous for $r \geq R$, $u_l^\pm(r,k)$ is related by (3.19) to the solutions $v_l^\pm(r,k)$ of an integral equation with a Hilbert–Schmidt kernel.

4. ASYMPTOTIC PROPERTIES OF Z

In conclusion we will prove the asymptotic condition (1.1). It is computationally more convenient, however, to deal with its equivalent form (2.11) and (2.13).

According to (2.13) one must prove

$$\text{s-lim}_{t\to\pm\infty} \int_0^\infty [u_l^\pm(r+\phi(r,k),k) - u_l(r,k)] \times \exp(-ik^2t)\tilde{f}(k)\, dk = 0, \tag{4.1}$$

where for convenience we take $\tilde{f}(k)$ belonging to the Schwartz $\mathcal{S}$ space on $[0,\infty)$ and vanishing in some neighborhood of the origin.

In (4.1) one has

$$u_l^\pm(r,k) = [f_l(k)\exp(il\pi/2)]^{-1}\sqrt{2/\pi}\,\Psi_l(r,k) \tag{4.2}$$

$$\Psi_l(r,k) = [f_l(r,k)f_l(k) - f_l(r,-k)f_l(-k)]/2i \tag{4.3}$$

and $u_l^- = \overline{u_l^+}$ where $\Psi_l(r,k)$ is the solution of

$$\left(-\frac{\partial^2}{\partial r^2} + \frac{l(l+1)}{r^2} + V(r)\right) y_l(r,k) = k^2 y_l(r,k), \tag{4.4}$$

which is regular at $r=0$ and satisfies $\Psi_l(0,k)=0$ and $(\partial r^{-l}\Psi_l/\partial r)(0,k)=1$, while $f_l(r,k)$ is defined as the Jost solution of (4.4) satisfying the asymptotic condition

$$\lim_{r\to\infty} f_l(r,k)\exp[-i\zeta(r,k)] = 1 \tag{4.5}$$

with

$$\zeta(r,k) = \int_0^r [k^2 - V_L(s)]^{1/2}\, ds, \tag{4.6}$$

and $f_l(k)$ is the Jost function

$$f_l(k) = k^{-1}\left[f_l(r,-k)\frac{\partial}{\partial r}\Psi_l(r,k) - \Psi_l(r,k)\frac{\partial}{\partial r} f_l(r,-k)\right]. \tag{4.7}$$

Let us define

$$a_l(r,k) = \sqrt{2/\pi}\,\sin[\zeta(r,k) - l\pi/2]. \tag{4.8}$$

We will prove (4.1) by showing that

$$\text{s-lim}_{t\to\infty} P_l^+(r,t) = 0 \tag{4.9}$$

and

$$\text{s-lim}_{t\to\infty} Q_l^+(r,t) = 0, \tag{4.10}$$

where

$$P_l^+(r,t) = \int_0^\infty \{u_l^+(r+\phi(r,k),k) - a_l(r+\phi(r,k),k)\} \times \exp(-ik^2t)\tilde{f}(k)\, dk, \tag{4.11}$$

and

$$Q_l^+(r,t) = \int_0^\infty [a_l(r+\phi(r,k),k) - u_l(r,k)] \times \exp(-ik^2t)\tilde{f}(k)\, dk. \tag{4.12}$$

We will only outline the argument which is based on the work of Matveev and Skriganov.[4]

From (4.11) one obtains the estimate

$$|P_l^+(r,t)| < \text{const}(r/t), \quad r \leq t^{1/2}, \tag{4.13}$$

which follows from an integration by parts with respect to $dk\exp(-ik^2t)$ and the estimates

$$\left|\frac{\partial}{\partial k} u_l^+(r+\phi(r,k),k)\right| < \text{const}\, r,$$
$$\left|\frac{\partial}{\partial k} a_l(r+\phi(r,k),k)\right| < \text{const}\, r. \tag{4.14}$$

For $r > t^{1/2}$ we split P_l^+ into the two terms

$$P^+_{l1}(r,t)=\sqrt{2/\pi}\int_0^\infty \{f_l(r+\phi(r,k),k) - \exp[i\xi(r+\phi(r,k),k)]\} \times\exp[-i(k^2t+l\pi/2)]\tilde f(k)\,dk/2i, \tag{4.15}$$

$$P^+_{l2}(r,t)=\sqrt{2/\pi}\int_0^\infty \{\exp[-i\xi(r+\phi(r,k),k)]-[f_l(-k)/f_l(k)] \times f_l(r+\phi(r,k),-k)\}\exp[-i(k^2t-l\pi/2)] \times\tilde f(k)\,dk/2i. \tag{4.16}$$

For t sufficiently large, one obtains the estimate

$$|P^+_{l1}(r,t)|<\text{const}\,r^{-\alpha}, \tag{4.17}$$

which follows from the estimate (see Ref. 4 with $\eta=1$ and the estimate on ϕ in Theorem 1)

$$|f_l(r+\phi(r,k),k)-\exp[i\xi(r+\phi(r,k),k)]|<[A(k)/|k|]r^{-\alpha} \tag{4.18}$$

with $A(k)$ bounded and r sufficiently large. Finally, from (4.16) one obtains the estimate

$$|P^+_l(r,t)|<\text{const}\,r^{-1}. \tag{4.19}$$

This is obtained via an integration by parts with respect to $dk\exp[-i(k^2t+\xi(r+\phi(r,k),k)]$ which yields

$$|P^+_{l2}(r,t)|<\text{const}\int_0^\infty\left|\frac{\partial}{\partial k}\left[\left(1-\frac{f_l(-k)}{f_l(k)}f_l(r+\phi,-k)\times\exp[i\xi(r+\phi,k)]\right)\tilde f(k)\left[2kt+\frac{\partial}{\partial k}\xi(r+\phi,k)\right]^{-1}\right]\right|dk \tag{4.20}$$

from which (4.19) follows since $(\partial/\partial k)\xi(r+\phi,k)=O(r)$ for $r\to\infty$.

From (4.13), (4.17), and (4.19) one has

$$\int_0^\infty|P^+_l(r,t)|^2\,dr=\left(\int_0^{t^{1/2}}+\int_{t^{1/2}}^\infty\right)|P^+_l(r,t)|^2\,dr < \text{const}\,t^{-1/2}+\text{const}\,t^{(1-2\alpha)/2}+\text{const}\,t^{-1/2}, \tag{4.21}$$

which tends to zero for $t\to\infty$ and $\alpha>\frac{1}{2}$.

An exactly similar argument proves (4.10) except that in place of the estimate (4.18) one uses the fact that for $kr\to\infty$,

$$|u_l(r,k)-\sqrt{2/\pi}\sin(kr-l\pi/2)|=O((kr)^{-1}) \tag{4.22}$$

and Eq. (3.11) which implies that

$$a_l(r+\phi(r,k),k)=\sqrt{2/\pi}\sin(kr-l\pi/2). \tag{4.23}$$

APPENDIX

The previously given[5,7] derivations of (2.15) from (1.1) were assuming that Z is a bounded operator. Since in general Z can be expected to be an unbounded operator[1-3] those proofs have to be reconsidered.

Let us recast[5] (1.1) in the form

$$\Omega^*_\pm\Psi_\pm=\text{s-}\lim_{\epsilon\to0+}(\pm\epsilon)\int_0^{\pm\infty}dt\exp(iH_0t)Z\exp[(\mp\epsilon-iH)t]\Psi_\pm \tag{A1}$$

The derivation of (2.15) from (A1) consists, in principle, in applying the spectral theorem to (A1), interchanging the order of integration in t and in the energy variable λ, and then performing explicitly the integration in t.

Since Z is unbounded, we can apply the spectral theorem only after taking a $\chi\in\mathcal{H}$ for which $\exp(-iH_0t)\chi$ is in the domain of Z^* at all $t\in\mathbb{R}^1$, and bringing Z to the left-hand side of the inner product,

$$\langle\chi|\Omega^*_\pm\Psi_\pm\rangle=\lim(\pm\epsilon)\int_0^{\pm\infty}dt \langle Z^*\exp(-iH_0t)\chi|\exp[(\mp\epsilon-iH)t]\Psi_\pm\rangle. \tag{A2}$$

Thus, in fact, we work with the Bochner integral

$$\chi_{\pm\epsilon}=\pm\epsilon\int_0^{\pm\infty}dt\exp\{(\mp\epsilon+iH)t\}Z^*\exp(-iH_0t)\chi. \tag{A3}$$

In order to prove the existence of this integral, as well as for the validation of the interchange in the order of integration in

$$\langle\chi_{\pm\epsilon}|\Psi_\pm\rangle=\pm\epsilon\int_0^{\pm\infty}dt\int_{-\infty}^{+\infty}\exp[(\mp\epsilon-i\lambda)t] \times d_\lambda\langle Z^*\exp(-iH_0t)\chi|E^H_\lambda\Psi_\pm\rangle, \tag{A4}$$

we are going to show that the vector-valued function

$$\chi(t)=Z^*\exp(-iH_0t)\chi \tag{A5}$$

is strongly continuous in $t\in\mathbb{R}^1$ and that

$$\|\chi(t)\|\le\text{const},\quad t\in\mathbb{R}^1. \tag{A6}$$

As a matter of fact, if that is the case, we can repeat verbatim the argument in Ref. 5, pp. 444–45 [with $Z^*\exp(-iH_0t)$ replacing $\exp(-iH_0t)$] and obtain that

$$\langle\chi_{\pm\epsilon}|\Psi_\pm\rangle =\pm\epsilon\int_{-\infty}^{+\infty}d_\lambda\int_0^{\pm\infty}dt\langle Z^*\exp[(\pm\epsilon+i\lambda-iH_0)t]\chi|E^H_\lambda\Psi_\pm\rangle \tag{A7}$$

$$=\pm\epsilon\int_{-\infty}^{+\infty}d_\lambda\langle(\pm\epsilon-i\lambda+iH_0)^{-1}\chi|ZE^H_\lambda\Psi_\pm\rangle, \tag{A7}$$

for vectors $\Psi_\pm$ for which $ZE^H_\lambda\Psi_\pm$ is defined at all $\lambda\in\mathbb{R}^1$. The above relation immediately implies (2.15) provided that our assumptions on χ are satisfied by a set $\mathcal{J}$ which is dense in $\mathcal{H}$.

We prove the earlier mentioned properties of $\chi(t)$ in (A5) for vectors $\chi\in\mathcal{D}_{lm}$. We write

$$\chi_{lm}(r,t)=[\exp(-iH_0t)\chi]_{lm}(r) \tag{A8}$$

and note that (since Z acts in the ranges of $\mathcal{R}_\pm$ of $\Omega_\pm$) it follows from (2.10) that in the momentum representation $(Z\Omega_\pm)^*_l\chi_{lm}(t)$ is given by

$$\tilde\chi^\pm_{lm}(k;t)=\int_0^\infty\overline{u^\pm(r+\phi_l(r,k);k)}\,\chi_{lm}(r,t)\,dr. \tag{A9}$$

To prove the strong continuity in t of (A9) [and therefore also of $\chi(t)$], we consider first

$$\hat\chi_{lm}(k;t)=(2\pi)^{-1/2}\exp[-i\pi(l+1)/2][f_l(k)]^{-1} \times\int_0^\infty[f_l(k)\exp(ikr)-f_l(-k)\exp(-ikr)]\chi_{lm}(r,t)\,dr. \tag{A10}$$

Since $|f_l(\pm k)/f_l(k)|=1$, the above function in k, con-

sidered as an element of $L^2([0,\infty))$ is obviously strongly continuous and norm-bounded in the parameter $t \in \mathbb{R}^1$. Consequently, the problem of establishing (A5) and (A6) is reduced to establishing these same two properties for the differences

$$D^{\pm}_{lm}(k;t) = \tilde{\chi}^{\pm}_{lm}(k;t) - \hat{\chi}_{lm}(k;t)$$
$$= (2\pi)^{-1/2} \exp[-i\pi(l+1)/2][f_l(k)]^{-1} \sum_{\sigma=-1}^{+1} (-1)^{(\sigma-1)/2} \times f_l(\sigma k) \int_0^\infty [f_l(r+\phi_l(r,k),\sigma k) - \exp(i\sigma k r)]\, dr \quad \text{(A11)}$$

for which an estimate of the form

$$\left(\int_0^\infty |D^{\pm}_{lm}(k;t)|^2\, dk\right)^{1/2} \leq \text{const} \sum_{\sigma=-1}^{+1} \left(\int_0^\infty dk \left| \int_0^\infty c_l(r,\sigma k)\chi_{lm}(r,t)\, dr \right|^2\right)^{1/2} \quad \text{(A12)}$$

can be derived, where

$$c_l(r,\sigma k) = f_l(r+\phi_l(r,k),\sigma k) - \exp(i\sigma k r). \quad \text{(A13)}$$

By combining (4.18) with the estimates in point (3) of Theorem 1, we obtain that

$$c_l(r,\sigma k) = O(k^{-1}). \quad \text{(A14)}$$

We show now that if $\chi_{lm}(r,t)$ is a Schwartz $\mathcal{S}$ function it can be majorized in any neighborhood of $t \neq 0$ by some t-independent function that is Lebesgue integrable in $r \in [0,\infty)$. Indeed, we have[5]

$$[\exp(-iH_0 t)\chi](\mathbf{r}) = (4\pi i t)^{-3/2} \int_{\mathbb{R}^3} \exp[(i/4t)(\mathbf{r}-\mathbf{r}')^2]\chi(\mathbf{r}')\, d\mathbf{r}'.$$

Integrating by parts we obtain

$$\left| \left(1 + \frac{r^2}{4t^2}\right)^2 [\exp(-iH_0 t)\chi](\mathbf{r}) \right|$$
$$= \left| (4\pi i t)^{-3/2} \exp\left(i\frac{\mathbf{r}^2}{4t}\right) \int_{\mathbb{R}^3} \exp\left(-i\frac{\mathbf{r}\cdot\mathbf{r}'}{2t}\right) \times (1-\nabla^2)^2 \exp\left(\frac{i\mathbf{r}'^2}{4t}\right) \chi(\mathbf{r}')\, d\mathbf{r}' \right|$$
$$\leq t^{-3/2} \sum_{n=0}^{4} \int |p_n(\mathbf{r}';t)\nabla^{2n}\chi(\mathbf{r}')|\, d\mathbf{r}', \quad \text{(A15)}$$

where p_n, $n = 0, \ldots, 4$, are polynomials in $\mathbf{r}$ of degree n having coefficients of the form $ct^{-\nu}$, $c > 0$, with ν assuming some integer value between 0 and 4.

The above result enables us to apply Lebesgue's dominated convergence theorem to the integrals in (A11) and thus conclude that $D^{\pm}_{lm}(k,t)$ is continuous in $t \neq 0$ for each fixed $k \in [0,\infty)$. In its turn, this result in conjunction with (A14) enables us to use again Lebesgue's dominated convergence theorem and infer that $D^{\pm}_{lm}(k,t)$ is strongly continuous in $t \neq 0$.

We note that the case $t = 0$ can be easily treated by suitable changes in our estimates.

The boundedness in norm (A6) for large values of $|t|$ can be obtained from (A12) by noting that for $|t| \geq 1$, (A15) yields

$$|\exp(-iH_0 t)\chi(\mathbf{r})| \leq \text{const} t^{-3/2}[1 + (r/2t)^2]^{-1},$$

since the coefficients of the polynomials p_n stay bounded as $|t| \to \infty$. Thus we get

$$\left| \int_0^\infty c_l(r,\sigma k)\chi_{lm}(r,t)\, dr \right| \leq 0(|k|^{-1}) \int_0^\infty [1 + \tfrac{1}{4}(r/t)^2]^{-1}\, d(r/t)$$

which establishes (A6).

We note that an alternative method for deriving (A6) can be applied to those vectors χ for which $\exp(iHt)Z^* \times \exp(-iH_0 t)\chi$ can be shown to converge strongly to $\Omega_{\pm}\chi$.

ACKNOWLEDGMENT

We would like to thank Dr. John Zorbas for his careful reading of the manuscript and several useful discussions.

*Supported in part by grants from the National Research Council of Canada.

[1]E. Prugovečki and J. Zorbas, J. Math. Phys. 14, 1398 (1973); Erratum *ibid.* 15, 268 (1974).
[2]E. Prugovečki and J. Zorbas, Nucl. Phys. A 213, 541 (1973).
[3]V. B. Matveev and M. M. Skriganov, Theor. Math. Phys. 10, 156 (1972).
[4]J. Zorbas, "Partial wave amplitudes for long-range scattering," University of Alberta, Preprint.
[5]E. Prugovečki, *Quantum Mechanics in Hilbert Space* (Academic, New York and London, 1971).
[6]G. Birkhoff and G.C. Rota, *Ordinary Differential Equations* (Bleisdell, Waltham, Mass., 1969), Chap. I, Theorem 6.
[7]E. Prugovečki, J. Math. Phys. 14, 957 (1973); Erratum *ibid.* 16, 442 (1975).

On the duality condition for quantum fields

Joseph J. Bisognano*, †

Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

Eyvind H. Wichmann‡

Department of Physics, University of California, Berkeley, California 94720
(Received 29 August 1975)

A general quantum field theory is considered in which the fields are assumed to be operator-valued tempered distributions. The system of fields may include any number of boson fields and fermion fields. A theorem which relates certain complex Lorentz transformations to the TCP transformation is stated and proved. With reference to this theorem, duality conditions are considered, and it is shown that such conditions hold under various physically reasonable assumptions about the fields. Extensions of the algebras of field operators are discussed with reference to the duality conditions. Local internal symmetries are discussed, and it is shown that these commute with the Poincaré group and with the TCP transformation.

I. INTRODUCTION

In an earlier paper,[1] hereafter referred to as BW I, the authors have discussed the duality condition for a Hermitian scalar field. It is the purpose of the present paper to extend the results in BW I to a general field theory, within the framework described in the monographs by Streater and Wightman[2] and by Jost.[3] We thus consider a theory in which there appears an arbitrary set of local and relatively local spinor and tensor fields. Each field has a *finite* number of components, and is assumed to be an operator-valued tempered distribution. In contrast to the situation in BW I we now have to consider fermion fields, and their characteristic anticommutation relations, which necessitates an obvious modification in the definitions of the duality conditions.

As we shall see, however, much of the reasoning in BW I applies in almost unchanged form to the issues in the present study. When this is the case we shall rely heavily on BW I, and not repeat arguments already given in that paper. The notation and terminology in BW I will be followed whenever applicable. We also refer to BW I for additional references to related work.

In Sec. II we review some aspects of the geometry of Minkowski space, and we also review some well-known facts about the quantum mechanical Poincaré group and its complex extension. In Sec. III we state our assumptions about the quantum fields, which are more or less standard. In these two sections we also explain the notation which we follow in the subsequent discussion.

The locality condition for the quantum fields is expressed in terms of the familiar (normal) commutation *and* anticommutation relations. For our purposes it would be extremely cumbersome to have to consider commutation and anticommutation relations simultaneously, and we therefore find it advantageous to restate the locality conditions in terms of the vanishing of certain *commutators*. The simple device through which this can be achieved is explained in Lemma 1 with reference to the field operators, and more generally, in Theorem 2 in Sec. V.

In Sec. IV we discuss the relationship between complex Lorentz transformations and the TCP transformation. The considerations are analogous to the considerations in Secs. III and IV in BW I, except that we now deal with spinor and tensor fields rather than with a single scalar field as in BW I. The main result in this section is presented in Theorem 1; this theorem is analogous to Theorem 1 in BW I. The form of this theorem is hardly surprising, in view of the analogous result in BW I, and some readers might feel that it would have been enough just to state the theorem. We felt, however, that an outline of the reasoning was in order and that some of the cumbersome details should be presented explicitly in writing and not left entirely to the imagination of the reader.

Sec. V in BW I was devoted to a discussion of some algebraic questions relating to Theorem 1. This discussion applies as such to the present study, and we do not repeat it here.

In Sec. V of the present paper we discuss the duality condition for the wedge regions W_R and W_L, where $W_R = \{x \mid x^3 > |x^4|\}$ and $W_L = \{x \mid x^3 < -|x^4|\}$. This discussion is analogous to the discussion in Sec. VI in BW I. The issue is the following. We wish to find two von Neumann algebras $\mathcal{A}(W_R)$ and $\mathcal{A}(W_L)$ such that $\mathcal{A}(W_R)$ can be regarded as locally associated with W_R and $\mathcal{A}(W_L)$ can be regarded as locally associated with W_L. Furthermore, the association should be consistent with the well-known TCP symmetry of the quantum fields. These notions are defined precisely in Definition 2 in Sec. V. If there are no fermion fields, then one aspect of locality is that $\mathcal{A}(W_R)$ is contained in the commutant $\mathcal{A}(W_L)'$ of $\mathcal{A}(W_L)$, and the *condition of duality* is that $\mathcal{A}(W_R) = \mathcal{A}(W_L)'$. In a theory in which fermion fields do occur these conditions have to be modified in an obvious way. The condition of duality is now that $\mathcal{A}(W_R) = (Z\mathcal{A}(W_L)Z^{-1})'$, where Z is the unitary operator defined by $Z = (I + iU_0)/(1 + i)$ in terms of the unitary operator U_0 which represents a rotation by angle 2π about any axis. In this paper we employ the notation $\mathcal{A}(W_L)^q = (Z\mathcal{A}(W_L)Z^{-1})'$, and we call $\mathcal{A}(W_L)^q$ the *quasicommutant* of the algebra $\mathcal{A}(W_L)$. The modified conditions of locality and duality are thus stated in terms of the notion of a quasicommutant. We note here that the second iterated quasicommutant is equal to the second iterated commutant, and that the

quasicommutant is equal to the commutant whenever $U_0 = I$, and hence $Z = I$. The reader who feels temporarily bewildered by the appearance of the superscript q in Secs. V and VI might find it helpful to ignore, at first, the distinction between a quasicommutant and a commutant, and hence to read the superscript q as the familiar von Neumann prime. This corresponds to the special case of no fermion fields. We feel that the modifications occasioned by the presence of fermion fields are really utterly trivial, although perhaps slightly distractive at first.

In a quantum field theory the local von Neumann algebras must be appropriately related to the field operators. Let $\mathcal{P}(W_R)$ denote the algebra of (in general unbounded) operators constructed from fields averaged with test functions with support in W_R, and let $\mathcal{P}(W_L)$ be analogously defined. A natural relationship between $\mathcal{A}(W_R)$ and $\mathcal{P}(W_R)$ is that the operators in the latter algebra shall have closed extensions affiliated to $\mathcal{A}(W_R)$, with the analogous relationship between $\mathcal{A}(W_L)$ and $\mathcal{P}(W_L)$. We have *not* been able to show that von Neumann algebras $\mathcal{A}(W_R)$ and $\mathcal{A}(W_L)$ with the above properties do exist for a *general* field theory, i. e., without further assumptions about the fields which go beyond the usual minimal assumptions. Hence we consider some special conditions on the fields which guarantee the existence of algebras $\mathcal{A}(W_R)$ and $\mathcal{A}(W_L)$ with physically satisfactory properties. Our conditions on the fields are not as such physically unreasonable, but it would clearly be desirable to settle the question of whether they are in fact necessary. The main results in Sec. V are presented in Theorems 3 and 4. We note here that these results, in the special case of a single Hermitian scalar field, are considerably stronger than our results in BW I.

In Sec. VI we discuss the construction of local von Neumann algebras associated with other regions than wedge regions in terms of algebras associated with W_R and W_L, and we show that the extended system of local algebras satisfy a condition of duality if the algebras $\mathcal{A}(W_R)$ and $\mathcal{A}(W_L)$ do. For reasons of simplicity we restrict our considerations to very special regions: double cones and their causal complements. Our results concerning the properties of the extended system of algebras in general are stated in Theorems 5 and 6. Theorem 7 describes the situation under specific assumptions about the fields. The discussion in Sec. VI is analogous to the discussion in Sec. VII in BW I, but the results in the present paper are considerably stronger than our earlier results. The paper concludes with Theorem 8, concerning local internal symmetries, in which we note that such symmetries commute with all Poincaré transformations and with the TCP transformation.

II. GEOMETRICAL PRELIMINARIES. ABOUT THE QUANTUM MECHANICAL POINCARE GROUP

Minkowski space $\mathcal{M}$ is parametrized by the customary Cartesian coordinates $x = (x^1, x^2, x^3, x^4)$. The Lorentz "metric" is so defined that $x \cdot y = x^4y^4 - x^1y^1 - x^2y^2 - x^3y^3$. The elements $\Lambda = \Lambda(M, y)$ of the *proper* Poincaré group $\bar{L}_0$ are parametrized by a 4×4 Lorentz matrix M, and a real 4-vector y, such that the image Λx of a point $x \in \mathcal{M}$ under any $\Lambda \in \bar{L}_0$ is given by $\Lambda x = \Lambda(M, y)x = Mx + y$. The image of any subset R of $\mathcal{M}$ under Λ is denoted ΛR.

The group of all 4×4 Lorentz matrices M, i. e., the group of all proper homogeneous Lorentz transformations, is denoted L_0. A rotation in L_0 by angle θ about the unit vector $\mathbf{e}$ is denoted $R(\mathbf{e}, \theta)$. We denote by $V(\mathbf{e}_3, t)$ the velocity transformation (in L_0) in the 3-direction given by

$$V(\mathbf{e}_3, t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh(t) & \sinh(t) \\ 0 & 0 & \sinh(t) & \cosh(t) \end{bmatrix}. \tag{1}$$

We define a "right wedge" W_R, and a "left wedge" W_L, as the following open subsets of Minkowski space:

$$W_R = \{x \,|\, x^3 > |x^4|\}, \quad W_L = \{x \,|\, x^3 < -|x^4|\}. \tag{2}$$

These two regions are bounded by two characteristic planes whose intersection is the 2-plane $\{x \,|\, x^3 = x^4 = 0\}$. We note that the one-parameter Abelian group of velocity transformations $V(\mathbf{e}_3, t)$, t real, maps W_R onto itself and W_L onto itself.

We next consider an *involutory* mapping $x \to \mathcal{J}x$ of Minkowski space onto itself, defined by

$$\mathcal{J}x = -R(\mathbf{e}_3, \pi)x,$$

or

$$\mathcal{J}(x^1, x^2, x^3, x^4) = (x^1, x^2, -x^3, -x^4), \tag{3}$$

where $R(\mathbf{e}_3, \pi)$ denotes the rotation by angle π about the 3-axis. We see that $\mathcal{J}$ maps W_R onto W_L, and the mapping can be described as a reflection in the common "edge" $\{x \,|\, x^3 = x^4 = 0\}$ of the pair of wedges W_R and W_L.

We note that $V(\mathbf{e}_3, t)$, as given in (1), is an entire analytic function of t. It is easily seen that

$$\mathcal{J} = V(\mathbf{e}_3, i\pi). \tag{4}$$

For *any* subset R of Minkowski space $\mathcal{M}$ we define the *causal complement* R^c of R by

$$R^c = \{x \,|\, (x - y) \cdot (x - y) < 0, \text{ all } y \in R\}. \tag{5}$$

We note that with this definition $W_R^c = \bar{W}_L$ and $W_L^c = \bar{W}_R$, where the bar denotes the closure. Two open regions R_1 and R_2 such that $R_1^c = \bar{R}_2$ *and* $R_2^c = \bar{R}_1$ form a pair of causally complementary open regions. Among such pairs the pair W_R and W_L is distinguished by the simple geometric relationships described above. Any pair of wedge-regions bounded by two nonparallel characteristic planes are distinguished in the same sense, and any such pair is in fact Poincaré-equivalent to the pair (W_R, W_L), i. e., of the form $(\Lambda W_R, \Lambda W_L)$ for some $\Lambda \in \bar{L}_0$. We shall here define $\mathcal{W}$ as the set of all (open) wedge regions bounded by two intersecting characteristic planes, i. e.,

$$\mathcal{W} = \{\Lambda W_R \,|\, \Lambda \in \bar{L}_0\}. \tag{6}$$

Although we shall at first be explicitly concerned with W_R, it is clear that analogous considerations apply to any $W \in \mathcal{W}$.

The regions W_R and W_L have further distinguishing properties, which are of crucial importance for our discussion, namely the following. Let $t = t_r + it_i$, with t_r,

t_i real. If $x \in W_R$, then the complex 4-vector $z(t) = V(\mathbf{e}_3, t)x$ is an element of the forward imaginary tube in C^4, i. e., $\mathrm{Im}(z(t)) \in V_+$, for all complex t in the open strip $0 < t_i < \pi$, and $z(t)$ is in the closure of the forward imaginary tube for all t in the closed strip $0 \leq t_i \leq \pi$. We here denote the forward lightcone with the origin as apex by V_+; the backward lightcone is denoted V_-. Similarly, if $x \in W_L$, then $z(t)$ is in the forward imaginary tube for all complex t in the strip $-\pi < t_i < 0$ and in the closed forward imaginary tube for all t in the closure of the above strip. These assertions are easily established through a simple computation. [See formula (45b) in BW I.] We note that the above facts were also of crucial importance in Jost's proof of the TCP theorem.[4]

For the reader's convenience we shall here review some well-known facts about the universal covering groups of the Lorentz and Poincaré groups, and about the complex extension of the covering group of the Lorentz group.[5]

The universal covering group of L_0, i. e., the group of all unimodular 2×2 complex matrices, is denoted $\mathcal{g}$. A specific two-to-one homomorphism of $\mathcal{g}$ onto L_0 is given by

$$g \to M(g), \quad M_{rs}(g) = \tfrac{1}{2}\mathrm{Tr}(g^\dagger \sigma_r g \sigma_s), \tag{7}$$

where $\sigma_1, \sigma_2, \sigma_3$ are the usual Pauli matrices and where $\sigma_4 = I$. The rotations and velocity transformations in $\mathcal{g}$ are denoted

$$u(\mathbf{e}, \theta) = \exp(-\tfrac{1}{2}i\theta \mathbf{e}\cdot\sigma), \quad v(\mathbf{e}, t) = \exp(\tfrac{1}{2}t\mathbf{e}\cdot\sigma), \tag{8}$$

and under the homomorphism (7) we thus have

$$R(\mathbf{e}, \theta) = M(u(\mathbf{e}, \theta)), \quad V(\mathbf{e}, t) = M(v(\mathbf{e}, t)). \tag{9}$$

The group $\mathcal{g}$ can be regarded as the complex extension of the group SU(2) of all *unitary* matrices (rotations) $u \in \mathcal{g}$, and every irreducible (unitary) representation $u \to D^s(u)$ of SU(2) can be analytically extended to a representation $g \to D^s(g)$ of $\mathcal{g}$, such that the matrix elements of $D^s(g)$ are homogeneous polynomials of degree $2s$ in the matrix elements of g. The most general finite-dimensional irreducible representation of $\mathcal{g}$ is of the form

$$g \to D^{s',s''}(g) = D^{s'}(g^r) \otimes D^{s''}(g), \tag{10}$$

where $g^r = (g^\dagger)^{-1}$. The mapping $g \to g^r$ is an outer automorphism of $\mathcal{g}$ which preserves every element in the subgroup SU(2).

In view of the complex structure of $\mathcal{g}$ it follows that the complex extension $\mathcal{g}_c$ of $\mathcal{g}$ is the direct product of $\mathcal{g}$ with itself, i. e., the group $\mathcal{g}_c = \mathcal{g} \times \mathcal{g}$ of all ordered pairs (g_1, g_2) of elements in $\mathcal{g}$ with the law of composition $(g_1', g_2')(g_1'', g_2'') = (g_1'g_1'', g_2'g_2'')$. The group $\mathcal{g}$ can be identified with a particular "real subgroup" of $\mathcal{g}_c$ through the one-one correspondence

$$g \longleftrightarrow (g^r, g). \tag{11}$$

To the set of all finite-dimensional irreducible representations $g \to D^{s',s''}(g)$ of $\mathcal{g}$ corresponds a particular family of finite-dimensional irreducible representations of $\mathcal{g}_c$, which can be regarded as the set of all finite-dimensional irreducible *analytic* representations of $\mathcal{g}_c$, namely the representations

$$(g_1, g_2) \to D_c^{s',s''}(g_1, g_2) = D^{s'}(g_1) \otimes D^{s''}(g_2). \tag{12}$$

With reference to the above definitions we define, for any complex number t, the complex velocity transformation $v_c(\mathbf{e}_3, t)$ in the 3-direction as the element

$$v_c(\mathbf{e}_3, t) = (\exp(-\tfrac{1}{2}t\sigma_3), \exp(\tfrac{1}{2}t\sigma_3)) \tag{13a}$$

of the group $\mathcal{g}_c$, and it follows from (12) that

$$D_c^{s',s''}(v_c(\mathbf{e}_3, t)) = D^{s'}(\exp(-\tfrac{1}{2}t\sigma_3)) \otimes D^{s''}(\exp(\tfrac{1}{2}t\sigma_3)). \tag{13b}$$

The matrix-valued function of t in (13b) is an entire analytic function of the complex variable t, and hence the unique analytic extension of the matrix-valued function $D^{s',s''}(v(\mathbf{e}_3, t))$ of the real variable t. We note in particular that

$$D_c^{s',s''}(v_c(\mathbf{e}_3, i\pi)) = (-1)^{2s''} D^{s',s''}(u(\mathbf{e}_3, \pi)), \tag{14a}$$

$$D_c^{s',s''}(v_c(\mathbf{e}_3, -i\pi)) = (-1)^{2s'} D^{s',s''}(u(\mathbf{e}_3, \pi)). \tag{14b}$$

The formula $V(\mathbf{e}_3, i\pi) = -R(\mathbf{e}_3, \pi)$ is a special case of (14a) (with $s' = s'' = \frac{1}{2}$), and with M_c denoting the analytic extension of the representation $g \to M(g)$ to the complex group $\mathcal{g}_c$ we have $M_c(v_c(\mathbf{e}_3, t)) = V(\mathbf{e}_3, t)$ for all complex t.

The universal covering group of $\bar{L}_0$ is denoted $\bar{\mathcal{g}}$. The elements $\lambda = \lambda(g, x)$ are the ordered pairs consisting of any $g \in \mathcal{g}$ and any $x \in \mathcal{H}$, with the law of composition $\lambda(g', x')\lambda(g'', x'') = \lambda(g'g'', x' + M(g')x'')$. We define an explicit homomorphism $\lambda \to \Lambda(\lambda)$ by $\Lambda(\lambda(g, x)) = \Lambda(M(g), x)$.

The Hilbert space $\mathcal{H}$ of physical states is assumed to be separable. It is assumed to carry a strongly continuous unitary representation $\lambda \to U(\lambda)$ of the quantum mechanical Poincaré group $\bar{\mathcal{g}}$. We write $U(g, x) = U(\lambda(g, x))$, and we also employ the special notation $T(x) = U(I, x)$ for the translations. The translations have the common spectral resolution

$$T(x) = U(I, x) = \int \exp(ix \cdot p)\mu(d^4p), \tag{15}$$

and it is assumed that the support of the spectral measure μ is contained in the closed forward lightcone $\bar{V}_+$ (in momentum space). This assumption about the support of μ will be referred to as the "spectral condition" in what follows.

We assume the existence of a vacuum state, represented by the unit vector Ω, *uniquely* characterized by its invariance under all translations. The vacuum state then satisfies $U(\lambda)\Omega = \Omega$ for all $\lambda \in \bar{\mathcal{g}}$. It is well known that the spectral condition allows the extension of the representation of the translation subgroup to a unique representation $z \to T(z)$ of the semigroup of complex translations for which $\mathrm{Im}(z) \in \bar{V}_+$, such that $T(z)$ is a bounded and strongly continuous function of z in the closed forward imaginary tube, and a strongly analytic function of z in the open forward imaginary tube.

The one-parameter group of velocity transformations in the 3-direction, as well as its analytic extension to the complex domain, will be of particular interest, and we shall therefore employ the shorter notation $V(t) = U(v(\mathbf{e}_3, t), 0)$ for the representatives of these velocity transformations. More generally we shall write

$$V(\tau) = \exp(-i\tau K_3) = \int \exp(-i\tau s)\mu_K(ds) \tag{16}$$

for any complex τ. Here μ_K is the spectral measure in

the simultaneous spectral resolution of the group of all $V(t)$, t real, and K_3 is the unique self-adjoint operator, with domain D_K, such that $V(t) = \exp(-itK_3)$. For a discussion of the domains of the normal operators $V(\tau)$ we refer to Sec. IV in BW I. We denote (as in BW I) by D_+ the domain on which $V(i\pi)$ is self-adjoint and by D_- the domain on which $V(-i\pi)$ is self-adjoint.

III. ASSUMPTIONS ABOUT THE QUANTUM FIELDS

We denote by $\mathcal{D}(R^n)$ the set of all complex-valued infinitely differentiable functions of compact support on n-dimensional Euclidean space R^n, and we denote by $\mathcal{S}(R^n)$ the space of test functions on R^n in terms of which tempered distributions are defined. The space $\mathcal{S}(R^n)$ is regarded as endowed with the particular topology appropriate to the definition of tempered distributions.[6]

For an unbounded linear or antilinear operator X defined on a domain D we shall employ the unorthodox notation (X, D), as in BW I. The adjoint of (X, D) is denoted $(X, D)^* = (X^*, D(X^*))$, where $D(X^*)$ is the domain of the adjoint. This notation will not be employed for manifestly bounded operators, for which the domain is taken to be the entire Hilbert space $\mathcal{H}$.

We shall next state the basic assumptions about the quantum fields. It is not our aim here to state a set of minimal independent assumptions for a field theory, but rather to describe the situation which prevails in a standard field theory.

(a) We assume the existence of a set of *boson fields* $\beta^{(b)}(x)$, where b is an element in an index set I_B, and a set of *fermion fields* $\phi^{(f)}(x)$, where f is an element in an index set I_F. The index sets are regarded as disjoint, and it is assumed that at least one of these sets is nonempty; otherwise they are arbitrary. We thus admit as possible special cases the cases when either I_B, or else I_F is empty. Each field $\beta^{(b)}(x)$ or $\phi^{(f)}(x)$ has a *finite* number of components, denoted $\beta_\mu^{(b)}(x)$, respectively $\phi_\mu^{(f)}(x)$, where μ is a suitable index distinguishing between the components.

(b) We also consider the set of all components of all the fermion fields and all the boson fields. An element in this set is denoted $\varphi_\mu(x)$, where μ is an element in an index set I_T such that when μ runs through I_T each component of each field is obtained once and once only. Each component $\varphi_\mu(x)$ is an operator-valued tempered distribution in the following sense. To each $f(x) \in \mathcal{S}(R^4)$, and each $\mu \in I_T$, corresponds a closable linear operator $(\varphi_\mu[f], D_1)$ on a dense domain D_1 (independent of f and μ) such that $\varphi_\mu[f]D_1 \subset D_1$. The mapping $f \to (\varphi_\mu[f], D_1)$ is linear, and for any $\xi \in D_1$ the vector $\varphi_\mu[f]\xi$ is a strongly continuous function of f on $\mathcal{S}(R^4)$.

Furthermore, if $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ is any ordered n-tuplet of indices from I_T, then there corresponds to every $f(x_1, x_2, \ldots, x_n) \in \mathcal{S}(R^{4n})$ a closable linear operator $(\varphi\{f;\sigma\}, D_1)$ on D_1 such that $\varphi\{f;\sigma\}D_1 \subset D_1$. The mapping $f \to (\varphi\{f;\sigma\}, D_1)$ is linear, and for any $\xi \in D_1$ the vector $\varphi\{f;\sigma\}\xi$ is a strongly continuous function of f on $\mathcal{S}(R^{4n})$. If f is of the particular form $f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$, with $f_k \in \mathcal{S}(R^4)$ for $k = 1, \ldots, n$, then, on D_1,

$$\varphi\{f;\sigma\} = \varphi_{\mu 1}[f_1]\varphi_{\mu 2}[f_2]\cdots\varphi_{\mu n}[f_n]. \tag{17}$$

This is consistent with the common notation for $\varphi\{f;\sigma\}$ in terms of the symbolic integral at right in

$$\varphi\{f;\sigma\} = \int_{(\infty)} d^4(x_1)d^4(x_2)\cdots d^4(x_n) \times f(x_1, x_2, \ldots, x_n)\varphi_{\mu 1}(x_1)\varphi_{\mu 2}(x_2)\cdots\varphi_{\mu n}(x_n). \tag{18}$$

(c) Let $\mathcal{P}(\mathcal{H})$ be the algebra, defined on D_1, which is the linear span of the identity I and all operators $(\varphi\{f;\sigma\}, D_1)$. The dense domain D_1 is assumed to be *precisely* equal to $\mathcal{P}(\mathcal{H})\Omega$.

(d) For any field component $\varphi_\mu(x)$ there exists a field component $\varphi_{\mu'}(x)$ such that for any $f \in \mathcal{S}(R^4)$

$$(\varphi_{\mu'}[f^*], D_1)^* \supset (\varphi_\mu[f], D_1). \tag{19}$$

The field component $\varphi_{\mu'}(x)$ is then also denoted $\varphi_\mu^\dagger(x)$.

(e) The domain D_1 is invariant under $\bar{\mathcal{P}}$, i.e., $U(\lambda)D_1 = D_1$, for any $\lambda \in \bar{\mathcal{P}}$. The action of $U(\lambda)$ by conjugation on the elements of $\mathcal{P}(\mathcal{H})$ is specified by the conditions

$$(\alpha)\quad T(x')\varphi_\mu(x)T(x')^{-1} = \varphi_\mu(x + x') \tag{20a}$$

for any field component $\varphi_\mu(x)$.

(β) For each $b \in I_B$,

$$U(g, 0)\beta_\mu^{(b)}(x)U(g, 0)^{-1} = \sum_{\mu'} \Gamma_{\mu\mu'}^{(b)}(g^{-1})\beta_{\mu'}^{(b)}(M(g)x), \tag{20b}$$

where $g \to \Gamma^{(b)}(g)$ is similar to one of the representations $g \to D^{s',s''}(g)$ for which $2(s' + s'')$ is an *even* integer.

(γ) For each $f \in I_F$,

$$U(g, 0)\phi_\mu^{(f)}(x)U(g, 0)^{-1} = \sum_{\mu'} \Gamma_{\mu\mu'}^{(f)}(g^{-1})\phi_{\mu'}^{(f)}(M(g)x), \tag{20c}$$

where $g \to \Gamma^{(f)}(g)$ is similar to one of the representations $g \to D^{s',s''}(g)$ for which $2(s' + s'')$ is an *odd* integer.

The sums at right in (20b) and (20c) extend over all the components of the field $\beta^{(b)}(x)$, respectively the field $\phi^{(f)}(x)$.

(f) All the fields satisfy the *normal conditions of locality*, i.e., they satisfy the conditions (in the sense of distributions)

$$\begin{aligned} &[\beta_\mu^{(b)}(x), \beta_{\mu'}^{(b')}(x')] = 0, \\ &[\beta_\mu^{(b)}(x), \phi_{\mu'}^{(f')}(x')] = 0, \\ &\{\phi_\mu^{(f)}(x), \phi_{\mu'}^{(f')}(x')\} = 0 \end{aligned} \tag{21}$$

on D_1 for all spacelike $x - x'$. Here the curly bracket denotes the anticommutator, i.e., $\{X, X'\} = XX' + X'X$.

The above formulation of the basic assumptions about the fields is more or less standard. The essence of the notion of a set of quantum fields is a certain kind of representation of a tensor algebra of multicomponent test functions by an operator algebra $\mathcal{P}(\mathcal{H})$. The precise formulation of a general field theory is unfortunately beset by considerable notational difficulties. We have tried to select a notation which is convenient for our particular purposes. Let us now elaborate further on the basic assumptions, and on some well-known immediate consequences.

(g) Whether the number of fields is finite, countably

infinite, or uncountably infinite is immaterial for the conclusions which we shall draw. That each field $\beta^{(b)}(x)$ or $\phi^{(f)}(x)$ has only a *finite* number of components, where the notion of "component," of course, refers specifically to the transformation laws (20b) and (20c), is, however, *essential*. Our purpose with introducing the specific "irreducible fields" $\beta^{(b)}(x)$ and $\phi^{(f)}(x)$ was to be able to state the transformation laws (20b) and (20c), as well as the locality conditions (21), with maximum clarity. For the subsequent discussion it will, however, be more convenient to employ a unified notation, in terms of the symbols $\varphi_\mu(x)$, for all the field components, and we shall therefore restate the conditions (20b) and (20c) in the form

$$U(g,0)\varphi_\mu(x)U(g,0)^{-1} = \sum_{\mu'} \Gamma_{\mu\mu'}(g^{-1})\varphi_{\mu'}(M(g)x) \tag{22}$$

The "matrix" $\Gamma(g)$ can be regarded as the direct sum of the finite-dimensional matrices $\Gamma^{(b)}(g)$ and $\Gamma^{(f)}(g)$ in an obvious sense. The sum in (22) is always a finite sum, and for each fixed μ (or each fixed μ') there is only a finite number of values of μ' (respectively of μ) for which $\Gamma_{\mu\mu'}$ is different from zero. We shall also consider the analytic extension of the representation $g \to \Gamma(g)$ of $\mathcal{G}$ to a representation $(g_1, g_2) \to \Gamma(g_1, g_2)$ of $\mathcal{G}_c$, defined as the direct sum of the corresponding analytic extensions of the representations $\Gamma^{(b)}(g)$ and $\Gamma^{(f)}(g)$ as described in Sec. II. To the complex velocity transformation $v_c(\mathbf{e}_3, t)$ thus corresponds the representative $\Gamma(v_c(\mathbf{e}_3, t))$, each matrix element of which is an entire analytic function of the complex variable t. With reference to this extension we thus define the *diagonal* "matrix" Γ'' (with eigenvalues $+1$ and -1) by

$$\Gamma'' = \Gamma(v_c(\mathbf{e}_3, -i\pi))\Gamma(u(\mathbf{e}_3, \pi)). \tag{23}$$

That Γ'' has the stated properties follows at once from (14a).

(h) The domain D_1 on which the "averaged fields" and the operators in $\mathcal{P}(\mathcal{M})$ are defined should be carefully noted. It follows readily from our assumptions that for any $(X, D_1) \in \mathcal{P}(\mathcal{M})$ the domain of the adjoint $(X, D_1)^*$ contains D_1. The restriction of the adjoint to D_1 shall be denoted $(X^\dagger, D_1)$, and called the Hermitian conjugate of X; the notion of the Hermitian conjugate of a field operator thus depends on the specific choice of D_1. It also follows from our assumptions that $(X^\dagger, D_1) \in \mathcal{P}(\mathcal{M})$ for all $(X, D_1) \in \mathcal{P}(\mathcal{M})$. In particular the Hermitian conjugate $\varphi_\mu[f]^\dagger$ of the averaged field $\varphi_\mu[f]$ is the averaged field $\varphi_\mu^\dagger[f^*]$. The mapping $(X, D_1) \to (X^\dagger, D_1)$ is an antilinear involution of $\mathcal{P}(\mathcal{M})$ [such that $(X_1X_2)^\dagger = X_2^\dagger X_1^\dagger$].

We note that every operator $(X, D_1) \in \mathcal{P}(\mathcal{M})$ satisfies

$$(X^\dagger, D_1)^{**} \subset (X, D_1)^* . \tag{24}$$

It is a hitherto unsolved problem whether the assumptions which we have made imply that the inclusion in (24) can be replaced by equality for some nontrivial set of operators in $\mathcal{P}(\mathcal{M})$.

(i) Let R be any subset of Minkowski space $\mathcal{M}$. We define $\mathcal{P}_0(R)$ as the polynomial algebra generated by the identity operator I and all operators $(\varphi_\mu[f], D_1)$, with $\mu \in I_T$, $f(x) \in \mathcal{S}(R^4)$ and $\mathrm{supp}(f) \subset R$. We define the algebra $\mathcal{P}(R)$ as the linear span of I and all operators $(\varphi\{f;\sigma\}, D_1)$, where $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ is any n-tuplet of indices in I_T, and where $f(x_1, x_2, \ldots, x_n) \in \mathcal{S}(R^{4n})$ with $\mathrm{supp}(f) \subset (\times R)^n$.

It is easily seen that $(X, D_1) \to (X^\dagger, D_1)$ is an involution of both $\mathcal{P}_0(R)$ and $\mathcal{P}(R)$. From the conditions (20a)–(20c) it follows that

$$U(\lambda)\,\mathcal{P}_0(R)U(\lambda)^{-1} = \mathcal{P}_0(\Lambda(\lambda)R),$$
$$U(\lambda)\,\mathcal{P}(R)U(\lambda)^{-1} = \mathcal{P}(\Lambda(\lambda)R) \tag{25}$$

for any $\lambda \in \bar{\mathcal{P}}$ and any R.

We trivially have $\mathcal{P}_0(R) \subset \mathcal{P}(R) \subset \mathcal{P}(\mathcal{M})$. According to a well-known theorem of Reeh and Schlieder[7] the linear manifold $\mathcal{P}_0(R)\Omega$ is dense in $\mathcal{H}$ for any open nonempty R.

(j) Let the unitary operators U_0 and Z be defined by

$$U_0 = U(-I, 0), \quad Z = (I + iU_0)/(1 + i). \tag{26}$$

These operators trivially satisfy

$$U_0^2 = I, \quad Z^2 = U_0, \quad U(\lambda)U_0U(\lambda)^{-1} = U_0, \quad U(\lambda)ZU(\lambda)^{-1} = Z \tag{27a}$$

and

$$U_0\Omega = Z\Omega = \Omega, \quad U_0D_1 = D_1, \quad ZD_1 = D_1. \tag{27b}$$

Furthermore, it follows from the assumptions in (e) above that

$$U_0\beta_\mu^{(b)}(x) = \beta_\mu^{(b)}(x)U_0, \quad Z\beta_\mu^{(b)}(x)Z^{-1} = \beta_\mu^{(b)}(x), \tag{28a}$$
$$U_0\phi_\mu^{(f)}(x) = -\phi_\mu^{(f)}(x)U_0, \quad Z\phi_\mu^{(f)}(x)Z^{-1} = iU_0\phi_\mu^{(f)}(x) \tag{28b}$$

for all boson fields $\beta^{(b)}(x)$ and all fermion fields $\phi^{(f)}(x)$.

The fact that the involution U_0 commutes with all boson fields, but anticommutes with all fermion fields permits a unique resolution of any field operator into a sum of a "boson operator" and a "fermion operator," and it also permits a restatement of the locality conditions (21) in terms of the vanishing of certain *commutators*. We shall state the important facts in the matter in the form of a lemma for later reference.

Lemma 1: (a) Let U_0 and Z be defined as in (26). For any subset R of $\mathcal{M}$, let

$$\mathcal{P}_B(R) = \{(X, D_1) \mid U_0XU_0 = X,\ (X, D_1) \in \mathcal{P}(R)\}, \tag{29a}$$
$$\mathcal{P}_F(R) = \{(X, D_1) \mid U_0XU_0 = -X,\ (X, D_1) \in \mathcal{P}(R)\}. \tag{29b}$$

Then every $(X, D_1) \in \mathcal{P}(R)$ has a unique resolution of the form

$$X = X_b + X_f, \quad X_b \in \mathcal{P}_B(R),\ X_f \in \mathcal{P}_F(R), \tag{30a}$$

where, in fact,

$$X_b = \tfrac{1}{2}(X + U_0XU_0), \quad X_f = \tfrac{1}{2}(X - U_0XU_0). \tag{30b}$$

The sets $\mathcal{P}_B(R)$ and $\mathcal{P}_F(R)$ are mapped onto themselves under the involution $(X, D_1) \to (X^\dagger, D_1)$. Furthermore,

$$ZX_bZ^{-1} = X_b, \quad ZX_fZ^{-1} = iU_0X_f \tag{31}$$

for all $X_b \in \mathcal{P}_B(R)$ and all $X_f \in \mathcal{P}_F(R)$.

(b) For any $(X, D_1) \in \mathcal{P}(R)$, let (X^z, D_1) be defined by

$$(X^z, D_1) = Z(X, D_1)Z^{-1} = (ZXZ^{-1}, D_1). \tag{32}$$

If R_1 and R_2 are two open subsets of $\mathcal{M}$ such that $R_1 \subset R_2^c$, then it follows from the locality conditions in (f) above that

$$[X_b, Y_b]=0, \quad [X_b, Y_f]=0, \quad [X_f, Y_b]=0, \quad \{X_f, Y_f\}=0 \tag{33a}$$

on D_1 for all $X_b \in \mathcal{P}_B(R_1)$, $X_f \in \mathcal{P}_F(R_1)$, $Y_b \in \mathcal{P}_B(R_2)$, and $Y_f \in \mathcal{P}_F(R_2)$. The conditions (33a) are equivalent to the condition

$$[X, Y^z]=0 \tag{33b}$$

on D_1 for all $X \in \mathcal{P}(R_1)$, $Y \in \mathcal{P}(R_2)$.

We omit the completely trivial proof. We note that the lemma is vacuous if $U_0 = I$, which is the case if and only if there is no fermion field.

IV. COMPLEX LORENTZ TRANSFORMATIONS AND THE *TCP* TRANSFORMATION

In this section we shall present the generalizations appropriate for the present situation of the considerations in Secs. III and IV in BW I. The main result is presented in Theorem 1, which corresponds to Theorem 1 in BW I. As in BW I we arrive at the main conclusion through a sequence of lemmas, arranged in such a way that the similarities with the discussion in BW I are pretty obvious.

For any $f(x_1, x_2, \ldots, x_n) \in \mathcal{S}(R^{4n})$ we define a Fourier transform $\tilde{f}$ by

$$\tilde{f}(p_1, \ldots, p_n) = \int_{(\infty)} d^4(x_1) \cdots d^4(x_n) f(x_1, \ldots, x_n) \exp\left(i \sum_{r=1}^{n} x_r \cdot p_r\right) . \tag{34}$$

For any positive integer n we denote by T_n the open tube region

$$T_n = \{(z_1, z_2, \ldots, z_n) \mid \mathrm{Im}(z_k) \in V_+, \ k = 1, \ldots, n\} \tag{35}$$

in complex $4n$-dimensional space, regarded as a direct sum of n replicas of complex Minkowski space and parametrized by an n-tuplet $(z_1, z_2, \ldots, z_n)$ of complex 4-vectors. The closure of T_n is denoted $\overline{T}_n$.

Lemma 2: Let $z \in \overline{T}_1$, i.e., z is any complex 4-vector in the closed forward imaginary tube. Then

(a) $$T(z) D_1 \subset D_1 \tag{36a}$$

(b) If $f \in \mathcal{S}(R^{4n})$ there exists an $f_z \in \mathcal{S}(R^{4n})$ such that

$$\tilde{f}_z(p_1, \ldots, p_n) = \tilde{f}(p_1, \ldots, p_n) \exp\left(iz \cdot \sum_{r=1}^{n} p_r\right) \tag{36b}$$

for $(p, \ldots, p_n) \in V_n$, where V_n is the subset of R^{4n} defined by

$$V_n = \left\{(p_1, \ldots, p_n) \mid \sum_{r=k}^{n} p_r \in \overline{V}_+, \quad k = 1, \ldots, n\right\} \tag{36c}$$

and for every such f_z we have

$$T(z) \varphi\{f; \sigma\}\Omega = \varphi\{f_z; \sigma\}\Omega, \tag{36d}$$

where σ is any ordered n-tuplet $(\mu 1, \mu 2, \ldots, \mu n)$ of indices from I_T.

Lemma 3: (a) For each $n \geq 1$, let E_n be the set of all functions $f(x_1, \ldots, x_n; z_1, \ldots, z_n)$ defined for $(x_1, \ldots, x_n) \in R^{4n}$ and $(z_1, \ldots, z_n) \in T_n$, such that $f \in \mathcal{S}(R^{4n})$ and such that the Fourier transform $\tilde{f}$ of f relative to the variables $(x_1, \ldots, x_n)$ satisfies the condition

$$\tilde{f}(p_1, \ldots, p_n; z_1, \ldots, z_n) = \exp\left(i \sum_{k=1}^{n} \sum_{r=k}^{n} z_k \cdot p_r\right) \tag{37a}$$

for all $(p_1, \ldots, p_n) \in V_n$, with V_n defined as in (36c). The set E_n is nonempty, and to every n-tuplet $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ of indices from I_T corresponds a *unique* vector-valued function $\phi(z_1, z_2, \ldots, z_n; \sigma)$ on T_n, defined by

$$\phi(z_1, z_2, \ldots, z_n; \sigma) = \varphi\{f; \sigma\}\Omega, \tag{37b}$$

where f is any element of E_n.

(b) The vector-valued function $\phi(z_1, z_2, \ldots, z_n; \sigma)$ is a strongly analytic function of $(z_1, z_2, \ldots, z_n)$ on T_n, and for each point in this domain it is an analytic vector for the Lie algebra of the group $U(\bar{\mathcal{P}})$.

(c) For any element $\lambda = \lambda(g, x)$ of the quantum mechanical Poincaré group $\bar{\mathcal{P}}$,

$$U(\lambda)\phi(z_1, z_2, \ldots, z_n; \sigma) = \sum_{\sigma'} \hat{\Gamma}_{\sigma, \sigma'}(g^{-1}) \phi(Mz_1 + x, Mz_2, Mz_3, \ldots, Mz_n; \sigma'), \tag{37c}$$

where $M = M(g)$, and where the sum is over the *finite* number of n-tuplets $\sigma' = (\mu 1', \mu 2', \ldots, \mu n')$ of indices from I_T for which

$$\hat{\Gamma}_{\sigma, \sigma'}(g) = \Gamma_{\mu 1, \mu 1'}(g) \Gamma_{\mu 2, \mu 2'}(g) \cdots \Gamma_{\mu n, \mu n'}(g) \tag{37d}$$

is not identically zero (as a function of g).

It may here be noted that

$$\varphi_{\mu 1}(z_1) \varphi_{\mu 2}(z_1 + z_2) \cdots \varphi_{\mu n}(z_1 + z_2 + \cdots + z_n)\Omega \tag{37e}$$

is a defensible notation (within the framework of distribution theory) for the vector $\phi(z_1, z_2, \ldots, z_n; \sigma)$.

Lemma 4: (a) Let $\{f_k \mid f_k \in \mathcal{S}(R^4), \ k = 1, \ldots, n\}$ be any n-tuplet of test functions, and let $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ be any ordered n-tuplet of indices from I_T. For $k = 1, \ldots, n$, let $X_k = \varphi_{\mu k}[f_k]$. Then the vector

$$T(z_1) X_1 T(z_2) X_2 \cdots T(z_n) X_n \Omega \tag{38a}$$

is well defined (through successive left multiplications) for all $(z_1, z_2, \ldots, z_n) \in \overline{T}_n$, and it is a strongly continuous function of the variables $(z_1, z_2, \ldots, z_n)$ on $\overline{T}_n$ and a strongly analytic function of these variables on T_n.

(b) There exist functions $f(x_1, \ldots, x_n; z_1, \ldots, z_n)$ defined for $(x_1, \ldots, x_n) \in R^{4n}$ and $(z_1, \ldots, z_n) \in \overline{T}_n$, such that $f \in \mathcal{S}(R^{4n})$ and such that the Fourier transform $\tilde{f}$ of f relative to the variables $(x_1, \ldots, x_n)$ satisfies the condition

$$\tilde{f}(p_1, \ldots, p_n; z_1, \ldots, z_n) = \exp\left(i \sum_{k=1}^{n} \sum_{r=k}^{n} z_k \cdot p_r\right) \prod_{k=1}^{n} \tilde{f}_k(p_k) \tag{38b}$$

for all $(p_1, \ldots, p_n) \in V_n$, with V_n defined as in (36c), and for all $(z_1, z_2, \ldots, z_n) \in \overline{T}_n$. For any such function f,

$$\varphi\{f; \sigma\}\Omega = T(z_1) X_1 T(z_2) X_2 \cdots T(z_n) X_n \Omega. \tag{38c}$$

(c) If $f_k \in \mathcal{D}(R^4)$ for $k = 1, 2, \ldots, n$, and $(z_1, z_2, \ldots, z_n) \in T_n$, then

$$\int_{(\infty)} d^4(x_1) \cdots d^4(x_n) f_1(x_1) f_2(x_2) \cdots f_n(x_n) \times \phi(z_1 + x_1, z_2 + x_2 - x_1, z_3 + x_3 - x_2, \ldots, z_n + x_n - x_{n-1}; \sigma) = T(z_1) X_1 T(z_2) X_2 \cdots T(z_n) X_n \Omega. \tag{38d}$$

(d) Let $\{R_n \mid n = 1, \ldots, \infty\}$ be any set of open, nonempty subsets of Minkowski space. For such a set, and for any $n \geq 1$, let S_n denote the linear span of all vectors of

the form $X_1X_2\cdots X_n\Omega$, with X_k defined as in (a) above, and with $f_k \in \mathcal{S}(R^4)$, $\text{supp}(f_k) \subset R_k$, for $k=1,\ldots,n$.

Then the linear span of the vacuum vector Ω and the union of all the linear manifolds S_n is dense in the Hilbert space $\mathcal{H}$.

About the proofs: Lemmas 2–4 in the present paper correspond to Lemmas 2–6 in Sec. III of BW I, and the reasoning there presented applies with very trivial modifications. The conclusions in Lemmas 2 and 4; the conclusion in part (a) of Lemma 3, and the conclusion [in part (b) of Lemma 3] that $\phi(z_1, z_2, \ldots, z_n; \sigma)$ is analytic as asserted, follow from the spectral condition, the action of the translation group by conjugation on the fields, and the assumption that the fields are *tempered* distributions on the domain D_1. That we now deal with an arbitrary number of field components instead of with a single field as in BW I is immaterial in the proofs. The formula (37c) is the trivial generalization of the formula (34) in BW I. Since the matrix $\hat{\Gamma}(g^{-1})$ in (37c) is in effect similar to a *finite* direct sum of matrices $D^{s',s''}(g^{-1})$, and hence an entire analytic function of g, it follows that $\phi(z_1, z_2, \ldots, z_n; \sigma)$ is an analytic vector for the Lie algebra of the group $U(\mathcal{J}, 0)$, and hence also for the Lie algebra of the group $U(\mathcal{J})$.

We next consider the action of the complex velocity transformations $V(t) = \exp(-itK_3)$, where t is complex, on the vectors $\phi(z_1, z_2, \ldots, z_n; \sigma)$. We denote by $D_V(\pi/2)$ the domain on which $V(i\pi/2)$ is self-adjoint and by $D_V(-\pi/2)$ the domain on which $V(-i\pi/2)$ is self-adjoint. The domain $D_V(\pi/2)$ is then a core for all operators $V(t)$ with $0 \leq \text{Im}(t) \leq \pi/2$, and the domain $D_V(-\pi/2)$ is a core for all operators $V(t)$ with $0 \geq \text{Im}(t) \geq -\pi/2$. The next lemma corresponds to Lemmas 8 and 9 in BW I, and it is proved, on the basis of Lemma 3, by a very trivial modification of the reasoning in BW I.

Lemma 5: Let $(z_1, \ldots, z_n)$ be an n-tuplet of complex 4-vectors $z_k = x_k + iy_k$, where x_k, y_k are real, $y_k^1 = y_k^2 = 0$, $y_k^4 > |y_k^3|$, for $k=1,\ldots,n$. Let $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ be any ordered n-tuplet of indices from I_T. For any k and any complex t we define $z_k(t)$ by

$$z_k(t) = V(\mathbf{e}_3, t) z_k. \tag{39a}$$

(a) If $x_k \in W_R$ (i. e., $x_k^3 > |x_k^4|$), for $k=1,\ldots,n$, then $(z_1(i\tau), \ldots, z_n(i\tau)) \in T_n$ for all $\tau \in [0, \pi/2]$. The vector $\phi(z_1, \ldots, z_n; \sigma)$ is in the domain $D_V(\pi/2)$, and

$$V(i\tau)\phi(z_1, \ldots, z_n; \sigma) = \sum_{\sigma'} \hat{\Gamma}_{\sigma,\sigma'}(v_c(\mathbf{e}_3, -i\tau))\phi(z_1(i\tau), \ldots, z_n(i\tau); \sigma') \tag{39b}$$

for all $\tau \in [0, \pi/2]$, where $\hat{\Gamma}$ is defined as in (37d).

(b) If $x_k \in W_L$ (i. e., $x_k^3 < -|x_k^4|$), for $k=1,\ldots,n$, then $(z_1(i\tau), \ldots, z_n(i\tau)) \in T_n$ for all $\tau \in [-\pi/2, 0]$. The vector $\phi(z_1, \ldots, z_n; \sigma)$ is in the domain $D_V(-\pi/2)$, and the relation (39b) holds for all $\tau \in [-\pi/2, 0]$.

(c) Let $(x_1, \ldots, x_n)$ be such that $x_k \in W_R$ for $k=1,\ldots,n$. Let v be the real forward timelike 4-vector with components $v = (0, 0, 0, 1)$, and let t be a real variable. Then

$$\text{s-}\lim_{t\to 0+} \sum_{\sigma'} \hat{\Gamma}_{\sigma,\sigma'}(c_+) V(i\pi/2)\phi(x_1 + itv, x_2 + itv, \ldots, x_n + itv; \sigma') = \text{s-}\lim_{t\to 0+} \sum_{\sigma'} \hat{\Gamma}_{\sigma,\sigma'}(c_-) V(-i\pi/2)$$
$$\times \phi(\mathcal{J}x_1 + itv, \mathcal{J}x_2 + itv, \ldots, \mathcal{J}x_n + itv; \sigma') = \phi(z_1, \ldots, z_n; \sigma), \tag{39c}$$

where $z_k = (x_k^1, x_k^2, ix_k^4, ix_k^3)$, for $k=1,\ldots,n$, and where c_+ and c_- are the elements $c_+ = v_c(\mathbf{e}_3, i\pi/2)$, $c_- = v_c(\mathbf{e}_3, -i\pi/2)$, of the group $\mathcal{J}_c$. Here $\mathcal{J}$ is defined as in (3).

The next lemma corresponds to Lemma 10 in BW I.

Lemma 6: Let R_1 be a bounded, open, nonempty subset of W_R, and let $x_0 \in W_R$ be such that $(x - x_0) \in W_L$ for all $x \in R_1$. For any integer $n > 1$ we define the set R_n by

$$R_n = \{x + (n-1)x_0 \,|\, x \in R_1\}. \tag{40a}$$

(a) Then $R_n \subset W_R$ for all n, and if $n > k$, then $(x' - x'') \in W_R$ for all $x' \in R_n$, $x'' \in R_k$. In particular R_n is spacelike separated from R_k (i. e., $R_n \subset R_k^c$) if $n \neq k$.

(b) Let $\{f_k \,|\, k=1,\ldots,n\}$ be an n-tuplet of test functions such that $f_k \in \mathcal{S}(R^4)$ and $\text{supp}(f_k) \subset R_k$, for $k=1,\ldots,n$. Let f_k^i denote the test function defined by $f_k^i(x) = f_k(-x)$. Let $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ be any ordered n-tuplet of indices from I_T. Let $c(s) \in \mathcal{D}(R^1)$. Then

$$V(i\pi)c(K_3)\varphi_{\mu 1}[f_1]\varphi_{\mu 2}[f_2]\cdots\varphi_{\mu n}[f_n]\Omega = \hat{\Gamma}''_{\sigma,\sigma} U(u(\mathbf{e}_3, \pi), 0)c(K_3)\varphi_{\mu 1}[f_1^i]\varphi_{\mu 2}[f_2^i]\cdots\varphi_{\mu n}[f_n^i]\Omega, \tag{40b}$$

where $\hat{\Gamma}''$ is the *diagonal* matrix given by

$$\hat{\Gamma}'' = \hat{\Gamma}(v_c(\mathbf{e}_3, -i\pi))\hat{\Gamma}(u(\mathbf{e}_3, \pi)). \tag{40c}$$

This lemma can be proved, on the basis of Lemmas 4 and 5, by a trivial modification of the reasoning by which we proved Lemma 10 in BW I; the modification, of course, has to do with the appearance of the matrices $\hat{\Gamma}$ in the formulas. To bring out the similarities with the discussion in BW I, we define the test function f_k^j by $f_k^j(x) = f_k(\mathcal{J}x)$, and we then have

$$U(u(\mathbf{e}_3, \pi), 0)\varphi_{\mu k}[f_k^i]U(u(\mathbf{e}_3, \pi), 0)^{-1} = \sum_{\mu'} \Gamma_{\mu k, \mu'}(u(\mathbf{e}_3, -\pi))\varphi_{\mu'}[f_k^j]. \tag{40d}$$

With reference to this formula it is easily seen that the formula (52) in BW I is a special case of (40b).

That the matrix $\hat{\Gamma}''$ in (40c) is diagonal (with diagonal elements $+1$ or -1) follows at once from the fact that the matrix Γ'' in (23) is diagonal (with diagonal elements $+1$ or -1).

Our conclusions up to this point in this section are completely independent of the locality conditions (f) in Sec. III. We shall now draw some further conclusions, in which we take the locality conditions into account. Before we state the relevant lemma, we recall that the domain of the *closed* and *normal* operator $V(t)$, t complex, depends only on $\text{Im}(t)$. We write the operator as $(V(t), D_V(\text{Im}(t)))$ when we wish to exhibit the domain explicitly.

Lemma 7: Let $\{R_n \,|\, n=1,\ldots,\infty\}$ be a *fixed* set of bounded, open, nonempty subsets of W_R, constructed as in Lemma 6. Let $\mathcal{Q}$ be the linear span of the identity operator I and all operators (Q, D_1) of the form

$$Q = \varphi_{\mu 1}[f_1]\varphi_{\mu 2}[f_2]\cdots\varphi_{\mu n}[f_n], \tag{41a}$$

where $\{f_k \,|\, k=1,\ldots,n\}$ is any n-tuplet of test functions

such that $f_k \in \mathcal{S}(R^4)$ and $\mathrm{supp}(f_k) \subset R_k$, for $k = 1, \ldots, n$, and where $\sigma = (\mu 1, \mu 2, \ldots, \mu n)$ is any ordered n-tuplet of indices from I_T. Then:

(a) The linear manifold $D_Q = \mathcal{Q}\Omega$ is dense in the Hilbert space $\mathcal{H}$, and $D_{Qc} = \mathrm{span}\{c(K_3)D_Q \mid c(s) \in \mathcal{D}(R^1)\}$ is a core for every operator $(V(t), D_V(\mathrm{Im}(t)))$.

(b) $(Q^*, D_1) \in \mathcal{Q}$ if $(Q, D_1) \in \mathcal{Q}$.

(c) There exists a *unique antiunitary* operator J such that if $(Q, D_1) \in \mathcal{Q}$ and $c(s) \in D(R^1)$, then

$$V(i\pi)c(K_3)Q\Omega = c(K_3)JQ^*\Omega. \tag{41b}$$

The operator J is an involution, i.e.,

$$J^2 = I, \tag{41c}$$

and it satisfies the conditions

$$J\Omega = \Omega, \quad JD_1 = D_1, \quad ZJXJZ^{-1} \in \mathcal{P}(\mathcal{H}) \tag{41d}$$

for all $(X, D_1) \in \mathcal{P}(\mathcal{H})$, and

$$\begin{aligned} &JZJ = Z^{-1}, \quad JU_0J = U_0, \\ &JV(t)J = V(t) \text{ for all real } t, \end{aligned} \tag{41e}$$

$$JD_+ = D_-, \quad J(V(i\pi), D_+)J = (V(-i\pi), D_-), \tag{41f}$$

$$JD_- = D_+, \quad J(V(-i\pi), D_-)J = (V(i\pi), D_+). \tag{41g}$$

(d) The antiunitary operator Θ_0 defined by

$$J = ZU(u(\mathbf{e}_3, \pi), 0)\Theta_0 \tag{41h}$$

is a TCP transformation which satisfies the conditions:

$$\Theta_0^2 = U_0, \quad \Theta_0\Omega = \Omega, \quad \Theta_0 U(g,x)\Theta_0^{-1} = U(g, -x), \tag{42a}$$

$$\Theta_0 D_1 = D_1, \quad \Theta_0 \mathcal{P}(\mathcal{H})\Theta_0^{-1} = \mathcal{P}(\mathcal{H}) \tag{42b}$$

and

$$\Theta_0 \varphi_\mu(x)\Theta_0^{-1} = \rho_\mu \Gamma''_{\mu,\mu}\varphi^\dagger_\mu(-x), \tag{42c}$$

where $\rho_\mu = +1$ if $\varphi_\mu(x)$ is a component of a boson field and $\rho_\mu = -i$ if $\varphi_\mu(x)$ is a component of a fermion field.

Proof: (1) This lemma corresponds to Lemma 11 in BW I. The reasoning in its proof is similar to our reasoning in BW I, but there are some important differences of detail which have to be discussed. We first note that the assertions (a) and (b) are trivial. The remaining assertions might be proved in the stated order, which in particular yields a proof of the TCP theorem. In order to shorten the discussion, we shall, however, base our proof of the assertion (c) on the well-known fact that under our general assumptions about the fields a TCP transformation Θ_0 which satisfies the conditions (42a)–(42c) does exist.[8] The relations (42a)–(42c) will thus be assumed, and we *define* the antiunitary operator J by (41h), where Z is given by (26). It is then trivial to show that J satisfies the relations (41c)–(41g).

(2) The formula (41b) holds trivially if Q is a multiple of I. Suppose now that Q is of the form (41a). We write $X_k = \varphi_{\mu k}[f_k]$ and $Y_k = \varphi_{\mu k}[f_k^i]$ for $k = 1, \ldots, n$, and we then have

$$\begin{aligned} JQ^*\Omega &= JX_n^\dagger \cdots X_2^\dagger X_1^\dagger \Omega \\ &= \hat{\rho}_\sigma^* \hat{\Gamma}''_{\sigma,\sigma} Z U(u(\mathbf{e}_3, \pi), 0) Y_n \cdots Y_2 Y_1 \Omega, \end{aligned} \tag{43a}$$

where $\hat{\rho}_\sigma = \rho_{\mu 1}\rho_{\mu 2}\cdots\rho_{\mu n}$, in view of (41h) and (42c). For any two operators Y_r and Y_s in the set $\{Y_1, Y_2, \ldots, Y_n\}$ the supports of the corresponding test functions f_r^i and f_s^i are spacelike separated, and hence Y_r anticommutes with Y_s if *both* operators are averaged fermion fields, whereas Y_r commutes with Y_s in all other cases. It is easily shown that under these circumstances

$$\hat{\rho}_\sigma^* Z Y_n \cdots Y_2 Y_1 \Omega = Y_1 Y_2 \cdots Y_n \Omega \tag{43b}$$

and hence

$$JQ^*\Omega = \hat{\Gamma}''_{\sigma,\sigma} U(u(\mathbf{e}_3, \pi), 0) Y_1 Y_2 \cdots Y_n \Omega. \tag{43c}$$

From this it follows, in view of (40b) in Lemma 6, that the operator Q satisfies (41b). From this it trivially follows that (41b) holds for *all* $Q \in \mathcal{Q}$.

We are now prepared to state the main theorem of this section. It will be convenient for the subsequent discussion to introduce the following notation. For any subset R of $\mathcal{H}$ we define the algebra $\mathcal{P}(R)^z$ by

$$\mathcal{P}(R)^z = \{(ZXZ^{-1}, D_1) \mid (X, D_1) \in \mathcal{P}(R)\}, \tag{44}$$

where Z is given by (26).

Theorem 1: (a) The algebras $\mathcal{P}(W_R)$ and $\mathcal{P}(W_L)^z$ are $*$-algebras with the antilinear involution $(X, D_1) \to (X^*, D_1)$. They commute on D_1, i.e.,

$$[X, Y]\psi = 0 \tag{45a}$$

for all $\psi \in D_1$ and for all $X \in \mathcal{P}(W_R)$, $Y \in \mathcal{P}(W_L)^z$.

(b) The vacuum vector Ω is cyclic and separating for both $\mathcal{P}(W_R)$ and $\mathcal{P}(W_L)^z$.

(c) With $V(t) = U(v(\mathbf{e}_3, t), 0)$ (a velocity transformation in the 3-direction),

$$\begin{aligned} &V(t)\,\mathcal{P}(W_R)V(t)^{-1} = \mathcal{P}(W_R), \\ &V(t)\,\mathcal{P}(W_L)^z V(t)^{-1} = \mathcal{P}(W_L)^z \end{aligned} \tag{45b}$$

for all real t, and with J defined as in Lemma 7,

$$J\,\mathcal{P}(W_R)J = \mathcal{P}(W_L)^z. \tag{45c}$$

(d) With the domains D_+ and D_- such that the operators $(V(i\pi), D_+)$ and $(V(-i\pi), D_-)$ are self-adjoint,

$$\mathcal{P}(W_R)\Omega \subset D_+, \quad V(i\pi)X\Omega = JX^*\Omega, \tag{45d}$$

for any $X \in \mathcal{P}(W_R)$, and

$$\mathcal{P}(W_L)^z\Omega \subset D_-, \quad V(-i\pi)Y\Omega = JY^*\Omega \tag{45e}$$

for any $Y \in \mathcal{P}(W_L)^z$.

(e) The condition

$$C_R X\Omega = X^*\Omega, \text{ all } X \in \mathcal{P}(W_R) \tag{46a}$$

defines an *antilinear* operator $(C_R, \mathcal{P}(W_R)\Omega)$, and the condition

$$C_L^z Y\Omega = Y^*\Omega, \quad \text{all } Y \in \mathcal{P}(W_L)^z \tag{46b}$$

defines an *antilinear* operator $(C_L^z, \mathcal{P}(W_L)^z\Omega)$.

These two operators satisfy the relations

$$(C_R, \mathcal{P}(W_R)\Omega)^{**} = (C_L^z, \mathcal{P}(W_L)^z\Omega)^* = (JV(i\pi), D_+), \tag{46c}$$

$$(C_L^z, \mathcal{P}(W_L)^z\Omega)^{**} = (C_R, \mathcal{P}(W_R)\Omega)^* = (JV(-i\pi), D_-). \tag{46d}$$

This theorem corresponds to Theorem 1 in BW I. The proof is *identical* with our proof in BW I, provided that we consistently substitute the operator C_L^z for the operator C_L, and the algebra $\mathcal{P}(W_L)^z$ for the algebra $\mathcal{P}(W_L)$.

In the particular case that there is *no* fermion field among the quantum fields we have $U_0 = I$ and $Z = I$, and hence $\rho(W_L)^z = \rho(W_L)$, in which case the present theorem is identical with Theorem 1 in BW I.

The algebra $\rho(W_R)$, respectively the algebra $\rho(W_L)$, can be regarded as consisting of field operators locally associated with the wedge region W_R, respectively the region W_L. We note that the role of these algebras is not quite as symmetric in the present theorem as in BW I, in the sense that the assertions are about the pair $(\rho(W_R), \rho(W_L)^z)$ rather than about the pair $(\rho(W_R), \rho(W_L))$. It is, however, easily seen that there is a completely equivalent formulation in terms of the pair $(\rho(W_L), \rho(W_R)^z)$, and we note, for instance, that

$$\rho(W_L)\Omega \subset D_-, \quad V(-i\pi)Y\Omega = J_L Y^* \Omega, \tag{47a}$$

for any $Y \in \rho(W_L)$, and

$$\rho(W_R)^z \Omega \subset D_+, \quad V(i\pi)X\Omega = J_L X^* \Omega, \tag{47b}$$

for any $X \in \rho(W_R)^z$, where

$$J_L = ZJZ^{-1} = U_0 J = JU_0. \tag{47c}$$

Furthermore,

$$J_L \rho(W_L) J_L = \rho(W_R)^z. \tag{47d}$$

We conclude this section with the remark that *all* the considerations in Sec. V in BW I also apply to the present situation, provided that $\rho(W_L)$ is replaced by $\rho(W_L)^z$ and that $\rho_0(W_L)$ is replaced by $\rho_0(W_L)^z = Z\rho_0(W_L)Z^{-1}$ everywhere in the discussion. In order to have a more suggestive notation it is then convenient to change the notation in BW I according to the scheme: $U(W_L) \to U(W_L)^z$, $\mathcal{A}_L \to \mathcal{A}_L^z$, etc.

V. THE DUALITY CONDITION FOR THE WEDGE REGIONS W_R AND W_L

The discussion in this section corresponds to the discussion in Sec. VI in BW I. We are thus concerned with the question of how the field operators in $\rho(W_R)$ might generate a von Neumann algebra of bounded operators which can be regarded as being locally associated with the region W_R. We must, of course, here define the term "locally associated with" precisely and in a manner appropriate for a field theory in which fermion fields might occur. To set the stage for the discussion, we begin with some algebraic considerations.

Definition 1: If $\mathcal{A}$ is a von Neumann algebra such that $U_0 \mathcal{A} U_0^{-1} = \mathcal{A}$, and if $\mathcal{A}^z = Z\mathcal{A}Z^{-1}$ with Z defined as in (26), then the *quasicommutant* $\mathcal{A}^q$ of $\mathcal{A}$ is defined as the von Neumann algebra $\mathcal{A}^q = (\mathcal{A}^z)'$.

In a theory in which fermion operators, i.e., operators X which satisfy $U_0 X U_0^{-1} = -X$, occur, the notion of quasicommutant[9] is the proper notion in terms of which one may formulate the conditions of locality and of duality. As an algebraic notion the notion of a quasicommutant is less general than the notion of a commutant in the sense that the former notion refers to a specific unitary involution U_0.

We formulate the pertinent facts about the notion of a quasicommutant as follows.

Theorem 2: Let $\mathcal{A}$ be a von Neumann algebra such that $U_0 \mathcal{A} U_0^{-1} = \mathcal{A}$, and let $\mathcal{A}^q = (Z\mathcal{A}Z^{-1})'$ be its quasicommutant. Let

$$\mathcal{A}_B = \{X \mid U_0 X U_0^{-1} = X, \ X \in \mathcal{A}\},$$
$$\mathcal{A}_F = \{X \mid U_0 X U_0^{-1} = -X, \ X \in \mathcal{A}\}, \tag{48a}$$

and

$$(\mathcal{A}^q)_B = \{Y \mid U_0 Y U_0^{-1} = Y, \ Y \in \mathcal{A}^q\},$$
$$(\mathcal{A}^q)_F = \{Y \mid U_0 Y U_0^{-1} = -Y, \ Y \in \mathcal{A}^q\}. \tag{48b}$$

Then:

(a) $U_0 \mathcal{A}^q U_0^{-1} = \mathcal{A}^q$, $\mathcal{A}^q = Z(\mathcal{A}')Z^{-1}$, $(\mathcal{A}^q)^q = \mathcal{A}$ (49)

(b) Every operator $X \in \mathcal{A}$ has the unique representation

$$X = X_b + X_f, \quad \text{with } X_b \in \mathcal{A}_B, \ X_f \in \mathcal{A}_F, \tag{50a}$$

where, in fact,

$$X_b = \tfrac{1}{2}(X + U_0 X U_0^{-1}), \quad X_f = \tfrac{1}{2}(X - U_0 X U_0^{-1}). \tag{50b}$$

Every operator $Y \in \mathcal{A}^q$ has the unique representation

$$Y = Y_b + Y_f, \quad \text{with } Y_b \in (\mathcal{A}^q)_B, \ Y_f \in (\mathcal{A}^q)_F, \tag{50c}$$

where, in fact,

$$Y_b = \tfrac{1}{2}(Y + U_0 Y U_0^{-1}), \quad Y_f = \tfrac{1}{2}(Y - U_0 Y U_0^{-1}). \tag{50d}$$

(c) The elements $X_b \in \mathcal{A}_B$, $X_f \in \mathcal{A}_F$, $Y_b \in (\mathcal{A}^q)_B$, and $Y_f \in (\mathcal{A}^q)_F$ satisfy the conditions

$$[X_b, Y_b] = 0, \tag{51a}$$

$$[X_b, Y_f] = 0, \tag{51b}$$

$$[X_f, Y_b] = 0, \tag{51c}$$

$$\{X_f, Y_f\} = X_f Y_f + Y_f X_f = 0. \tag{51d}$$

The set $(\mathcal{A}^q)_B$ is a von Neumann algebra, precisely equal to the set of all bounded operators Y_b which satisfy the condition $U_0 Y_b U_0^{-1} = Y_b$, and the conditions (51a) and (51c) for all $X_b \in \mathcal{A}_B$, $X_f \in \mathcal{A}_F$. The set $\mathcal{A}_B$ is a von Neumann algebra, precisely equal to the set of all bounded operators X_b which satisfy the condition $U_0 X_b U_0^{-1} = X_b$, and the conditions (51a) and (51b) for all $Y_b \in (\mathcal{A}^q)_B$, $Y_f \in (\mathcal{A}^q)_F$. The set $(\mathcal{A}^q)_F$ is precisely equal to the set of all bounded operators Y_f which satisfy the condition $U_0 Y_f U_0^{-1} = -Y_f$, and the conditions (51b) and (51d) for all $X_b \in \mathcal{A}_B$, $X_f \in \mathcal{A}_F$. The set $\mathcal{A}_F$ is precisely equal to the set of all bounded operators X_f which satisfy the condition $U_0 X_f U_0^{-1} = -X_f$, and the conditions (51c) and (51d) for all $Y_b \in (\mathcal{A}^q)_B$, $Y_f \in (\mathcal{A}^q)_F$.

(d) The vector Ω is cyclic (respectively separating) for $\mathcal{A}$ if and only if it is separating (respectively cyclic) for $\mathcal{A}^q$.

We omit the very trivial proofs of these assertions. We stated the above facts in the form of a formal theorem in view of their importance for our discussion. The situation might be illustrated as follows. Suppose that two von Neumann algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ are "locally associated with" two regions R_1, respectively R_2, which are causally independent. The "local" nature of the association can then be expressed through the relation $\mathcal{A}_1 \subset \mathcal{A}_2^q$, which, in view of the theorem, is equivalent to the customary conditions in terms of commutators and anticommutators, i.e., the fermion operators in $\mathcal{A}_1$

anticommute with the fermion operators in $\mathcal{A}_2$ and commute with the boson operators in $\mathcal{A}_2$, whereas the boson operators in $\mathcal{A}_1$ commute with all operators in $\mathcal{A}_2$. Now $\mathcal{A}_1 \subset \mathcal{A}_2^g$ is equivalent to the condition that $[X, Y] = 0$ for all $X \in \mathcal{A}_1$ and all $Y \in \mathcal{A}_2^g = Z\mathcal{A}_2 Z^{-1}$, which means that the locality conditions are expressible in terms of the vanishing of certain *commutators*, irrespective of whether fermion operators occur or do not occur in the theory. This has the important practical consequence, from our point of view, that we do not have to create a new algebraic theory in order to deal with the case of fermion operators; as in BW I it suffices to consider the relationships between von Neumann algebras and their commutants.[10] Let us also note here that according to the fermion-superselection principle only a boson operator can be a physical observable. This means, with reference to our illustration above, that the observables in $\mathcal{A}_2$ and $\mathcal{A}_2^g$ are precisely the same, and thus that the observables associated with the region R_1 commute with the observables associated with R_2.

Definition 2: (a) A set $\mathcal{K}(W_R)$ of bounded operators such that $X^* \in \mathcal{K}(W_R)$ for all $X \in \mathcal{K}(W_R)$ shall be said to be *covariantly associated with* W_R if and only if

$$U(\lambda)\mathcal{K}(W_R)U(\lambda)^{-1} \subset \mathcal{K}(W_R) \tag{52a}$$

for all elements λ in the semigroup $\sigma(W_R)$ consisting of all $\lambda \in \bar{\mathcal{P}}$ such that $\Lambda(\lambda)W_R \subset W_R$. In particular,

$$V(t)\mathcal{K}(W_R)V(t)^{-1} = \mathcal{K}(W_R), \quad \text{all real } t, \tag{52b}$$

and, more generally,

$$U(\lambda)\mathcal{K}(W_R)U(\lambda)^{-1} = \mathcal{K}(W_R), \quad \text{all } \lambda \in \bar{\mathcal{P}}(W_R), \tag{52c}$$

where $\bar{\mathcal{P}}(W_R)$ is the group of all elements $\lambda \in \bar{\mathcal{P}}$ such that $\Lambda(\lambda)W_R = W_R$, i. e., all Poincaré transformations which map W_R onto W_R.

(b) A set $\mathcal{K}(W_L)$ of bounded operators such that $Y^* \in \mathcal{K}(W_L)$ for all $Y \in \mathcal{K}(W_L)$ shall be said to be *covariantly associated with* W_L if and only if

$$\mathcal{K}(W_L) = U(u(\mathbf{e}_1, \pi), 0)\mathcal{K}(W_R)U(u(\mathbf{e}_1, \pi), 0)^{-1}, \tag{53}$$

where $\mathcal{K}(W_R)$ is a set covariantly associated with W_R.

(c) Let $\mathcal{K}(W_R)$ be a set of bounded operators, co-covariantly associated with W_R as above. The association shall be said to be *TCP-symmetric* if and only if

$$\Theta_0 \mathcal{K}(W_R)\Theta_0^{-1} = \mathcal{K}(W_L) \tag{54a}$$

or, equivalently,

$$J\mathcal{K}(W_R)J^{-1} = \mathcal{K}(W_L)^g, \tag{54b}$$

where $\mathcal{K}(W_L)$ is given by (53).

(d) A set $\mathcal{K}(W_R)$ of bounded operators which contains X^* if it contains X shall be said to be *locally associated with* W_R if and only if $\mathcal{K}(W_R)$ is covariantly associated with W_R *and*

$$\mathcal{K}(W_R) \subset \mathcal{K}(W_L)^q, \tag{55}$$

where $\mathcal{K}(W_L)$ is given by (53) and where the von Neumann algebra $\mathcal{K}(W_L)^q$ is defined as $(\mathcal{K}(W_L)^g)'$.

(e) A von Neumann algebra $\mathcal{A}(W_R)$, locally associated with W_R, shall be said to satisfy the *condition of duality* if and only if

$$\mathcal{A}(W_R) = \mathcal{A}(W_L)^q, \tag{56}$$

where $\mathcal{A}(W_L)$ is defined in terms of $\mathcal{A}(W_R)$ in analogy with (53).

We present these formal definitions for later reference as we will repeatedly encounter sets which satisfy one, or several, of these defining relations. The geometrical significance of these definitions is obvious and need not be discussed here. Concerning the physical interpretation, we note that the conditions in (d) are minimum conditions which a set of "local observables for W_R" would have to satisfy. In a quantum field theory these conditions are not, however, by themselves enough; the *bounded* local operators should also satisfy some condition of locality relative to the local field operators.

Lemma 8: Let $\mathcal{J}$ be a set of *closable* operators, such that $U_0 \mathcal{J} U_0^{-1} = \mathcal{J}$. We define the set $\mathcal{J}^q$ as the set of all *bounded* operators X such that

$$\begin{aligned} &X^g(Y, D(Y))^* \subset (Y, D(Y))^* X^g, \\ &X^g(Y, D(Y))^{**} \subset (Y, D(Y))^{**} Z^g \end{aligned} \tag{57}$$

for all $(Y, D(Y)) \in \mathcal{J}$. Or, *equivalently*, the set $\mathcal{J}^q$ is precisely equal to the set of all bounded operators X such that for all $(Y, D(Y)) \in \mathcal{J}$,

$$\begin{aligned} &X^g(Y, D(Y)) \subset (Y, D(Y))^{**} X^g, \\ &(X^g)^*(Y, D(Y)) \subset (Y, D(Y))^{**} (X^g)^*. \end{aligned} \tag{58}$$

(a) The set $\mathcal{J}^q$ is a von Neumann algebra, and it satisfies the relation $U_0(\mathcal{J}^q)U_0^{-1} = \mathcal{J}^q$.

(b) Let the set $\mathcal{J}^{qq}$ of bounded operators be defined by

$$\mathcal{J}^{qq} = (\mathcal{J}^q)^q. \tag{59}$$

Then $\mathcal{J}^{qq}$ is a von Neumann algebra precisely equal to the von Neumann algebra generated by the operators V and the spectral projections of the operators K for all pairs of operators $\{V, K\}$, where V is the unique partial isometry, and K is the unique nonnegative definite self-adjoint operator, defined through the polar decomposition

$$(Y, D(Y))^{**} = V(K, D(Y^{**})) \tag{60}$$

of the closure of any $(Y, D(Y)) \in \mathcal{J}$.

This lemma is a paraphrase of well-known facts about the commutant in the sense of von Neumann[11] of a set of closed operators. An equivalent definition for $\mathcal{J}^q$ is thus

$$\mathcal{J}^q = (Z\mathcal{J}^{**}Z^{-1})' \tag{61a}$$

with the prime notation of von Neumann, and the set $\mathcal{J}^{qq}$ is then given by

$$\mathcal{J}^{qq} = (\mathcal{J}^{**})'' \tag{61b}$$

where $\mathcal{J}^{**}$ denotes the set of all closures of the operators in $\mathcal{J}$. That the assertion in (b) above about the algebras $\mathcal{J}^{qq}$ [regarded as given by (61b)] holds is well known[12] (and easily proved). That $\mathcal{J}^q$ (and hence $\mathcal{J}^{qq}$) is invariant under conjugation by U_0 follows trivially from the corresponding property of $\mathcal{J}$.

We shall call $\mathcal{J}^q$ the quasicommutant of the set of adjoints and closures of the possibly unbounded opera-

tors in $\mathcal{J}$; this is consistent with our earlier terminology in the case that $\mathcal{J}$ is actually a von Neumann algebra. We shall say that the von Neumann algebra $\mathcal{J}^{qq}$ is *generated* by the set $\mathcal{J}$.

We shall next consider some special sets of bounded operators defined in terms of field operators in $\mathcal{P}(R)$, where R is any subset of $\mathcal{M}$. In this section we are primarily interested in the wedge regions W_R and W_L, but for later reference it will be convenient to consider other regions R as well. We note here that it would be reasonable to restrict the regions R such that they satisfy the condition $R^{cc}=R$, but we shall not do so since we do not here wish to investigate the geometrical implications of this restriction.

Definition 3: Let R be any subset of Minkowski space, and let R^c be its causal complement [as defined in (5)].

(a) The set $\mathcal{L}(R)$ is defined as the set of all finite linear combinations of operators of the form $(\varphi_\mu[f], D_1)$, where $\mu \in I_T$ and where $f \in \mathcal{S}(R^4)$, with $\mathrm{supp}(f) \subset R$.

(b) The set $\mathcal{G}(R)$ is defined as the von Neumann algebra *generated by* $\mathcal{L}(R)$, i.e.,

$$\mathcal{G}(R)=\mathcal{L}(R)^{qq}, \tag{62}$$

where the superscript "qq" denotes the mapping $\mathcal{J}\to\mathcal{J}^{qq}$ defined in Lemma 8.

(c) The von Neumann algebra $\mathcal{C}(R)$ is defined as the *quasicommutant of* $\mathcal{L}(R^c)$, i.e.,

$$\mathcal{C}(R)=\mathcal{L}(R^c)^{q}=\mathcal{G}(R^c)^{q}, \tag{63}$$

where the superscript "q" denotes the mapping $\mathcal{J}\to\mathcal{J}^{q}$ defined in Lemma 8.

(d) The *weak quasicommutant* $\mathcal{C}_w(R)$ of $\mathcal{P}(R^c)$ is defined as the set of all bounded operators X such that

$$\langle Y^*\phi \,|\, X\psi\rangle = \langle X^*\phi \,|\, Y\psi\rangle \tag{64}$$

for all $\phi, \psi \in D_1$ and all $(Y, D_1) \in \mathcal{P}(R^c)^{\varepsilon} = Z\mathcal{P}(R^c)Z^{-1}$.

We introduce the new term "weak quasicommutant" with some reluctance, but it does seem fairly appropriate to describe the nature of the sets $\mathcal{C}_w(R)$. The adjective "weak" is here intended to convey an impression of the "weak" nature of the "commutation relations" (64), as contrasted with the more restrictive conditions (57). It should be noted, however, that the operators in $\mathcal{C}_w(R)$ commute in the weak sense of (64) with *all* the operators in $\mathcal{P}(R^c)^{\varepsilon}$, whereas the operators in $\mathcal{C}(R)$ commute in the strong sense of (57) only with the operators in the subset $\mathcal{L}(R^c)^{\varepsilon}$ of $\mathcal{P}(R^c)^{\varepsilon}$.

We shall next consider some fairly elementary properties of the sets defined above.

Lemma 9: Let R be any subset of Minkowski space, and let the sets $\mathcal{L}(R)$, $\mathcal{C}(R)$, $\mathcal{C}_w(R)$, and $\mathcal{G}(R)$ be defined as in Definition 3. Then:

(a) Each one of these four sets satisfies the condition (65a) of *covariance*, the condition (64b) of *TCP symmetry*, and the condition (65c) of *isotony*, i.e., if $\mathcal{Q}(R)$ is any one of the sets $\mathcal{L}(R)$, $\mathcal{C}(R)$, $\mathcal{C}_w(R)$, or $\mathcal{G}(R)$, then

$$U(\lambda)\mathcal{Q}(R)U(\lambda)^{-1}=\mathcal{Q}(\Lambda(\lambda)R), \quad \text{all } \lambda \in \bar{\mathcal{P}}, \tag{65a}$$

$$\Theta_0\mathcal{Q}(R)\Theta_0^{-1}=\mathcal{Q}(-R), \tag{65b}$$

where $-R$ denotes the set $-R=\{-x \,|\, x\in R\}$, and

$$\mathcal{Q}(R)\supset\mathcal{Q}(R_1), \quad \text{whenever } R\supset R_1 \tag{65c}$$

(b) The set $\mathcal{C}_w(R)$ is a weakly closed linear manifold, closed under the $*$-operation, i.e., it contains X^* if it contains X.

A bounded operator X is in $\mathcal{C}_w(R)$ if and only if

$$X(Y^*, D_1)\subset(Y, D_1)^*X \tag{66}$$

for all $(Y, D_1)\in\mathcal{P}(R^c)^{\varepsilon}$.

(c) A bounded operator X is in $\mathcal{C}_w(R)$ if and only if the condition (64) holds for all $\phi, \psi \in D_1$ and all (Y, D_1) $\in \mathcal{L}(R^c)^{\varepsilon}$, or, equivalently, if and only if the condition (66) holds for all $(Y, D_1)\in\mathcal{L}(R^c)^{\varepsilon}$.

(d)

$$X_1XX_2\in\mathcal{C}_w(R) \tag{67a}$$

for all $X\in\mathcal{C}_w(R)$ and all $X_1, X_2\in\mathcal{C}(R)$. In particular,

$$\mathcal{C}(R)\in\mathcal{C}_w(R). \tag{67b}$$

(e) If R^c has a nonempty interior, then Ω is separating for $\mathcal{C}_w(R)$, i.e., if $X\in\mathcal{C}_w(R)$ and $X\Omega=0$, then $X=0$.

If R has a nonempty interior, then $\mathcal{G}(R)\Omega$ is dense in the Hilbert space $\mathcal{H}$.

(f) If (for a particular subset R) the "linear field operators" in the set $\mathcal{L}(R^c)$ satisfy the condition that D_1 is a core for the adjoints of the operators in the set, i.e., $(Y^\dagger, D_1)^*=(Y, D_1)^{**}$ for all $(Y, D_1)\in\mathcal{L}(R^c)$, then $\mathcal{C}(R)=\mathcal{C}_w(R)$.

Proof: (1) The assertions (a) and (b) are trivial. We note here that the condition (66) [which is a trivial restatement of the condition (64)] is equivalent to the condition that

$$X(Y^*, D_1)^{**}\subset(Y, D_1)^*X \tag{68}$$

for all $(Y, D_1)\in\mathcal{P}(R^c)^{\varepsilon}$.

(2) To prove the assertion (c), we assume that X is a bounded operator which satisfies the condition (64) for all $\phi, \psi\in D_1$ and all $(Y, D_1)\in\mathcal{L}(R^c)^{\varepsilon}$. It follows at once that the condition (64) then also holds for all (Y, D_1) $\in\mathcal{P}_0(R^c)^{\varepsilon}$. For such an X, let $\phi, \psi\in D_1$, and let (Y, D_1) $\in\mathcal{P}(R^c)^{\varepsilon}$. Since we have $ZD_1=D_1$, and since the quantum fields are operator-valued tempered distributions, it follows from the fact that $(\otimes\mathcal{D}(R^4))^n$ is dense in $\mathcal{S}(R^{4n})$ that there exists a sequence $\{(Y_k, D_1)\,|\,(Y_k, D_1)\in\mathcal{P}_0(R^c)^{\varepsilon}, k=1,\ldots,\infty\}$ of operators such that

$$\text{s-}\lim_{k\to\infty}Y_k\psi=Y\psi, \quad \text{s-}\lim_{k\to\infty}Y_k^*\phi=Y^*\phi. \tag{69}$$

It readily follows that the relation (64) holds for the above operator (Y, D_1), and hence $X\in\mathcal{C}_w(R)$ as asserted.

(3) We consider the assertion (d). Let $X\in\mathcal{C}(R)$, X_w $\in\mathcal{C}_w(R)$, and $(Y, D_1)\in\mathcal{L}(R^c)^{\varepsilon}$. We then have, in view of (57) and (68),

$$XX_w(Y^*, D_1)^{**}\subset X(Y, D_1)^*X_w\subset(Y, D_1)^*XX_w, \tag{70}$$

which means that $XX_w\in\mathcal{C}_w(R)$. From this (67a) follows readily, and, since $I\in\mathcal{C}_w(R)$, the relation (67b) follows.

(4) If $X \in \mathcal{C}_w(R)$, then $X\Omega = 0$ implies that

$$\langle Y_1\Omega | XY_2\Omega\rangle = \langle Y_2^* Y_1\Omega | X\Omega\rangle = 0 \tag{71}$$

for all $Y_1, Y_2 \in \mathcal{P}(R^c)^s$. By the Reeh–Schlieder theorem the set $\mathcal{P}(R^c)^s\Omega$ is dense if R^c has a nonempty interior, which implies that in this case $X = 0$ if (71) holds. This proves the first assertion in (e), and in view of (67b) it follows that Ω is a separating vector for the von Neumann algebra $\mathcal{C}(R)$, and hence a cyclic vector for its quasicommutant $\mathcal{G}(R^c)$ whenever the interior of R^c is nonempty. It readily follows, since $\mathcal{G}(R)$ satisfies the condition of isotony (65c), that $\mathcal{G}(R)\Omega$ is dense whenever R has a nonempty interior.

(5) We consider the assertion (f). If $(Y^*, D_1)^* = (Y, D_1)^{**}$ for all $(Y, D_1) \in \mathcal{L}(R^c)$, and if $X \in \mathcal{C}_w(R)$, then the relation (68) implies that $X \in \mathcal{C}(R)$. In view of (67b) this implies that $\mathcal{C}_w(R) = \mathcal{C}(R)$, as asserted. This completes the proof.

We note that it does not follow from the definition of $\mathcal{C}_w(R)$ as a weak quasicommutant of an algebra $\mathcal{P}(R^c)$ of unbounded operators [or equivalently as the "weak commutant" of the operator algebra $\mathcal{P}(R^c)^s$] that $\mathcal{C}_w(R)$ is a von Neumann algebra; the set need not be closed under multiplication. What the actual situation is in quantum field theory we do not know. In the case of free fields the premises in part (f) of the lemma are trivially satisfied, and $\mathcal{C}_w(R)$ is then identical with the von Neumann algebra $\mathcal{C}(R)$. In this connection we refer to the work of Powers on algebras of unbounded operators, their "weak commutants," and related subjects.[13]

Lemma 10: Let R be any subset of Minkowski space, and let the notation be as in Definition 3 and Lemma 9. Let $\mathcal{A}_0(R)$ be defined as the set of all bounded operators X such that XX_w and X_wX are both in $\mathcal{C}_w(R)$ for all $X_w \in \mathcal{C}_w(R)$. Then:

(a) The set $\mathcal{A}_0(R)$ is a von Neumann algebra, and

$$\mathcal{C}(R) \subset \mathcal{A}_0(R) \subset \mathcal{C}_w(R). \tag{72}$$

(b) The mapping $R \to \mathcal{A}_0(R)$ satisfies the condition of covariance (65a) and the condition of TCP symmetry (65b) in Lemma 9. In particular $U_0\mathcal{A}_0(R)U_0^{-1} = \mathcal{A}_0(R)$.

(c) All operators $(Y, D_1) \in \mathcal{P}(R^c)$ have *closable* extensions defined by

$$(Y, D_1) \to (a(Y), D_a) = (Y^{\dagger *}, D_a) = (Y^{\dagger *}, D_a), \tag{73a}$$

where D_a is the domain defined by

$$D_a = \text{span}\{X\phi \,|\, X \in \mathcal{A}_0(R), \phi \in D_1\}. \tag{73b}$$

These extensions satisfy the conditions

$$(Y^*, D_1)^* \supset (a(Y)^*, D_a)^* \supset (a(Y), D_a) \supset (Y, D_1). \tag{73c}$$

(d) Let $\mathcal{P}_a(R^c)$ be the set of all operators $(a(Y), D_a)$ with $(Y, D_1) \in \mathcal{P}(R^c)$. Then, with the notation in Lemma 8,

$$\begin{aligned} &\mathcal{A}_0(R) = \mathcal{P}_a(R^c)^q, \\ &\mathcal{P}_a(R^c)^{qq} = \mathcal{P}_a(R^c)'' = \mathcal{A}_0(R)^q \subset \mathcal{G}(R^c) \end{aligned} \tag{74a}$$

and the closures and adjoints of the operators $(a(Y), D_a)$ in $\mathcal{P}_a(R^c)$ are thus *affiliated to* the von Neumann algebra $\mathcal{A}_0(R)^q$.

The weak quasicommutant of $\mathcal{P}_a(R^c)$ relative to the domain D_a, i.e., the set of all bounded operators X such that

$$\langle X^*\phi \,|\, a(Y)^s\psi\rangle = \langle (a(Y)^s)^*\phi \,|\, X\psi\rangle \tag{74b}$$

for all $\phi, \psi \in D_a$, all $(a(Y), D_a) \in \mathcal{P}_a(R^c)$, is precisely equal to the set $\mathcal{C}_w(R)$.

(e) The mapping $(Y, D_1) \to (a(Y), D_a)$ of the algebra $\mathcal{P}(R^c)$ onto $\mathcal{P}_a(R^c)$ is a *representation*, and it is a $*$-*representation* of the $*$-algebra $\mathcal{P}(R^c)$ in the sense that

$$(a(Y^\dagger), D_a) = (a(Y)^*, D_a). \tag{75a}$$

The representation is *continuous* in the sense that

$$\text{s-}\lim_{k\to\infty} a(Y_k)\psi = 0 \tag{75b}$$

for all $\psi \in D_a$ whenever

$$\text{s-}\lim_{k\to\infty} Y_k\phi = 0 \tag{75c}$$

for all $\phi \in D_1$.

Proof: (1) $\mathcal{A}_0(R)$ is trivially a $*$-algebra since $\mathcal{C}_w(R)$ is closed under the $*$-operation. From the fact that $\mathcal{C}_w(R)$ is weakly closed, it follows that $\mathcal{A}_0(R)$ is also weakly closed, and hence a von Neumann algebra. The relation (72) is trivial in view of (67b). The assertions (b) are obvious.

(2) It follows from (66) that if $X \in \mathcal{C}_w(R)$ and $\phi \in D_1$, then $X\phi \in D(Y^*)$, for any $(Y, D_1) \in \mathcal{P}(R^c)^s$. In view of (72) this implies that D_a, as defined in (73b), is contained in the domain of the adjoint of any operator (Y, D_1) in $\mathcal{P}(R^c)^s$ *or* in $\mathcal{P}(R^c)$, since $ZD_a = D_a$. It follows that the extensions $(a(Y), D_a)$ are well defined by (73a). Furthermore, (73a) also defines an extension of every operator $(Y^s, D_1) \in \mathcal{P}(R^c)^s$, and we have

$$(a(ZYZ^{-1}), D_a) = Z(a(Y), D_a)Z^{-1} \tag{76a}$$

for all $(Y, D_1) \in \mathcal{P}(R^c)$.

(3) Let $X_1, X_2 \in \mathcal{A}_0(R)$, $\phi \in D_1$, and $(Y, D_1) \in \mathcal{P}(R^c)^s$. Then $X_1X_2 \in \mathcal{A}_0(R)$, and since $\mathcal{A}_0(R) \subset \mathcal{C}_w(R)$, we have

$$\begin{aligned} a(Y)X_1X_2\phi &= Y^{\dagger *}X_1X_2\phi = X_1X_2Y\phi \\ &= X_1Y^{\dagger *}X_2\phi = X_1a(Y)X_2\phi, \end{aligned} \tag{76b}$$

which implies that X_1 commutes with $(a(Y), D_a)^{**}$ in the strong sense of (57), and we have thus proved that $\mathcal{A}_0(R) \subset \mathcal{P}_a(R^c)^q$. It, furthermore, readily follows that the relations (73c) hold for all $(Y, D_1) \in \mathcal{P}(R^c)^s$, and hence for all $(Y, D_1) \in \mathcal{P}(R^c)$. The relation (75a) is then trivial.

(4) We next consider the weak quasicommutant $\mathcal{C}_{wa}(R)$ of $\mathcal{P}_a(R^c)$ relative to the domain D_a. It is easily seen from the condition (74b) that a bounded operator X is in $\mathcal{C}_{wa}(R)$ if and only if $X_1XX_2 \in \mathcal{C}_w(R)$ for all $X_1, X_2 \in \mathcal{A}_0(R)$. This implies that $\mathcal{C}_{wa}(R) = \mathcal{C}_w(R)$, as asserted. We obviously have $XX_w, X_wX \in \mathcal{C}_{wa}(R)$ for all $X_w \in \mathcal{C}_{wa}(R)$, $X \in \mathcal{P}_a(R^c)^q$, and in view of the results in step (3) above the first relation (74a) follows. The remaining relations (74a) then follow trivially, in view of (72).

(5) The remaining assertions in part (e) of the lemma are trivial, and we omit the detailed proofs.

We must here state that we know much less about the relationships between the sets $\mathcal{C}(R)$, $\mathcal{C}_w(R)$, and $\mathcal{A}_0(R)$

than we would like to know. We note here that $\mathcal{C}(R)$ was defined as the quasicommutant of the *subset* $\mathcal{L}(R^c)$ of $\mathcal{P}(R^c)$, which means that the closures and adjoints of the operators in $\mathcal{L}(R^c)$ are affiliated to the von Neumann algebra $\mathcal{G}(R^c)=\mathcal{C}(R)^q$, but we see no obvious reason why this would imply that the closures and adjoints of the operators in $\mathcal{P}(R^c)$ are also affiliated to this same von Neumann algebra. The lemma now shows that there exists a "natural" extension $(a(Y), D_a)$ of all the operators in $\mathcal{P}(R^c)$ such that the closures and adjoints of the extended operators are affiliated to $\mathcal{G}(R^c)$, or to the possibly smaller von Neumann algebra $\mathcal{A}_0(R)^q$. It is here important to note that this extension depends on the set R^c, although this is not shown explicitly in our notation. A field operator which can be associated with different regions might thus have different extensions constructed as in the lemma.

In view of our present lack of understanding of the general structure of a quantum field theory the possible physical interpretation of the weak quasicommutant $\mathcal{C}_w(R)$ of $\mathcal{P}(R^c)$ is far from clear. With reference to the discussion by Licht of strict localization[14] we note here the following. Let V be a partial isometry in $\mathcal{C}_w(R)^g$ such that $V^*V=I$, and let $\psi=V\Omega$. Then ψ is in the domain of $(Y, D_1)^*$ for any $(Y, D_1)\in\mathcal{P}(R^c)$ and we have, for any such (Y, D_1),

$$\langle\psi\,|\,Y^{\dagger *}\psi\rangle=\langle\Omega\,|\,Y\Omega\rangle \tag{77a}$$

and, more generally,

$$\langle Y_1^{\dagger *}\psi\,|\,Y_2^{\dagger *}\psi\rangle=\langle Y_1\Omega\,|\,Y_2\Omega\rangle \tag{77b}$$

for any two (Y_1, D_1), $(Y_2, D_1)\in\mathcal{P}(R^c)$. We here assume that both R and R^c have nonempty interiors. It is then not hard to show that if a vector ψ satisfies the conditions (77b), then ψ is of the above form.

The expression at left in (77a) might be loosely regarded as the "expectation value of the field operator Y in the state ψ," and the "local character" of the state then manifests itself in the fact that the expectation value in the state equals the vacuum expectation value, for all operators $(Y, D_1)\in\mathcal{P}(R^c)$. Note, however, that the operator $Y^{\dagger *}$ at left in (77a) cannot in general be replaced by Y^{**} or by Y, as ψ might not be in the domains of these operators. We furthermore note that the condition (77a) also holds for all the bounded operators in the von Neumann algebra $\mathcal{C}_w(R)^q$, but not necessarily for the operators in $\mathcal{G}(R^c)$. In our opinion (77a) is a necessary condition for a local state (localized in the complement of R^c) but by no means a sufficient condition.

We shall next consider the properties of the sets $\mathcal{C}(R)$, $\mathcal{A}_0(R)$, $\mathcal{C}_w(R)$, and $\mathcal{G}(R)$ for the special case that $R\in\mathcal{W}$. The lemma which follows corresponds in part to our Theorem 3 in BW I, with some added refinements which we overlooked before.

Lemma 11: Let $\mathcal{C}(R)$, $\mathcal{C}_w(R)$, $\mathcal{A}_0(R)$, and $\mathcal{G}(R)$ be defined as in Definition 3 and Lemma 10. Then:

(a) $\mathcal{C}(W_R)=\mathcal{C}(\overline{W}_R)$, $\mathcal{C}_w(W_R)=\mathcal{C}_w(\overline{W}_R)$,

$$\mathcal{G}(W_R)=\mathcal{G}(\overline{W}_R),\quad \mathcal{A}_0(W_R)=\mathcal{A}_0(\overline{W}_R) \tag{78a}$$

with analogous identities for the corresponding objects associated with W_L, and

$$\mathcal{C}(W_R)\subset\mathcal{A}_0(W_R)\subset\mathcal{C}_w(W_R)\subset\mathcal{G}(W_R)=\mathcal{C}(W_L)^q. \tag{78b}$$

(b) The von Neumann algebra $\mathcal{C}(W_R)$ is *locally associated with* W_R and the association is *TCP symmetric*, in the sense of Definition 2.

(c) The set $\mathcal{C}_w(W_R)$ and the von Neumann algebra $\mathcal{G}(W_R)$ are *covariantly associated with* W_R, and the association is *TCP symmetric*, in the sense of Definition 2.

(d) For every $X\in\mathcal{C}_w(W_R)$ [and hence for every X in $\mathcal{C}(W_R)$ or $\mathcal{A}_0(W_R)$] we have

$$X\Omega\in D_+,\quad V(i\pi)X\Omega=JX^*\Omega. \tag{79}$$

(e) The von Neumann algebra $\mathcal{A}_0(W_R)$ satisfies the conditions:

$$\mathcal{A}_0(W_L)=\Theta_0\mathcal{A}_0(W_R)\Theta_0^{-1}$$
$$=U(u(\mathbf{e}_1,\pi),0)\mathcal{A}_0(W_R)U(u(\mathbf{e}_1,\pi),0)^{-1} \tag{80a}$$

and

$$U(\lambda)\mathcal{A}_0(W_R)U(\lambda)^{-1}=\mathcal{A}_0(W_R) \tag{80b}$$

for all $\lambda\in\bar{\mathcal{P}}$ such that $\Lambda(\lambda)W_R=W_R$, i.e., for all Poincaré transformations which map W_R *onto* W_R.

(f) $$[X, JX_wJ]\Omega=0 \tag{81}$$

for all $X\in\mathcal{A}_0(W_R)$, $X_w\in C_w(W_R)$.

Proof: (1) We consider the identities (78a). Let $x\in W_R$. Then we have $\mathcal{C}(\overline{W}_R)\supset\mathcal{C}(W_R)\supset T(x)\mathcal{C}(\overline{W}_R)T(x)^{-1}$, in view of the fact that $\mathcal{C}(R)$ satisfies the condition of isotony. Since $\mathcal{C}(R)$ is weakly closed, and since $T(x)$ is a strongly continuous function of x, it follows at once that the first identity in (78a) holds. The next two identities are proved by exactly the same reasoning. The fourth identity follows from the second, and from the definition of $\mathcal{A}_0(R)$ in terms of $\mathcal{C}_w(R)$.

(2) The inclusion relations between the first three sets at left in (78b) correspond to (72) in Lemma 10. The assertions (e) also follow from Lemma 10. [Note that we do not assert that (80b) holds for all Poincaré transformations λ which map W_R *into* W_R.] The assertion (c) is trivial.

(3) The relation $\mathcal{C}_w(W_R)\subset\mathcal{C}(W_L)^q$ is not trivial; it is equivalent to the condition that all operators in $\mathcal{C}_w(W_R)$ commute with all operators in $\mathcal{C}(W_L)^g$. To prove this relation, we first consider the assertion (d) of the lemma. The relations (79) follows readily from the definition of $\mathcal{C}_w(W_R)$, and Lemma 13 in BW I. (In this argument we depend, of course, ultimately on Theorem 1 of the present paper in place of Theorem 1 in BW I.)

(4) Let $X\in\mathcal{A}_0(W_R)$ and let $X_w\in\mathcal{C}_w(W_R)$. Since, by (c) above, $\mathcal{C}_w(W_R)$ is invariant under conjugation by $V(t)$, it follows that $XV(t)X_w^*V(t)^{-1}\in\mathcal{C}_w(W_R)$ for all real t. In view of (d) above it then follows from Lemma 14 in BW I that the relation (81) holds.

(5) Let $X\in\mathcal{C}(W_R)$, and let $X_w\in\mathcal{C}_w(W_R)$. We write $Y=ZJX_wJZ^{-1}$, and we then have $Y\in\mathcal{C}_w(W_L)$. Let $x\in W_R$, and let $X(x)=T(x)XT(x)^{-1}$. Then $X(x)\in\mathcal{C}(W_R)$, and (81) holds with X replaced by $X(x)$. We consider the special cases when each one of the operators X and Y is either a boson operator (i.e., a bounded operator which

commutes with U_0), or else a fermion operator (i. e., a bounded operator which anticommutes with U_0). The relation (81) then implies that

$$(X(x)Y + sYX(x))\Omega = 0, \tag{82}$$

where $s = +1$ if both X and Y are fermion operators, and $s = -1$ if at least one of the operators X and Y is a boson operator.

We note that the operator $Q(x) = X(x)Y + sYX(x)$ is included in the set $\mathcal{C}_w(R)$, where $R = W_L \cup \Lambda(I,x)W_R$; this follows from Lemma 9 since $X(x) \in \mathcal{C}(\Lambda(I,x)W_R) \subset \mathcal{C}(R)$ and $Y \in \mathcal{C}_w(W_L) \subset \mathcal{C}_w(R)$. Since the interior of R^c is nonempty, it follows from Lemma 9 that $Q(x) = 0$. Since $Q(x)$ is a strongly continuous function of x, we conclude that $(XY + sYX) = Q(0) = 0$. This in turn implies that $[X, JX_wJ] = 0$. From the fact that this relation holds in the special cases considered it readily follows that it holds for *all* $X \in \mathcal{C}(W_R)$, $X_w \in \mathcal{C}_w(W_R)$. This means that $\mathcal{C}_w(W_R) \subset \mathcal{C}(W_L)^q = \mathcal{G}(W_R)$, as asserted in (78b). This completes the proof of the lemma.

The relations (78a) should be carefully noted. The algebraic objects appearing in these relations are thus the *same* for the *closed* wedge W_R as for the *open* wedge W_R, which fact leads to a considerable simplification of the subsequent discussion. We employ a notation in the following according to which the objects are labeled by the open wedges W_R and W_L.

The facts stated in part (b) of the lemma correspond, in a sense, to a well-known result of Borchers concerning the local nature of quantum fields which are local relative to an irreducible set of local fields.[15]

Theorem 3: Let the notation be as in Definition 3 and Lemmas 10 and 11.

(A) *If the quantum fields are such* that $\mathcal{A}_0(W_R)\Omega$ is dense in the Hilbert space $\mathcal{H}$, then $\mathcal{A}_0(W_R)$ is *locally associated with* W_R, and the association is *TCP symmetric*, in the sense of Definition 2. Furthermore, $\mathcal{A}_0(W_R)$ satisfies the condition of *duality*, and

$$\mathcal{C}(W_R) \subset \mathcal{A}_0(W_R) = \mathcal{C}_w(W_R) = \mathcal{A}_0(W_L)^q \subset \mathcal{G}(W_R). \tag{83}$$

(B) *If the quantum fields are such* that there exists a von Neumann algebra $\mathcal{A}(W_R) \subset \mathcal{C}_w(W_R)$ such that $\mathcal{A}(W_R)\Omega$ is dense, and such that $\mathcal{A}(W_R)$ is either *locally associated with* W_R, or else *covariantly and TCP symmetrically associated with* W_R, in the sense of Definition 2, then:

(a) The algebra $\mathcal{A}(W_R)$ is *locally*, and *TCP symmetrically*, associated with W_R. Furthermore, $\mathcal{A}(W_R)$ satisfies the condition of *duality*, and

$$\mathcal{A}_0(W_R) \subset \mathcal{A}(W_R) = \mathcal{A}(W_L)^q \subset \mathcal{C}_w(W_R), \tag{84a}$$

where

$$\mathcal{A}(W_L) = U(u(\mathbf{e}_1, \pi), 0)\mathcal{A}(W_R)U(u(\mathbf{e}_1, \pi), 0)^{-1} \tag{84b}$$

as in Definition 2. The relation $\mathcal{A}_0(W_R) = \mathcal{A}(W_R)$ holds if and only if $\mathcal{A}_0(W_R)\Omega$ is dense.

(b) The algebra $\mathcal{A}(W_R)$ is a *factor*, with Ω as a cyclic and separating vector. For any $X \in \mathcal{A}(W_R)$,

$$X\Omega \in D_+, \quad V(i\pi)X\Omega = JX^*\Omega, \tag{85a}$$

and

$$J\mathcal{A}(W_R)J = \mathcal{A}(W_R)'. \tag{85b}$$

(c) There exists an extension of the operators in $\mathcal{P}(W_R)$ defined by

$$(X, D_1) \to (e_R(X), D_{1R}) = (X^{\dagger *}, D_{1R}), \tag{86a}$$

where

$$D_{1R} = \text{span}\{Y\phi \mid Y \in \mathcal{A}(W_L),\ \phi \in D_1\} \tag{86b}$$

such that the extension satisfies the conditions

$$(X^\dagger, D_1)^* \supset (e_R(X)^*, D_{1R})^* \supset (e_R(X), D_{1R}) \supset (X, D_1). \tag{86c}$$

The mapping $(X, D_1) \to (e_R(X), D_{1R})$ of $\mathcal{P}(W_R)$ onto the set $\mathcal{P}_e(W_R)$ of the extended operators is a continuous $*$-representation in the sense described in Lemma 10.

The closures and adjoints of all operators $(e_R(X), D_{1R}) \in \mathcal{P}_e(W_R)$ are *affiliated to* the von Neumann algebra $\mathcal{A}(W_R)$.

(d) The *weak quasicommutant* $\mathcal{C}_{we}(W_L)$ of $\mathcal{P}_e(W_R)$ relative to the domain D_{1R}, i. e., the set of all bounded operators Y such that for all $(X, D_1) \in \mathcal{P}(W_R)$,

$$Y^g(e_R(X)^*, D_{1R}) \subset (e_R(X), D_{1R})^* Y^g \tag{87}$$

is precisely equal to the quasicommutant $\mathcal{A}(W_L)$ of $\mathcal{P}_e(W_R)$.

Proof: (1) Let $\mathcal{A}(W_R)$ be a von Neumann algebra such that $\mathcal{A}(W_R) \subset \mathcal{C}_w(W_R)$ and $V(t)\mathcal{A}(W_R)V(t)^{-1} = \mathcal{A}(W_R)$ for all real t. The algebra $\mathcal{A}_0(W_R)$, in particular, satisfies these conditions, in view of Lemma 11. If now $\mathcal{A}(W_R)\Omega$ is dense, then it follows from Theorem 2 in BW I that (85a) and (85b) hold. It furthermore follows from Lemma 15 in BW I that $\mathcal{A}(W_R)$ is a factor. We have thus proved the assertions (Bb).

(2) We consider the relation (81) in Lemma 11, with $X_w = X_1X_2$, where X_1 and X_2 are elements of a von Neumann algebra $\mathcal{A}(W_R)$ which satisfies the premises in step (1) above, and where $X \in \mathcal{A}_0(W_R)$. By repeated application of (81) it readily follows that $[X, JX_1J]JX_2\Omega = 0$, and, if $\mathcal{A}(W_R)\Omega$ is dense, it follows that $[X, JX_1J] = 0$ for all $X \in \mathcal{A}_0(W_R)$, $X_1 \in \mathcal{A}(W_R)$. In view of (85b) this implies that $\mathcal{A}_0(W_R) \subset \mathcal{A}(W_R)$, as asserted in (84a).

(3) We consider again the relation (81), with $X = X_3X_4$, where $X_3, X_4 \in \mathcal{A}_0(W_R)$ and $X_w \in \mathcal{C}_w(W_R)$. By repeated application of (81) we easily show that

$$[X_3, JX_wJ]X_4\Omega = 0. \tag{88}$$

In the particular case that $\mathcal{A}_0(W_R)\Omega$ is dense the relation (88) implies that $\mathcal{C}_w(W_R) \subset (J\mathcal{A}_0(W_R)J)' = \mathcal{A}_0(W_R)$, where the equality between the last two members follows from step (1) above. In view of (78b) in Lemma 11 it then follows that the relations (83) hold. We have thus shown that the premises in (A) imply the relations (83). Since $\mathcal{C}_w(W_R)$ is covariantly associated with W_R, we then conclude that $\mathcal{A}_0(W_R)$ is locally associated with W_R. We have thus proved the assertions (A).

(4) We consider a von Neumann algebra $\mathcal{A}(W_R)$ which satisfies the premises in part (B). If $\mathcal{A}(W_R)$ is *locally* associated with W_R, then $\mathcal{A}(W_L) \subset \mathcal{A}(W_R)^q = (\mathcal{A}(W_R)')^g = (J\mathcal{A}(W_R)J)^g$ in view of (85b), and this means that the association of $\mathcal{A}(W_R)$ with W_R is *TCP* symmetric. Conversely, if $\mathcal{A}(W_R)$ is *TCP symmetrically* associated with

W_R, then (85b) implies at once that $\mathcal{A}(W_R)=\mathcal{A}(W_L)^a$, and in particular the association is local. It readily follows from the results in steps (2) and (3) above that $\mathcal{A}_0(W_R)=\mathcal{A}(W_R)$ if and only if $\mathcal{A}_0(W_R)\Omega$ is dense. We have thus proved the assertions (Ba).

(5) The assertions (Bc) are proved in the same manner as the corresponding assertions about the extension $(Y,D_1)\rightarrow(a(Y),D_a)$ in Lemma 10, and we need not repeat the arguments.

(6) We finally consider the assertion (d). It readily follows from (87) that a bounded operator Y_w is in $\mathcal{C}_{we}(W_L)$ if and only if $Y_1Y_wY_2\in\mathcal{C}_w(W_L)$ for all $Y_1, Y_2\in\mathcal{A}(W_L)$. We can restate this as follows. The operator X_w is in $(J\mathcal{C}_{we}(W_L)J)^s$ if and only if $X_1X_wX_2\in\mathcal{C}_w(W_R)$ for all $X_1, X_2\in\mathcal{A}(W_R)$.

An operator X_w which satisfies the above condition is thus included in $\mathcal{C}_w(W_R)$. By the same reasoning as in the proof of (81) in Lemma 11 we show that $[X,JX_wJ]\Omega=0$ for all $X\in\mathcal{A}(W_R)$, $X_w\in(J\mathcal{C}_{we}(W_L)J)^s$. By the same reasoning as in step (3) in the present proof we conclude that $[X,JX_wJ]=0$, which means that $\mathcal{C}_{we}(W_L)^s\subset\mathcal{A}(W_R)'=\mathcal{A}(W_L)^s$. Since the set $\mathcal{A}(W_L)$ is trivially included in $\mathcal{C}_{we}(W_L)$ it follows that the two sets are equal, as asserted.

This completes the proof of the theorem. We postpone the discussion of this result until after the next theorem.

Theorem 4: Let the notation be as in Theorem 3 (i.e., as in Definition 3 and Lemma 10).

(a) The following six conditions are equivalent:

(1) $\mathcal{G}(W_R)\subset\mathcal{G}(W_L)^a$. (89a)

(2) $\mathcal{C}(W_R)=\mathcal{C}(W_L)^a$. (89b)

(3) $\mathcal{G}(W_R)\subset\mathcal{C}_w(W_R)$. (89c)

(4) Ω is a cyclic vector for $\mathcal{C}(W_R)$.

(5) Ω is a separating vector for $\mathcal{G}(W_R)$.

(6) $\mathcal{G}(W_R)\Omega\subset D_+$, and $V(i\pi)X\Omega=JX^*\Omega$ (89d)

for all $X\in\mathcal{G}(W_R)$.

(b) If these conditions are satisfied, then

$$\mathcal{A}_0(W_R)=\mathcal{C}(W_R)=\mathcal{C}_w(W_R)=\mathcal{G}(W_R). \tag{90}$$

The von Neumann algebra $\mathcal{A}_0(W_R)$ satisfies the premises of part (A) of Theorem 3, and all the conclusions of that theorem apply. In particular $\mathcal{A}_0(W_R)$ is a factor with Ω as a cyclic and separating vector. It is locally and *TCP* symmetrically associated with W_R, and it satisfies the condition of duality.

Proof: (1) We first note that since $\mathcal{G}(W_R)\Omega$ is dense by part (e) of Lemma 9, the relations (90) imply that $\mathcal{A}_0(W_R)$ satisfies the premises of part (A) of Theorem 3, and it then follows trivially from that theorem that the six conditions in part (a) of the present theorem are satisfied.

(2) Since $\mathcal{G}(W_L)^a=\mathcal{C}(W_R)$, the condition (89a), in view of (78b) in Lemma 11, at once implies the condition (90). Similarly (89b) implies (90). The condition (89c) implies, in view of (78b), that $\mathcal{C}_w(W_R)=\mathcal{G}(W_R)$, and hence $\mathcal{C}_w(W_R)$ is a von Neumann algebra, which, by the definition of $\mathcal{A}_0(W_R)$ must be equal to $\mathcal{A}_0(W_R)$. Since this von Neumann algebra now has Ω as a cyclic vector, it readily follows from Theorem 3 that all the conditions (90) hold.

(3) The conditions (4) and (5) in part (a) of the theorem are obviously equivalent. If condition (4) holds, then $\mathcal{A}(W_R)=\mathcal{C}(W_R)$ satisfies the premises of part (B) of Theorem 3, and it follows trivially that the conditions (90) are satisfied.

(4) If condition (6) is satisfied, it follows from Theorem 2 in BW I that $J\mathcal{G}(W_R)J=\mathcal{G}(W_R)'$, which implies (89b), and hence (90). This completes the proof.

As the symbolism in Theorems 3 and 4, and in the preceding lemmas, might appear bewildering, we shall now discuss the situation in plain English. Part (b) of Theorem 4 describes what we regard as highly desirable properties of a quantum field theory, and these properties are thus implied by either one of the six equivalent conditions in part (a). We consider the first of these, namely the relation (89a). The von Neumann algebra $\mathcal{G}(W_R)$ is "generated" by the quantum fields $(\varphi_\mu[f],D_1)$ with the support of f in the right wedge W_R, and $\mathcal{G}(W_L)$ is defined analogously. The condition (89a) is simply the condition that these algebras are *local*, i.e., one is contained in the quasicommutant of the other. These algebras are always sufficiently "large" in the sense that each one of them has the vacuum vector as a cyclic vector, and according to (78) in Lemma 11 it is always the case that the quasicommutant of either one is contained in the other. We do not know, however, whether (89a) holds generally; in a particular field theory it could be the case that these algebras are "too large" in the sense that they fail to be locally associated with the wedges. The theorem now shows that the condition that the algebra $\mathcal{G}(W_R)$ not be too large in the above sense is precisely the condition that Ω is a separating vector for $\mathcal{G}(W_R)$, i.e., the condition that $\mathcal{G}(W_R)$ does not contain any nonzero operators which annihilate the vacuum vector.

The algebra $\mathcal{C}(W_R)$ is defined as a "strong" quasicommutant of the field operators $(\varphi_\mu[f],D_1)$, with $\mathrm{supp}(f)\subset W_L$, i.e., $\mathcal{C}(W_R)$ is precisely equal to the set of all bounded operators which commute with the closures of the operators $(\varphi_\mu[f],D_1)^s$, $\mathrm{supp}(f)\subset W_L$, in the strong sense of von Neumann. The algebra $\mathcal{C}(W_R)$ is then trivially equal to the quasicommutant of $\mathcal{G}(W_L)$. According to Lemma 11 the algebra $\mathcal{C}(W_R)$ is always *locally* associated with W_R, and the association is furthermore *TCP* symmetric. These circumstances correspond to a well-known result of Borchers which we referred to earlier.[15] The algebra $\mathcal{C}(W_R)$ is a reasonable choice for "the algebra of all bounded operators locally associated with W_R" unless it so happens that this algebra is "too small" in the sense that it fails to satisfy the duality condition. By the theorem the algebra is too small in the above sense if and only if it does not have the vacuum vector as a cyclic vector, i.e., if and only if $\mathcal{C}(W_R)\Omega$ is a proper subspace of the Hilbert space $\mathcal{H}$.

We have already discussed (following Lemma 10) the possible physical interpretation of the set $C_w(W_R)$, defined (in Definition 3) as the "weak quasicommutant" of all the operators in $\mathcal{P}(W_L)$. Now it is interesting to note that, by Lemma 11, the wedge region W_R has the special property that $C_w(W_R)$ is *included* in $\mathcal{G}(W_R)$. This result, which we derived on the basis of Theorem 1, is not a triviality in our opinion. We also know that an analogous inclusion relation does not hold for arbitrary open regions R. It is, furthermore, interesting to note that, by Theorem 4, the seemingly weak condition $\mathcal{G}(W_R) \subset C_w(W_R)$, i.e., the condition that the operators in $\mathcal{G}(W_R)$ commute at least in the weak sense of (64) with the operators $(\varphi_\mu[f], D_1)^-$ for which $\mathrm{supp}(f) \subset W_L$, in fact, implies that $C(W_R) = C_w(W_R) = \mathcal{G}(W_R)$, i.e., that $C_w(W_R)$ is a von Neumann algebra, *identical* with $\mathcal{G}(W_R)$, and that $\mathcal{G}(W_R)$ is locally associated with W_R and satisfies the condition of duality. This result is also ultimately based on Theorem 1, and it does not seem to follow from some more trivial considerations.

We do not know at this time whether $C_w(W_R)$ is always a von Neumann algebra, i.e., closed under multiplication, without further conditions on the quantum fields. The set $C_w(W_R)$ is trivially equal to the von Neumann algebra $C(W_R)$ if $(X^\dagger, D_1)^* = (X, D_1)^{**}$ for all $(X, D_1) \in \mathcal{L}(W_L)$. One might thus say that the relation $C_w(W_R) \neq C(W_R)$ (if there are quantum field theories for which this is the case) in some sense reflects the inadequacy of the domain D_1 for the definition of the field operators. Let us here note that with our present understanding of the situation the equality $C_w(W_R) = C(W_R)$ does not by itself seem to imply the duality condition. In particular we have not shown that it might not happen that $C_w(W_R)$ consists of multiples of the identity only.

The sixth condition in part (a) of Theorem 4 is of a "technical" nature, without any immediate physical interpretation. We stated this condition because its form suggests a possible direct connection with Theorem 1. We note, for instance, that, in the very special case that the vacuum vector is an analytic vector for the field operators $(\varphi_\mu[f], D_1)$ (as is the case for a free field), then the sixth condition follows trivially from the facts in Theorem 1. We are *not*, however, here conjecturing that the sixth condition follows *in general* from Theorem 1 alone.

Even if the premises of Theorem 4 are not satisfied, it is conceivable, according to Theorem 3, that the quantum fields nevertheless have extensions which are affiliated to von Neumann algebras which satisfy a duality condition, at least for the wedge regions in $\mathcal{W}$. It is easily seen that if $(X, D_1) \to (e_R(X), D_{1R})$ is an extension of a set of field operators which satisfies the condition (86c), then the weak quasicommutant (relative to D_{1R}) of the set of extended operators is necessarily contained in the weak quasicommutant of the original set. The premises in part (B) of Theorem 3 thus seem to us to express *minimal* conditions which a "local" algebra "generated" by the fields must satisfy.

In Sec. VI of BW I we considered four particular conditions on the quantum field, called Conditions I–IV, which were shown to imply the duality condition for the wedge regions. We shall not state the generalizations of these conditions here, but we assert that our earlier Conditions I, II, and IV trivially imply the premises of Theorem 4, and that our Condition III implies the premises of part (B) of Theorem 3.

VI. THE DUALITY CONDITION FOR VON NEUMANN ALGEBRAS ASSOCIATED WITH DOUBLE CONES AND THEIR CAUSAL COMPLEMENTS

In this section we shall generalize the discussion in Sec. VII of BW I. We shall thus consider the construction of von Neumann algebras locally associated with a particular family of regions, namely double cones and their causal complements, in terms of a von Neumann algebra $\mathcal{A}(W_R)$ locally associated with W_R. The scheme is the same as in BW I.

Definition 4: Let the von Neumann algebra $\mathcal{A}(W_R)$ be locally associated with W_R, in the sense of Definition 2.

(a) For any $W \in \mathcal{W}$, i.e., for any wedge region W bounded by two nonparallel characteristic planes, we define a von Neumann algebra $\mathcal{A}(W)$ by

$$\mathcal{A}(\Lambda(\lambda) W_R) = U(\lambda) \mathcal{A}(W_R) U(\lambda)^{-1}, \quad \text{any } \lambda \in \bar{\mathcal{J}}. \tag{91}$$

(b) For any two points x_1 and x_2 in Minkowski space such that $x_2 \in V_+(x_1)$ [where $V_+(x_1)$ is the forward light cone with x_1 as apex], we define the double cone $C = C(x_1, x_2)$ by

$$C(x_1, x_2) = V_+(x_1) \cap V_-(x_2), \tag{92}$$

where $V_-(x_2)$ is the backward light cone with x_2 as apex. The double cones so defined are thus open and nonempty. We denote by $\mathcal{D}_c$ the set of all double cones.

For any double cone C we define a von Neumann algebra $\mathcal{B}(\bar{C})$ by

$$\mathcal{B}(\bar{C}) = \cap \{\mathcal{A}(W) \mid W \in \mathcal{W}, \ W \supset \bar{C}\}. \tag{93}$$

(c) For any $C \in \mathcal{D}_c$ we define the von Neumann algebra $\mathcal{A}(\bar{C}^c)$ by

$$\mathcal{A}(\bar{C}^c) = \{\mathcal{A}(W) \mid W \in \mathcal{W}, \ W \subset \bar{C}^c\}''. \tag{94}$$

(d) A set of von Neumann algebras, defined as above, shall be called a *local AB-system*.

It is easily seen that the definition in part (a) above is consistent, i.e., that the algebras defined by the right-hand side of (91) for two different λ', λ'', are equal whenever $\Lambda(\lambda') W_R = \Lambda(\lambda'') W_R$. We remark here that, as in BW I, we prefer to regard $\mathcal{B}(\bar{C})$ as associated with the closed set $\bar{C}$, and hence the above notation.

We shall next state a theorem corresponding to Theorem 5 and part of Theorem 6 in BW I.

Theorem 5: Given a local AB-system, defined as in Definition 4 in terms of a von Neumann algebra $\mathcal{A}(W_R)$ locally associated with W_R, then:

(a) The algebras in the AB-system satisfy the conditions of *covariance* and *isotony*, i.e., if $\mathcal{Q}(R)$ denotes $\mathcal{A}(R)$ or $\mathcal{B}(R)$, with the appropriate restriction on R, then the conditions (65a) and (65c) hold. Furthermore,

$$\mathcal{B}(\bar{C}_1) \subset \mathcal{A}(W) \subset \mathcal{A}(\bar{C}_2^s) \tag{95}$$

for all $W \in \mathcal{W}$, $C_1, C_2 \in \mathcal{D}_c$, such that $C_1 \subset W \subset \bar{C}_2^s$.

(b) The algebras $\mathcal{B}(\bar{C})$ are *local*, in the sense that

$$\mathcal{B}(\bar{C}_1) \subset \mathcal{B}(\bar{C}_2)^q \tag{96a}$$

for any $C_1, C_2 \in \mathcal{D}_c$, such that $C_1 \subset \bar{C}_2^c$. Furthermore,

$$\mathcal{B}(\bar{C})^q \supset \mathcal{A}(\bar{C}^c) \tag{96b}$$

for any $C \in \mathcal{D}_c$.

(c) The mapping $W \to \mathcal{A}(W)$ is *continuous from the outside* in the sense that

$$\mathcal{A}(W) = \cap \{\mathcal{A}(W_0) \mid W_0 \in \mathcal{W},\ W_0 \supset \bar{W}\} \tag{97a}$$

and it is *continuous from the inside* in the sense that

$$\mathcal{A}(W) = \{\mathcal{A}(W_i) \mid\ W_i\ \in \mathcal{W},\ \bar{W}_i \subset W\}''. \tag{97b}$$

The mapping $\bar{C} \to \mathcal{B}(\bar{C})$ is continuous from the outside in the sense that

$$\mathcal{B}(\bar{C}) = \cap \{\mathcal{B}(\bar{C}_0) \mid C_0 \in \mathcal{D}_c,\ \bar{C} \subset C_0\}. \tag{97c}$$

The mapping $\bar{C}^c \to \mathcal{A}(\bar{C}^c)$ is continuous from the inside in the sense that

$$\mathcal{A}(\bar{C}^c) = \{\mathcal{A}(\bar{C}_i^c) \mid C_i \in \mathcal{D}_c,\ C_i \supset \bar{C}\}''. \tag{97d}$$

(d) If the algebra $\mathcal{A}(W_R)$ satisfies, in addition, the condition of TCP symmetry, as stated in Definition 2, then the AB-system is TCP symmetric in the sense that

$$\begin{aligned} &\Theta_0 \mathcal{A}(W) \Theta_0^{-1} = \mathcal{A}(-W),\quad \Theta_0 \mathcal{B}(\bar{C}) \Theta_0^{-1} = \mathcal{B}(-\bar{C}), \\ &\Theta_0 \mathcal{A}(\bar{C}^c) \Theta_0^{-1} = \mathcal{A}(-\bar{C}^c) \end{aligned} \tag{98}$$

for all $W \in \mathcal{W}$, $C \in \mathcal{D}_c$, and where $-R = \{x \mid -x \in R\}$ for any subset R of Minkowski space.

(e) If the algebra $\mathcal{A}(W_R)$ satisfies, in addition, the condition of duality, as stated in Definition 2, then the algebras $\mathcal{B}(C)$ satisfy a condition of duality in the sense that

$$\mathcal{B}(\bar{C})^q = \mathcal{A}(\bar{C}^c) \tag{99}$$

for any $C \in \mathcal{D}_c$.

The assertions (a)—(d) in the theorem correspond to Theorem 5 in BW I, and the assertion (e) to the assertion (a) in Theorem 6 in BW I. The above assertions are proved by a very trivial modification of the reasoning whereby we proved the corresponding assertions in BW I, and we do not feel that it is necessary to repeat the arguments here. The modifications, of course, have to do with the circumstance that the locality conditions in the present theorem refer to the notion of a quasi-commutant, rather than to the notion of a commutant as in BW I.

The above theorem is of interest because it shows how a "wedge algebra" $\mathcal{A}(W_R)$ with physically desirable properties gives rise to a *system* of algebras (associated with other regions) with physically desirable properties, such as covariance, isotony, TCP symmetry, and duality. In our study of a general quantum field theory the crux of the matter is thus to establish the existence of an algebra $\mathcal{A}(W_R)$ which is locally associated with W_R and which satisfies the conditions of TCP symmetry and duality.

Now it should be noted that nothing said so far guarantees that $\mathcal{B}(\bar{C})$, for some particular $C \in \mathcal{D}_c$, contains other elements than multiples of the identity. In a physically satisfactory "local" theory it must clearly be the case that at least *some* of the algebras $\mathcal{B}(C)$ are nontrivial. In a quantum field theory one might in fact demand that *all* the algebras $\mathcal{B}(\bar{C})$ are nontrivial, and furthermore one might demand that the algebras $\mathcal{B}(\bar{C})$ associated with all $\bar{C} \subset \bar{C}_0^c$, for some C_0, should generate the algebra $\mathcal{A}(\bar{C}_0^c)$. We shall show that this is in fact the case if the quantum fields satisfy the conditions in part (a) of Theorem 4. We do not have a corresponding result for fields which merely satisfy the premises of Theorem 3. The situation in the latter case is complicated by the fact that the extensions of the field operators described in Theorem 3 depend on the region with which the operators are associated, and to clarify the situation it would be necessary to investigate the relationship between the domains of the extensions for different regions. This we have not done, and we shall therefore restrict our considerations to the case when the premises of Theorem 4 are satisfied. We note, however, that we do obtain a satisfactory local theory if the fields satisfy the premises of Theorem 3, and some additional condition which guarantees that $\mathcal{B}(\bar{C})\Omega$ is dense. We refer here to the assertions (b) and (d) in Theorem 6 in BW I, which can readily be generalized to the present situation. It is of interest to state the generalization of the first one of these assertions as follows.

Theorem 6: Let the von Neumann algebra $\mathcal{A}(W_R)$ satisfy the premises of Theorem 5, and let a local AB-system be defined in terms of $\mathcal{A}(W_R)$ as in Definition 4. Let $\mathcal{A}(W_R)$ satisfy the condition of duality, as well as the additional condition that

$$X\Omega \in D_+,\quad V(i\pi)X\Omega = JX^*\Omega \tag{100}$$

for all $X \in \mathcal{A}(W_R)$.

If there exists a double cone C_0 such that $\mathcal{B}(\bar{C}_0)\Omega$ is dense in the Hilbert space $\mathcal{H}$, then

$$\mathcal{A}(\bar{C}_1^c) = \{\mathcal{B}(\bar{C}) \mid C \in \mathcal{D}_c, \bar{C} \subset \bar{C}_1^c\}'' \tag{101a}$$

for every $C_1 \in \mathcal{D}_c$, and

$$\mathcal{A}(W) = \{\mathcal{B}(\Lambda \bar{C}_0) \mid \Lambda \in \bar{L}_0,\ \Lambda \bar{C}_0 \subset W\}'', \tag{101b}$$

$$\mathcal{A}(\bar{C}_1^c) = \{\mathcal{B}(\Lambda \bar{C}_0) \mid \Lambda \in \bar{L}_0,\ \Lambda \bar{C}_0 \subset \bar{C}_1^c\}'' \tag{101c}$$

for every $C_1 \in \mathcal{D}_c$, $W \in \mathcal{W}$. If, furthermore, $\bar{C}_0 \subset W_R$, then

$$\mathcal{A}(W_R) = \{V(t)\mathcal{B}(\bar{C}_0)V(t)^{-1} \mid t \in R^1\}''. \tag{101d}$$

These assertions are proved by the same reasoning as in our proof of the corresponding assertions in Theorem 6 in BW I, and we shall not repeat the arguments. We note here that the premises of the theorem at once imply that Ω is a cyclic and separating vector for $\mathcal{A}(W_R)$, as well as for $\mathcal{B}(\bar{C}_0)$. We furthermore note that the condition (100) is not required for the conclusion in part (e) of Theorem 5. It is, however, essential for the present theorem, and in particular for the conclusion (101d). We refer here to our discussion in Sec. V of BW I of the connection between our considerations and the Tomita—Takesaki theory of modular Hilbert algebras.[16] The relation (101d) can thus be understood with reference to the fact that because of (100) the group $\{V(t) \mid t \in R^1\}$ is precisely the modular automorphism group for $\mathcal{A}(W_R)$.

In preparation for Theorem 7 we prove a lemma about the nature of the weak quasicommutant $\mathcal{C}_w(R)$ in the special case that R is the closure of a double cone in $\mathcal{D}_c$.

Lemma 12: Let $C \in \mathcal{D}_c$. Then

$$\mathcal{C}_w(\bar{C}) = \cap \{\mathcal{C}_w(W) \mid W \in \mathcal{W},\ W \supset \bar{C}\}. \tag{102}$$

Proof: (1) Let $\mathcal{C}_t$ denote the set defined by the right side of (102). It is at once obvious that $\mathcal{C}_w(\bar{C}) \subset \mathcal{C}_t$, and we thus have to prove that if $X \in \mathcal{C}_t$, then $X \in \mathcal{C}_w(\bar{C})$.

(2) Let $\mu \in I_T$, and let $f(x) \in \mathcal{D}(R^4)$ such that $\mathrm{supp}(f) = R_0 \subset \bar{C}^c$. The support R_0 of the test function f is thus a *compact* subset of the open set $\bar{C}^c$. For any x we denote by $b(x;\rho)$ the open ball of radius $\rho > 0$ centered at x [where Minkowski space is regarded as a Euclidean space with Cartesian coordinates $x = (x^1, x^2, x^3, x^4)$]. Now, for each $x \in R_0$ we can select a $\rho(x) > 0$ such that $b(x; 2\rho(x)) \subset W$ for some $W \in \mathcal{W}$ such that $W \subset \bar{C}^c$. The set $\{b(x;\rho(x)) \mid x \in R_0\}$ of open balls covers R_0, and, since R_0 is compact, this open covering contains a *finite* subcovering. There thus exists a finite set $\{x_k \mid x_k \in R_0,\ k = 1, \ldots, n\}$ of points, and a set $\{W_k \mid W_k \in \mathcal{W},\ k = 1, \ldots, n\}$ of wedges, such that

$$R_0 \subset \cup \{b(x_k; \rho(x_k)) \mid k = 1, \ldots, n\}, \tag{103a}$$

$$b(x_k; 2\rho(x_k)) \subset W_k \subset \bar{C}^c, \quad k = 1, \ldots, n. \tag{103b}$$

In view of (103a) there then exists a set $\{g_k(x) \mid g_k \in \mathcal{D}(R^4),\ k = 1, \ldots, n\}$ of functions such that $\mathrm{supp}(g_k) \subset b(x_k; 2\rho(x_k))$ for $k = 1, \ldots, n$, and

$$\sum_{k=1}^{n} g_k(x) = 1, \quad \text{all } x \in R_0. \tag{103c}$$

Let $(Y, D_1) = (\varphi_\mu[f], D_1)$ and $(Y_k, D_1) = (\varphi_\mu[fg_k], D_1)$ for $k = 1, \ldots, n$. We then have

$$(Y, D_1) = \sum_{k=1}^{n} (Y_k, D_1), \tag{103d}$$

where $(Y_k, D_1) \in \mathcal{L}(W_k)$. If now $X \in \mathcal{C}_i$, then $X \in \mathcal{C}_w(\overline{W_k^c})$ and hence X commutes in the weak sense (64) with $(Y_k, D_1)^\varepsilon$ for $k = 1, \ldots, n$. If follows, in view of (103d), that

$$\langle (Y^\dagger)^\varepsilon \phi \mid X\psi \rangle = \langle X^* \phi \mid Y^\varepsilon \psi \rangle \tag{103e}$$

for all $\phi, \psi \in D_1$.

(3) For any $X \in \mathcal{C}_i$ the relation (103e) thus holds for all $(Y, D_1) = (\varphi_\mu[f], D_1) \in \mathcal{L}(\bar{C}^c)$ such that $\mathrm{supp}(f)$ is compact. The set $\mathcal{D}(R^4)$ is dense in $\mathcal{S}(R^4)$ in the topology of the space of tempered test functions, and, since the quantum fields are operator-valued tempered distributions, it readily follows that (103e) holds for *all* $(Y, D_1) = (\varphi_\mu[f], D_1) \in \mathcal{L}(\bar{C}^c)$ such that $f \in \mathcal{S}(R^4)$, $\mathrm{supp}(f) \subset \bar{C}^c$, i. e., for all elements of $\mathcal{L}(\bar{C}^c)$. It then follows, in view of Lemma 9, part (c), that $X \in \mathcal{C}_w(\bar{C})$. This, in effect, completes the proof of the lemma.

We are now prepared to present the main result of this section.

Theorem 7: Let the quantum fields be such that the conditions in part (a) of Theorem 4 are satisfied, i. e., the von Neumann algebra $\mathcal{A}(W_R) = \mathcal{A}_0(W_R)$ satisfies the relations

$$\mathcal{A}(W_R) = \mathcal{C}(W_R) = \mathcal{C}_w(W_R) = \mathcal{G}(W_R), \tag{104}$$

and hence the algebra is locally and TCP-symmetrically associated with W_R. Furthermore, $\mathcal{A}(W_R)$ satisfies the condition of duality, and the conditions (100). Let a local AB-system be constructed from $\mathcal{A}(W_R)$, as in Definition 4. Then:

(a) The algebra $\mathcal{A}(W_R)$ satisfies all the general and special premises of Theorems 5 and 6, and all the conclusions of these theorems apply. In particular $\beta(\bar{C}_0)\Omega$ is dense for *any* $C_0 \in \mathcal{D}_c$. Furthermore, for any $C_0 \in \mathcal{D}_c$ such that $\bar{C}_0 \subset W_R$,

$$\mathcal{A}(W_R) = \{V(t)\mathcal{G}(C_0)V(t)^{-1} \mid t \in R^1\}'', \tag{105a}$$

$$\mathcal{A}(\bar{C}_1^c) = \{\mathcal{G}(\Lambda C_0) \mid \Lambda \in \bar{L}_0,\ \Lambda \bar{C}_0 \subset \bar{C}_1^c\}''. \tag{105b}$$

(b) For any $C \in \mathcal{D}_c$,

$$\mathcal{C}(\bar{C}) \subset \mathcal{C}_w(\bar{C}) = \beta(\bar{C}), \quad \mathcal{G}(\bar{C}) \subset \beta(\bar{C}), \tag{106a}$$

$$\mathcal{C}_w(\bar{C}^c) \supset \mathcal{C}(\bar{C}^c) \supset \mathcal{A}(\bar{C}^c), \quad \mathcal{G}(\bar{C}^c) \supset \mathcal{A}(\bar{C}^c). \tag{106b}$$

(c) With the notation of Lemma 10, $\mathcal{A}_0(\bar{C}) = \mathcal{C}_w(\bar{C}) = \beta(\bar{C})$ for all $C \in \mathcal{D}_c$. For any such C the operators in $\mathcal{P}(\bar{C}^c)$ have extensions constructed as in part (c) of Lemma 10, and these extensions have the properties described in the lemma. In particular the closures and adjoints of the extended operators are affiliated to the von Neumann algebra $\mathcal{A}(\bar{C}^c)$.

(d) With the notation of Lemma 10, $\mathcal{C}_w(\bar{C}^c) \supset \mathcal{A}_0(\bar{C}^c) \supset \mathcal{C}(\bar{C}^c)$ for all $C \in \mathcal{D}_c$. For any such C the operators in $\mathcal{P}(\bar{C})$ have extensions constructed as in part (c) of Lemma 10, and these extensions have the properties described in the lemma. In particular the closures and adjoints of the extended operators are affiliated to the von Neumann algebra $\mathcal{A}_0(\bar{C}^c)^a \subset \mathcal{G}(\bar{C}) \subset \beta(\bar{C})$.

Proof: (1) The algebra $\mathcal{A}(W_R)$ trivially satisfies the general premises of Theorem 5. From the construction of the AB-system, and from (104), it follows, in view of Lemma 12, that $\mathcal{C}_w(\bar{C}) = \beta(\bar{C})$.

Since the mapping $R \to \mathcal{G}(R)$ satisfies the condition of isotony, the inclusion relation at right in (106a) follows from (104). The remaining relations (106a) and (106b) are then trivial.

(2) Since, by Lemma 9, $\mathcal{G}(C)\Omega$ is dense for any $C \in \mathcal{D}_c$ it follows that $\beta(\bar{C})\Omega$ is dense, as asserted in part (a) of the theorem. Let now $C_0 \in \mathcal{D}_c$ and $\bar{C}_0 \subset W_R$. Let $\mathcal{A}_R$ denote the von Neumann algebra defined by the right member in (105a). The vector Ω is then a cyclic vector for $\mathcal{A}_R$, and in view of the construction we have $V(t)\mathcal{A}_R V(t)^{-1} = \mathcal{A}_R$ for all real t. Furthermore, it is trivially the case that $\mathcal{A}(W_R) \supset \mathcal{A}_R$. It then follows from Theorem 2 in BW I that $\mathcal{A}(W_R) = \mathcal{A}_R$, as asserted in (105a). The relation (105b) follows trivially from the relation (105a).

(3) The assertions (c) and (d) of the theorem are trivial in view of Lemma 10.

As we see from this theorem, a very satisfactory "local" theory results if the quantum fields satisfy the premises of Theorem 4, i. e., any one of the six conditions in part (a) of that theorem. There thus exists a local AB-system which satisfies the condition of TCP symmetry and the condition of duality $\beta(\bar{C})^a = \mathcal{A}(\bar{C}^c)$.

Furthermore, for any $C \in \mathcal{O}_c$, the von Neumann algebra $\mathcal{B}(\bar{C})$ has Ω as a cyclic and separating vector. The relations (101a)—(101d) hold, which means that the set of local operators associated with the bounded regions C is sufficiently large in the sense that these operators generate all the algebras of the AB-system, as described by the relations (101a)—(101d). Now it is interesting to note that the algebra $\mathcal{B}(\bar{C})$ is in fact equal to the weak quasicommutant $\mathcal{C}_w(\bar{C})$ of the set of all field operators of the form $(\varphi_\mu[f], D_1)$, where $f \in \mathcal{S}(R^4)$, $\text{supp}(f) \subset \bar{C}^c$. We thus have a conceptually simple prescription for "finding" the algebras $\mathcal{B}(\bar{C})$ provided that it has first been established that the quantum fields do satisfy the premises of Theorem 4.

We note here that this is the case under what we called Condition I in BW I, because this condition says that $\mathcal{C}(W_R)\Omega$ is dense. It follows that all the conclusions in Theorem 7 hold under our earlier Condition I. We overlooked this fact in our previous paper.

We infer from the work of Landau[17] that $\mathcal{G}(\bar{C})$ is in general smaller than $\mathcal{B}(\bar{C})$. The study of Landau is concerned with generalized free fields, in which case we have the further simplification that $\mathcal{C}_w(R) = \mathcal{C}(R)$ for any subset R of $\mathcal{M}$. We then have $\mathcal{A}(\bar{C}^c) = \mathcal{G}(\bar{C}^c)$ and $\mathcal{B}(\bar{C}) = \mathcal{C}(\bar{C})$, but it can well happen that $\mathcal{G}(\bar{C}) \neq \mathcal{B}(\bar{C})$.

We conclude by stating a theorem about local internal symmetries.

Theorem 8: Let $\mathcal{A}(W_R)$ be a von Neumann algebra locally and TCP-symmetrically associated with W_R. It is assumed that $\mathcal{A}(W_R)$ satisfies the condition of duality, and that furthermore

$$X\Omega \subset D_+, \quad V(i\pi)X\Omega = JX^*\Omega \tag{107}$$

for all $X \in \mathcal{A}(W_R)$. Let a local AB-system be constructed in terms of $\mathcal{A}(W_R)$ as in Definition 4.

Let G be a unitary operator such that

$$G\Omega = \Omega, \quad G\mathcal{A}(W)G^{-1} = \mathcal{A}(W), \quad \text{all } W \in \mathcal{W}. \tag{108a}$$

Then:

(a) The operator G commutes with the TCP transformation, and with all Poincaré transformations, i.e.,

$$\Theta_0 G \Theta_0^{-1} = G, \quad U(\lambda)GU(\lambda)^{-1} = G, \quad \text{all } \lambda \in \bar{\mathcal{P}} \tag{108b}$$

(b) For all double cones C,

$$G\mathcal{B}(\bar{C})G^{-1} = \mathcal{B}(\bar{C}), \quad G\mathcal{A}(\bar{C}^c)G^{-1} = \mathcal{A}(\bar{C}^c). \tag{108c}$$

(c) The set of all unitary operators G which satisfy the conditions (108a) forms a group, *the group of all local internal symmetries*.

This theorem is proved by the same reasoning as in our proof of the corresponding Theorem 7 in BW I, and it is not necessary to repeat the arguments here. We note here that the conclusions of the theorem do not follow (as far as we know) merely from the assumptions that $\mathcal{A}(W_R)$ satisfies the condition of duality and is locally and TCP-symmetrically associated with W_R. Our proof in BW I depends on the specific conditions (107), which presumably characterize local von Neumann algebras in a quantum field theory. Without the conditions (107) it can be shown[18] that G commutes with all translations, but it appears that further assumptions are necessary for the conclusion that G also commutes with homogeneous Lorentz transformations.[19]

We finally note that the "group of all local internal symmetries," as defined above, will in general include superselection symmetries with no observable physical effects.

ACKNOWLEDGMENTS

It is a pleasure to thank Mr. Ray T. Hagstrom and Dr. Marc A. Rieffel for discussions about some of the topics of this paper.

*Work supported by the U.S. Energy Research and Development Administration.

†Present address: Belfer Graduate School of Science, Yeshiva University, New York, New York, 10033.

‡Work supported in part by the National Science Foundation under Grant No. MPS 74-08175-A01.

[1] J.J. Bisognano and E.H. Wichmann, J. Math. Phys. **16**, 985 (1975).

[2] R.F. Streater and A.S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[3] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R.I., 1965).

[4] R. Jost, Helv. Phys. Acta **30**, 409 (1957).

[5] For further details we refer to the discussion in Ref. 2, Chap. 1, and in Ref. 3, Appendix 1.

[6] For a discussion of this issue, see Ref. 2, Chap. 2.

[7] H. Reeh and S. Schlieder, Nuovo Cimento **22**, 1051 (1961).

[8] See pp. 131 and 146 in Ref. 2. We note that it is immaterial if we define Θ_0 as we have defined it, or with an extra factor U_0. We refer to the quoted monograph for comments on the question whether this "*TCP* transformation" always has the physical interpretation suggested by the name: In certain kinds of field theories there could conceivably exist other antiunitary transformations which deserve the name better.

[9] Our term "quasicommutant" in Definition 1 has no offical standing in the literature. We felt a need for a term to describe the relationship in question, and we selected the term quasicommutant (not without some misgivings) as one which to some extent suggests the relationship, and as one unencumbered by well-established mathematical connotations.

[10] It should be noted that our device of "turning anticommutators into commutators" with the help of the operator Z is intimately related to a so-called Klein transformation: O. Klein, J. Phys. Radium **9**, 1 (1938). For discussions of the Klein transformations in quantum field theory, see also G. Lüders, Z. Naturforsch. **13a**, 254 (1958), and H. Araki, J. Math. Phys. **2**, 267 (1961).

[11] J. von Neumann, Math. Ann. **102**, 370 (1929).

[12] F.J. Murray and J. von Neumann, Ann. Math. **37**, 116 (1936), Lemma 4.4.1.

[13] R.T. Powers, Commun. Math. Phys. **21**, 85 (1955), and Trans. Amer. Math. Soc. **187**, 261 (1974). It should be noted that in these papers the prime notation is employed for what we would call the weak commutant.

[14] A.L. Licht, J. Math. Phys. **4**, 1443 (1963).

[15] H.J. Borchers, Nuovo Cimento **15**, 784 (1960).

[16] M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras and Its Applications* (Springer-Verlag, Berlin, 1970).

[17] L.J. Landau, Commun. Math. Phys. **39**, 49 (1974).

[18] L.J. Landau and E.H. Wichmann, J. Math. Phys. **11**, 306 (1970).

[19] L.J. Landau, Commun. Math. Phys. **17**, 156 (1970).

Gauge-covariant differentiation and Green's functions for the Yang–Mills field

Heinz Frey

Physikalisches Institut der Universität Bonn, Germany
(Received 22 July 1975)

A gauge-independent definition of differentiation is given for non-Abelian gauge fields in terms of parallel translation. This is achieved by a suitable definition of time-ordered operator products. Equal time commutation relations are used to derive the differential equations for the related Green's functions. The Green's functions are discussed for general linear gauges. In comparison with electrodynamics the Green's functions have the well-known ghost-loop terms.

1. INTRODUCTION

The purpose of this paper is to extend a formalism developed by Meetz[1] for scalar electrodynamics to non-Abelian gauge theories. In this formalism the concept of covariant differentiation (or parallel displacement) is transferred to suitably defined time-ordered products. The field equations obeyed by time-ordered products and Green's functions then naturally arise. Because this concept is not restricted to the Abelian gauge group one should be able to treat non-Abelian gauge groups with their complications in the same way. We will show that this in fact is the case.

Throughout the paper we shall consider a gauge theory with local group SU(2). This is without loss of generality as the results are easily translated to the case of any compact simple local gauge group. Now in the usual formalism one defines the gauge-covariant derivative of a field ϕ^α by

$$(\nabla_\mu\phi)^\alpha(x) := \nabla_\mu^{\alpha\beta}(x)\,\phi^\beta(x) := [\,\delta_{\alpha\beta}\partial_\mu - ig\mathbf{T}\mathbf{A}_\mu^{\alpha\beta}(x)]\,\phi^\beta(x)$$
$$= [\delta_{\alpha\beta}\partial_\mu - ig(T^a)^{\alpha\beta}A_\mu^a(x)]\,\phi^\beta(x). \tag{1.1}$$

Here A_μ^a is the gauge field potential (with the index a taking the values 1, 2, 3), T^a are the generators of the Lie algebra of SU(2) in the representation corresponding to the transformation properties of the field ϕ^α, and g is the coupling constant. In the adjoint representation, which for simplicity we always consider, we have $(T^a)^{bc} = i\epsilon_{abc}$. $\nabla_\mu\phi$ is covariant under the gauge transformation

$$\phi^a(x) \to \exp[ig\mathbf{T}\boldsymbol{\theta}(x)]^{ab}\phi^b(x), \tag{1.2a}$$
$$\mathbf{T}\mathbf{A}_\mu(x) \to \exp[ig\mathbf{T}\boldsymbol{\theta}(x)]\,\mathbf{T}\mathbf{A}_\mu(x)\exp[-ig\mathbf{T}\boldsymbol{\theta}(x)]$$
$$-(1/ig)\exp[ig\mathbf{T}\boldsymbol{\theta}(x)]\,\partial_\mu\exp[-ig\mathbf{T}\boldsymbol{\theta}(x)], \tag{1.2b}$$

where $\theta^a(x)$ are arbitrary functions. The relation between the gauge field potentials and the field strengths is given by

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)^{ab}\phi^b(x) = -g\epsilon_{abc}F_{\mu\nu}^b(x)\,\phi^c(x), \tag{1.3a}$$

where

$$F_{\mu\nu}^a(x) = \partial_\mu A_\nu^a(x) - \partial_\nu A_\mu^a(x) - g\epsilon_{abc}A_\mu^b(x)A_\nu^c(x). \tag{1.3b}$$

A further covariant derivation ∇_ρ and application of the Jacobi identity gives the analog of the homogeneous Maxwell equations

$$\nabla_\rho^{ab}(x)F_{\mu\nu}^b(x) + \nabla_\mu^{ab}(x)F_{\nu\rho}^b(x) + \nabla_\nu^{ab}(x)F_{\rho\mu}^b(x) = 0. \tag{1.4}$$

$F_{\mu\nu}^a$ transforms under gauge transformations as an isovector

$$F_{\mu\nu}^a(x) \to \exp[ig\mathbf{T}\boldsymbol{\theta}(x)]^{ab}F_{\mu\nu}^b(x). \tag{1.5}$$

All the complications of the non-Abelian gauge field stem from the fact that, because the adjoint representation is not trivial, field strengths are gauge dependent. Hence the covariant derivatives do not reduce to ordinary ones.

We can also consider this formalism from a more differential geometric point of view. We regard (1.1) as a covariant derivative in the sense of differential geometry with linear connection $g\epsilon_{abc}A_\mu^c(x)$. Parallel translation of a vector $\phi^a(x)$ along a path $\xi(s)$ is then defined by the differential equation

$$\frac{d}{ds}\phi^a(\xi(s)) - g\epsilon_{abc}\frac{d\xi^\mu(s)}{ds}A_\mu^b(\xi(s))\,\phi^c(\xi(s)) = 0. \tag{1.6}$$

The solution of (1.6) for parallel translation along a finite path is given by

$$\phi^a(x) \to \phi^a(x(s)) = T\exp[ig\int_0^s d\xi^\mu\mathbf{T}\mathbf{A}_\mu(\xi)]^{ab}\phi^b(x)$$
$$=: \chi_s^{ab}(x)\phi^b(x), \tag{1.7}$$

where T means ordering along the path.

Now one should consider two vectors $\phi^a(x)$, $\phi^b(y)$ as physically equivalent if they can be transferred into each other by parallel translation. In the presence of a gauge field this concept of physical equivalence does depend on the path chosen because under a deformation of the path, with the endpoints fixed, we get

$$\delta_t\chi_s^{ab}(x) = -g\int_0^s d\xi^\mu(s')\,\delta\xi^\nu(s')\,\chi_{s'}^{ac}(x)\epsilon_{cde}F_{\mu\nu}^d(\xi(s'))\,\chi_s^{eb}(\xi(s')), \tag{1.8}$$

where $\chi_s^{eb}(\xi(s')) = T\exp(ig\int_{s'}^s d\xi^\mu\mathbf{T}\mathbf{A}_\mu)^{eb}$. The dependence on the path prevents, in our view, this notion of physical equivalence to be of relevance. But locally we are able to define the gauge-covariant derivative unam-

biguously by parallel translation,

$$\lim_{s\to 0}\frac{1}{s}[\chi_s^{-1ab}(x)\,\phi^b(x(s))-\phi^a(x)]$$
$$=:\left.\frac{d\xi^\mu}{ds}\right|_x(\nabla_\mu\phi)^a(x)=\left.\frac{d\xi^\mu}{ds}\right|_x(\delta_{ab}\partial_\mu-g\epsilon_{acb}A_\mu^c(x))\,\phi^b(x). \tag{1.9}$$

From our point of view only the local implications of parallelism should be ingredients of a physical theory, in particular with respect to the quantized theory. This is in contrast to the view presented by Mandelstam in Refs. 2 and 3. He introduced field variables that are gauge independent, but depend on a path leading from the space-time point x to infinity. For an isovector field $\phi^a(x)$ the path dependent field $\Phi^a(x,P)$ is constructed by parallel translation of ϕ^a along the path P from x to infinity,

$$\Phi^a(x,P):=T^+\exp[-ig\int_P d\xi^\mu\mathbf{T}\mathbf{A}_\mu(\xi)]^{ab}\phi^b(x). \tag{1.10}$$

Here T^+ means antipath-ordering.

We prefer a formalism that makes use of the concept of parallelism only locally. In Sec. 2 we shall derive the equal time commutation relations for gauge field variables and a scalar isovector field. The procedure will be closely analogous to that given in Ref. 1. We also restrict ourselves to linear gauges. Our program then requires the definition of parallel translation or gauge-covariant differentiation of time-ordered products or Green's functions respectively. These concepts will be introduced in Sec. 3 and the resulting field equations will be determined. In Sec. 4 we shall summarize these equations in the condensed notation invented by Mandelstam[3] and write them in integrated form. The equations differ from those naively expected by an additional term, which in perturbation theory produces the well-known "ghost" loops in Feynman diagrams, found by Feynman,[4] De Witt,[5] Faddeev and Popov,[6] Mandelstam,[3] and others.

2. COMMUTATION RELATIONS

Our starting point is the Lagrangian of a non-Abelian gauge field with structure group SU(2), coupled to a scalar isovector field

$$L(x)=-\tfrac{1}{4}F^a_{\mu\nu}(x)F_a^{\mu\nu}+\tfrac{1}{2}[\nabla^\mu_{ab}\phi^b(x)\,\nabla^{ac}_\mu\phi^c(x)-m^2\phi^a(x)\,\phi^a(x)]. \tag{2.1}$$

The additional isovector field will not add appreciably to the complexity of our system, but will help us to infer the commutation relations in accordance with the classical Poisson-brackets that can be derived from (2.1), e.g., by the method of Peierls.[7] From the Lagrangian (2.1) we derive the classical equations of motion

$$\nabla^\nu_{ab}(x)F_{\nu\mu}{}^b(x)+g\nabla^{bc}_\mu(x)\phi^c(x)\epsilon_{abd}\phi^d(x)$$
$$=:\nabla^\nu_{ab}(x)\,F_{\nu\mu}{}^b(x)-j^a_\mu(x)=0, \tag{2.2a}$$

$$\nabla^\mu_{ab}(x)\nabla^{bc}_\mu(x)\phi^c(x)+m^2\phi^a(x)=0. \tag{2.2b}$$

We shall assume that these equations, suitably symmetrized, are still valid in the quantized theory. Here we want to remark on our treatment of noncommuting operators in operator products. We always regard such products as symmetrized and indicate this by a dot, $A\circ B:=\frac{1}{2}(AB+BA)$, and appropriate brackets if more than two factors are involved. Some of our calculations require a reordering of symmetrized products. This can be achieved by the formula $A\circ(B\circ C)=(A\circ B)\circ C-\frac{1}{4}[[A,C],B]$. In our calculations the double commutator always vanishes.

Now we proceed with the identification of the equal time commutation relations. The commutators that can be written down from the outset are those between the scalar variables and their conjugate momenta. Non-vanishing are

$$[\phi^a(\mathbf{x},x^0),(\nabla_0\phi)^b(\mathbf{y},x^0)]=i\delta_{ab}\delta(\mathbf{x}-\mathbf{y}). \tag{2.3}$$

In contrast to the Abelian case the commutators between field strength operators are gauge dependent and we have to derive them. To do so we follow closely the procedure given in Ref. 1 that relies on the method of Peierls and consistency considerations.

First we determine the commutation relations between the field strength operators and the scalar variables. A^a_i, hence F_{ij}^a, and ϕ^a commute, because these variables are independent. [This is not generally true. Rather it is this assumption that confines us to "coulomblike" gauges in the following. These gauges are defined by imposing a gauge condition on the spatial components A^a_i of the potentials and treating A^α_0 as dependent variable.]

The commutation relations between F_{0i}^a and ϕ^b can be derived by considering the zero-component of Eq. (2.2a), which is an equation of constraint,

$$\nabla^i_{ab}(x)\circ F_{0i}{}^b(x)=-j^a_0(x)=g(\nabla^{bc}_0(x)\,\phi^c(x))\,\epsilon_{abd}\phi^d(x). \tag{2.4}$$

Relation (2.3) implies

$$[-j^a_0(\mathbf{x},x^0),\ \phi^b(\mathbf{y},x^0)]=-ig\delta(\mathbf{x}-\mathbf{y})\epsilon_{abc}\phi^c(\mathbf{y},x^0), \tag{2.5a}$$
$$[-j^a_0(\mathbf{x},x^0),(\nabla_0\phi)^b(\mathbf{y},x^0)]=-ig\delta(\mathbf{x}-\mathbf{y})\epsilon_{abc}(\nabla_0\phi)^c(\mathbf{y},x^0). \tag{2.5b}$$

In view of (2.4) and our assumption that A^a_i and the scalar variables commute, we conclude from (2.5a)

$$[F_{0i}^a(\mathbf{x},x^0),\ \phi^b(\mathbf{y},x^0)]=-g\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y})\,\phi^c(\mathbf{y},x^0), \tag{2.6a}$$

where $\hat{\Gamma}_i$ fulfills the equation

$$\nabla^i_{ad}(x)\circ\hat{\Gamma}_i^{dbc}(\mathbf{x},\mathbf{y})=i\epsilon_{abc}\,\delta(\mathbf{x}-\mathbf{y}). \tag{2.7a}$$

Because of the hermiticity of F_{0i}^a and ϕ^b it follows from (2.6a) that

$$\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y})\,\phi^c(\mathbf{y},x^0)=-\phi^c(\mathbf{y},x^0)\,\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y})^\dagger. \tag{2.8a}$$

Now we consider the commutator between F_{0i}^a and $\phi^b\phi^b$. The Eqs. (2.6a) and (2.8a) imply

$$[F_{0i}^a(\mathbf{x},x^0),\ \phi^b(\mathbf{y},x^0)\,\phi^b(\mathbf{y},x^0)]$$
$$=-g\,\phi^b(\mathbf{y},x^0)\,[\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y})-\hat{\Gamma}_i^{acb}(\mathbf{x},\mathbf{y})^\dagger]\,\phi^c(\mathbf{y},x^0).$$

On the other hand we can calculate this commutator by the method of Peierls. We find that the gauge invariant quantities $F_{0i}^a\circ F_{jk}^a$ and $\phi^b\phi^b$ commute. But F_{jk}^a

and $\phi^b\phi^b$ commute as well and therefore the same holds for F_{0i}^a and $\phi^b\phi^b$. To satisfy this condition we demand

$$\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y}) = \hat{\Gamma}_i^{acb}(\mathbf{x},\mathbf{y})^\dagger. \tag{2.8b}$$

In the same way the method of Peierls shows that F_{0i}^a and $(\nabla_0\phi)^b\phi^b$, $(\nabla_0\phi)^b(\nabla_0\phi)^b$ respectively, commute. This, together with (2.6a), (2.8b), and the hermiticity of $(\nabla_0\phi)^b$ gives

$$[F_{0i}^a(\mathbf{x},x^0),(\nabla_0\phi)^b(\mathbf{y},x^0)] = g(\nabla_0\phi)^c(\mathbf{y},x^0)\,\hat{\Gamma}_i^{acb}(\mathbf{x},\mathbf{y}) = -g\,\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y})(\nabla_0\phi)^c(\mathbf{y},x^0). \tag{2.6b}$$

From now on we shall assume that $\hat{\Gamma}_i^{abc}$ is a function only of the spatial components A_i^a of the potential. Because $\hat{\Gamma}_i^{abc}$ then commutes with ϕ^b and $(\nabla_0\phi)^b$ the relations (2.8a) and (2.8b) show that it has to be antisymmetric in the indices b and c. Therefore we can write

$$\hat{\Gamma}_i^{abc}(\mathbf{x},\mathbf{y}) = i\Gamma_i^{ad}(\mathbf{x},\mathbf{y})\epsilon_{dbc}. \tag{2.9}$$

From (2.8b) we see that Γ_i^{ab} has to be Hermitian. It obeys a differential equation that can be derived from (2.7a),

$$\nabla_{ad}^i(x)\,\Gamma_i^{db}(\mathbf{x},\mathbf{y}) = \delta_{ab}\,\delta(\mathbf{x}-\mathbf{y}). \tag{2.7b}$$

In terms of Γ_i^{ab} the commutation relations (2.6a) and (2.6b) now read

$$[F_{0i}^a(\mathbf{x},x^0),\ \phi^b(\mathbf{y},x^0)] = -ig\,\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\,\epsilon_{cbd}\,\phi^d(\mathbf{y},x^0), \tag{2.10a}$$

$$[F_{0i}^a(\mathbf{x},x^0),\ (\nabla_0\phi)^b(\mathbf{y},x^0)] = -ig\,\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\epsilon_{cbd}(\nabla_0\phi)^d(\mathbf{y},x^0). \tag{2.10b}$$

This is in close analogy to the Abelian case. The important difference is that now, according to Eq. (2.7), Γ_i generally is a q-number function. An exception to this is the axial gauge, as will be seen below.

Next we derive the commutation relations between F_{0i}^a and the spatial covariant derivatives $(\nabla_j\phi)^b$. To achieve this we first compute the commutator of $F_{0i}^a \circ F_{jk}^a$ and $(\nabla_0\phi)^b(\nabla_j\phi)^b$ by the method of Peierls. As F_{jk}^a commutes with $(\nabla_0\phi)^b(\nabla_j\phi)^b$ we can, as in our previous calculation "divide" by it and we get

$$[F_{0i}^a(\mathbf{x},x^0),\ (\nabla_0\phi)^b(\mathbf{y},x^0)(\nabla_j\phi)^b(\mathbf{y},x^0)] = -ig\,\delta_{ij}\delta(\mathbf{x}-\mathbf{y})\epsilon_{abc}(\nabla_0\phi)^b(\mathbf{y},x^0)\,\phi^c(\mathbf{y},x^0). \tag{2.11}$$

Together with (2.10b) this implies

$$[F_{0i}^a(\mathbf{x},x^0),\ (\nabla_j\phi)^b(\mathbf{y},x^0)] = -ig\delta_{ij}\,\delta(\mathbf{x}-\mathbf{y})\epsilon_{abc}\,\phi^c(\mathbf{y},x^0) - ig\,\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\epsilon_{cbd}(\nabla_j\phi)^d(\mathbf{y},x^0). \tag{2.12}$$

If we express ∇_j explicitly in terms of the potential

$$\nabla_j^{bc}\phi^c(\mathbf{y},x^0) = [\delta_{bc}\partial_j - g\,\epsilon_{bdc}A_j^d(\mathbf{y},x^0)]\,\phi^c(\mathbf{y},x^0),$$

and expand the left hand side of (2.12) with the help of (2.10a) we find

$$[F_{0i}^a(\mathbf{x},x^0),\ A_j^b(\mathbf{y},x^0)] = -i[\delta_{ab}\delta_{ij}\,\delta(\mathbf{x}-\mathbf{y}) - \Gamma_i^{ac}(\mathbf{x},\mathbf{y})\,\overleftarrow{\nabla}_j^{cb}(\mathbf{y},x^0)]. \tag{2.13}$$

Here $\overleftarrow{\nabla}_j$ is defined by $\overleftarrow{\nabla}_j^{cb}(\mathbf{y},x^0) = [\delta_{cb}\overleftarrow{\partial}_j + g\epsilon_{cdb}A_j^d(\mathbf{y},x^0)]$, where the derivative acts on the function to its left. Because of $F_{jk}^a = \partial_j A_k^a - \partial_k A_j^a - g\epsilon_{abc}A_j^bA_k^c$, (2.13) leads to

$$[F_{0i}^a(\mathbf{x},x^0),\ F_{jk}^b(\mathbf{y},x^0)] = i[\delta_{ij}\nabla_k^{ab}(\mathbf{x},x^0) - \delta_{ik}\nabla_j^{ab}(\mathbf{x},x^0)]\,\delta(\mathbf{x}-\mathbf{y}) - ig\,\Gamma^{ac}(\mathbf{x},\mathbf{y})\,\epsilon_{cbd}\,F_{jk}^d(\mathbf{y},x^0). \tag{2.14}$$

Equations (2.13) and (2.14) are consistent with (2.4).

Finally we have to determine the commutator between F_{0i}^a and F_{0j}^b. To this end we calculate

$$[F_{0i}^a(\mathbf{x},x^0),\ \nabla_{bc}^j(\mathbf{y},x^0)\circ F_{0j}^c(\mathbf{y},x^0)] = -[F_{0i}^a(\mathbf{x},x^0),\ j_0^b(\mathbf{y},x^0)] = ig\,\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\epsilon_{cbd}j_0^d(\mathbf{y},x^0). \tag{2.15}$$

On the other hand, because of (2.13), we can evaluate the left hand side as

$$[F_{0i}^a(\mathbf{x},x^0),\ \nabla_{bc}^j(\mathbf{y},x^0)\circ F_{0j}^c(\mathbf{y},x^0)] = \nabla_{bc}^j(\mathbf{y},x^0)\circ[F_{0i}^a(\mathbf{x},x^0),\ F_{0j}^c(\mathbf{y},x^0)] + ig\{-\delta(\mathbf{x}-\mathbf{y})\epsilon_{bac} \times F_{0i}^c(\mathbf{x},x^0) + [\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\,\overleftarrow{\nabla}_j^{cd}(\mathbf{y},x^0)]\circ\epsilon_{dbe}\,F_{0j}^e(\mathbf{y},x^0)\}. \tag{2.16}$$

Using (2.7b) and a reordering of the symmetrization, we can transform the last term,

$$ig\{-\delta(\mathbf{y}-\mathbf{x})\,\epsilon_{bac}F_{0i}^c(\mathbf{x},x^0) + [\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\,\overleftarrow{\nabla}_j^{cd}(\mathbf{y},x^0)]\circ\epsilon_{dbe}F_{0j}^e(\mathbf{y},x^0)\} = ig\nabla_{bc}^j(\mathbf{y},x^0)\circ[-\Gamma_j^{cd}(\mathbf{y},\mathbf{x})\circ\epsilon_{dae}F_{0i}^e(\mathbf{x},x^0) + \Gamma_i^{ad}(\mathbf{x},\mathbf{y})\circ\epsilon_{dce}F_{0j}^e(\mathbf{y},x^0)] + ig\,\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\,\epsilon_{cbd}j_0^d(\mathbf{y},x^0). \tag{2.17}$$

Together with (2.15) we have

$$\nabla_{bc}^j(\mathbf{y},x^0)\circ[F_{0i}^a(\mathbf{x},x^0),\ F_{0j}^c(\mathbf{y},x^0)] = \nabla_{bc}^j(\mathbf{y},x^0)\circ[-ig\,\Gamma_i^{ad}(\mathbf{x},y)\circ\epsilon_{dce}F_{0j}^e(\mathbf{y},x^0) + ig\,\Gamma_j^{cd}(\mathbf{y},\mathbf{x})\circ\epsilon_{dae}F_{0i}^e(\mathbf{x},x^0)]. \tag{2.18}$$

Hence the commutator reads

$$[F_{0i}^a(\mathbf{x},x^0),\ F_{0j}^b(\mathbf{y},x^0)] = -ig\Gamma_i^{ac}(\mathbf{x},\mathbf{y})\circ\epsilon_{cbd}F_{0j}^d(\mathbf{y},x^0) + ig\Gamma_j^{bc}(\mathbf{y},\mathbf{x})\circ\epsilon_{cad}F_{0i}^d(\mathbf{x},x^0). \tag{2.19}$$

This is what can be derived without more definite assumptions on the gauge defining operator Γ_i. To go on we shall require in the following that A_i^a obeys a linear gauge condition,

$$\int d\tau(\mathbf{z})A_i^b(\mathbf{z},y^0)\,\gamma_{ba}^i(\mathbf{z},\mathbf{y}) = B^a(\mathbf{y},y^0). \tag{2.20}$$

Here B^a is a c-number function and γ_i^{ab} is a c-number matrix that is assumed to satisfy

$$\partial^i\gamma_i^{ab}(\mathbf{x},\mathbf{y}) = \delta_{ab}\,\delta(\mathbf{x}-\mathbf{y}). \tag{2.21}$$

Substituting (2.20) in (2.13) we find that this gauge condition requires

$$\Gamma_i^{ab}(\mathbf{x},\mathbf{y}) = \int d\tau(z)\,\gamma_i^{ac}(\mathbf{x},\mathbf{z})\,(\nabla^l\gamma_l)^{-1}(\mathbf{z},\mathbf{y})^{cb}, \tag{2.22}$$

provided that the operators involved exist.

To solve for the dependent variables A_0^a we start from

the equation

$$F_{0i}^a(x) = \partial_0 A_i^a(x) - \partial_i A_0^a(x) - g\epsilon_{abc} A_i^c(x) \circ A_0^b(x)$$
$$= \partial_0 A_i^a(x) - \nabla_i^{ab}(x) \circ A_0^b(x), \tag{2.23}$$

which solves the field equation (1.4). If we integrate this equation with γ_i, we get

$$\int d\tau(\mathbf{z})\, F_{0i}^b(\mathbf{z}, x^0)\, \gamma_{ba}^i(\mathbf{z}, \mathbf{x})$$
$$= \partial_0 B^a(\mathbf{x}, x^0) + \int d\tau(\mathbf{z}) A_0^b(\mathbf{z}, x^0) \circ [\nabla_i^{bc}(\mathbf{z}, x^0)\gamma_{ca}^i(\mathbf{z}, \mathbf{x})]. \tag{2.24}$$

Under the assumption that the commutator of A_0^a with A_i^b is a function of A_i^b we can remove the symmetrization of the product, which gives an additional commutator that is a function of A_i^b, and integrate with $(\nabla^l \gamma_l)^{-1}$. Upon symmetrization of the resulting expression the additional term vanishes and we finally arrive at

$$A_0^a(\mathbf{x}, x^0) = \int d\tau(\mathbf{z}) \{F_{0i}^b(\mathbf{z}, x^0) \circ \Gamma_{ba}^i(\mathbf{z}, \mathbf{x})$$
$$- [\partial_0 B^b(\mathbf{z}, x^0)](\nabla^l \gamma_l)^{-1}(\mathbf{z}, \mathbf{x})^{ba}\}. \tag{2.25}$$

This is in agreement with our assumptions on the commutator of A_0^a and A_i^b.

Lastly we calculate with the help of Eqs. (2.10), (2.13), (2.22), (2.7), and (2.25) the commutators of A_0^a with the other field variables:

$$[A_i^a(\mathbf{x}, x^0),\ A_0^b(\mathbf{y}, x^0)]$$
$$= -i[\Gamma_i^{ab}(\mathbf{x}, \mathbf{y}) - \nabla_i^{ac}(\mathbf{x}, x^0) \int d\tau(\mathbf{z})\, \Gamma_r^{dc}(\mathbf{z}, \mathbf{x})\, \Gamma_r^{db}(\mathbf{z}, \mathbf{y})], \tag{2.26a}$$

$$[F_{0i}^a(\mathbf{x}, x^0),\ A_0^b(\mathbf{y}, x^0)]$$
$$= i\partial_0 \Gamma_i^{ab}(\mathbf{x}, \mathbf{y}) + ig\epsilon_{acd} F_{0i}^d(\mathbf{x}, x^0)$$
$$\circ [\int d\tau(\mathbf{z})\, \Gamma_r^{ec}(\mathbf{z}, \mathbf{x})\, \Gamma_r^{eb}(\mathbf{z}, \mathbf{y})] - ig\Gamma_i^{ac}(\mathbf{x}, \mathbf{y})$$
$$\circ \epsilon_{cbd} A_0^d(\mathbf{y}, x^0) + igA_0^c(\mathbf{x}, x^0) \circ \epsilon_{cad}\, \Gamma_i^{db}(\mathbf{x}, \mathbf{y}), \tag{2.26b}$$

$$[A_0^a(x, x^0),\ \phi^b(\mathbf{y}, x^0)] = ig \int d\tau(\mathbf{z}) \Gamma_r^{ca}(\mathbf{z}, \mathbf{x})\, \Gamma_r^{cd}(\mathbf{z}, \mathbf{y}) \epsilon_{dbe} \phi^e(\mathbf{y}, x^0). \tag{2.26c}$$

To derive (2.26b), one has to use the formula

$$[F_{0i}^a(\mathbf{x}, x^0),\ \Gamma_j^{bc}(\mathbf{y}, \mathbf{z})]$$
$$= ig\Gamma_j^{bd}(\mathbf{y}, \mathbf{x})\epsilon_{dae}\Gamma_i^{ec}(\mathbf{x}, \mathbf{z})$$
$$+ ig \int d\tau(\mathbf{z}')$$
$$\times [\Gamma_i^{ad}(\mathbf{x}, \mathbf{z}')\overleftarrow{\nabla}_k^{de}(\mathbf{z}')\, \Gamma_j^{bf}(\mathbf{y}, \mathbf{z}')\epsilon_{feg}\Gamma_{gc}^k(\mathbf{z}', \mathbf{z})]$$
$$= ig\, \Gamma_j^{bd}(\mathbf{y}, \mathbf{x})\epsilon_{dae}\Gamma_i^{ec}(\mathbf{x}, \mathbf{z})$$
$$+ ig \int d\tau(\mathbf{z}')[\Gamma_j^{bd}(\mathbf{y}, \mathbf{z}')\, \overleftarrow{\nabla}_k^{de}(\mathbf{z}')$$
$$\times \Gamma_i^{af}(\mathbf{x}, \mathbf{z}')\epsilon_{feg}\Gamma_{gc}^k(\mathbf{z}', \mathbf{z})]$$
$$+ ig\Gamma_j^{bd}(\mathbf{y}, \mathbf{z})\Gamma_i^{ae}(\mathbf{x}, \mathbf{z})\epsilon_{edc}, \tag{2.26d}$$

that follows from (2.22). In addition (2.23) and (2.25) are needed to cast (2.26b) into the special form given above.

The commutators we have derived are a generalization of those obtained for the coulomb gauge by Schwinger.[8] Our formulas cover all linear "coulomb like" gauge conditions. For the coulomb gauge we have to choose

$$\gamma_{ab}^i(\mathbf{x}, \mathbf{y}) = -\delta_{ab} \frac{\partial}{\partial x^i} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|}. \tag{2.27}$$

Alternatively, we can choose a path gauge, as mentioned in Ref. 1, with

$$\gamma_{ab}^i(\mathbf{x}, \mathbf{y}) = -\delta_{ab} \int_{-\infty}^0 d\xi^i\, \delta(\mathbf{x} - \mathbf{y} - \xi), \tag{2.28}$$

where $\xi(s)$ denotes a fixed spatial path from zero to infinity. For $\xi(s) = \mathbf{n}s$, $\mathbf{n}$ an arbitrary unit vector, we recover the axial gauge used by Arnowitt, Fickler,[9] and Schwinger.[10] In the case $B^a = 0$ this gauge has the property that the component of the three-vector A_i^a in the direction of $\mathbf{n}$ vanishes. From this it follows that the corresponding $\gamma_{ab}^i(x, y)$,

$$\gamma_{ab}^i(\mathbf{x}, \mathbf{y})_A = -\delta_{ab} n^i \int_{-\infty}^0 ds\, \delta(\mathbf{x} - \mathbf{y} - \mathbf{n}s), \tag{2.29}$$

fulfills $A_i^b(x)\gamma_{ba}^i(\mathbf{x}, \mathbf{y})_A = 0$ without integration. This implies

$$\Gamma_{ab}^i(\mathbf{x}, \mathbf{y})_A = \gamma_{ab}^i(\mathbf{x}, \mathbf{y})_A, \tag{2.30}$$

as we see from the explicit expression (2.22). We shall remark on the implications of (2.30) in Sec. 4.

3. TIME-ORDERED PRODUCTS AND FIELD EQUATIONS

In this section we shall find a definition of time-ordered products of field operators, denoted by $\tilde{T}$, so that these products with a suitable definition of covariant differentiation (or parallel translation) obey Eqs. (1.3)–(1.4) and yield covariant field equations. $\tilde{T}$ products will differ from ordinary time-ordered products, denoted by T and defined explicitly by

$$T(A(x),\ B(y)) = \theta(x^0 - y^0) A(x) B(y) + \theta(y^0 - x^0) B(y) A(x), \tag{3.1}$$

by additional terms similar to those found in Ref. 1.

$\tilde{T}$ is equal to T, if only scalar variables are involved. For field strength operators the generalization from electrodynamics is straightforward, i.e.,

$$\tilde{T}(F_{\mu\nu}^a(x),\ F_{\rho\sigma}^b(y)) = T(F_{\mu\nu}^a(x),\ F_{\rho\sigma}^b(y))$$
$$+ i\delta_{ab}(g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)(g_\rho^0 g_\sigma^j - g_\rho^j g_\sigma^0)\delta_{ij}\, \delta(x - y). \tag{3.2}$$

This is to be generalized to products of several field strength operators as follows,

$$\tilde{T}(F_{\mu\nu}^a(x) \cdots F_{\rho\sigma}^b(y) \circ \cdots) = \tilde{T}(\hat{F}_{\mu\nu}^a(x) \cdots F_{\rho\sigma}^b(y) \cdots)$$
$$+ i\, \delta_{ab}(g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)(g_\rho^0 g_\sigma^j - g_\rho^j g_\sigma^0)$$
$$\times \delta_{ij}\delta(x - y)\, \tilde{T}(\cdots) + \cdots, \tag{3.3}$$

where the symbol $\hat{F}_{\mu\nu}^a$ means that the denoted operator is not to be included in $\tilde{T}$ ordering.

Next we consider the inclusion of one potential operator A_μ^a in $\tilde{T}$ ordering. For this we calculate with the

help of (2.13)

$$\partial'_\rho T(F_{\mu\nu}^a(z),\ A_\sigma^b(z')) - \partial'_\sigma T(F_{\mu\nu}^a(z),\ A_\rho^b(z')) - g\epsilon_{bcd} \times T(F_{\mu\nu}^a(z),\ A_\rho^c(z'),\ A_\sigma^d(z'))$$
$$= T(F_{\mu\nu}^a(z),\ F_{\rho\sigma}^b(z'))$$
$$+ i(g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)(g_\rho^0 g_\sigma^j - g_\rho^j g_\sigma^0)\{\delta_{ab}\delta_{ij}\,\delta(\mathbf{z}-\mathbf{z}')$$
$$- [\delta_{bc}\partial'_j - g\epsilon_{bdc}A_j^d(z')]\,\Gamma_i^{ac}(\mathbf{z},\mathbf{z}')\}\,\delta(z^0 - z'^0).$$

We now see that with the definition

$$\tilde{T}(F_{\mu\nu}^a(z),\ A_\sigma^b(z')) = T(F_{\mu\nu}^a(z),\ A_\sigma^b(z')) - i(g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)\,g_\sigma^0\,\Gamma_i^{ab}(\mathbf{z},\mathbf{z}')\delta(z^0 - z'^0), \tag{3.4}$$

the connection (1.3b) between field strengths and potentials is also true for $\tilde{T}$ products,

$$\partial'_\rho \tilde{T}(F_{\mu\nu}^a(z),\ A_\sigma^b(z')) - \partial'_\sigma \tilde{T}(F_{\mu\nu}^a(z),\ A_\rho^b(z')) - g\epsilon_{bcd}\,\tilde{T}(F_{\mu\nu}^a(z),\ A_\rho^c(z'),\ A_\sigma^d(z')) = \tilde{T}(F_{\mu\nu}^a(z),\ F_{\rho\sigma}^b(z')). \tag{3.5}$$

The extension of (3.4) to products with more than one field strength operator is obvious and is given by the recurrence relation

$$\tilde{T}(A_\rho^b(y),\ F_{\mu\nu}^a(z)\cdots) = \tilde{T}(\hat{A}_\rho^b(y),\ F_{\mu\nu}^a(z)\cdots) - i(g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)g_\rho^0\,\tilde{T}(\hat{\Gamma}_i^{ab}(\mathbf{z},\mathbf{y}) \times \delta(z^0 - y^0),\ \cdots) + \cdots. \tag{3.6}$$

Now we are able to define the gauge-covariant derivative of $\tilde{T}$ products,

$$\tilde{\nabla}_\mu^{ab}(x)\,\tilde{T}(\phi^b(x)\cdots,\ F_{\rho\sigma}^d(z)\cdots) = \partial_\mu \tilde{T}(\phi^a(x)\cdots,\ F_{\rho\sigma}^d(z)\cdots) - g\epsilon_{acb}\tilde{T}(A_\mu^c(x),\ \phi^b(x)\cdots, \times F_{\rho\sigma}^d(z)\cdots). \tag{3.7}$$

We have written the gauge-covariant derivative of a scalar variable. In this case the derivation results in

$$\tilde{\nabla}_\mu^{ab}(x)\,\tilde{T}(\phi^b(x)\cdots,\ F_{\rho\sigma}^d(z)\cdots) = \tilde{T}([\delta_{ab}\partial_\mu - g\epsilon_{acb}\hat{A}_\mu^c(x)]\ \phi^b(x),\ \cdots,\ F_{\rho\sigma}^d(z)\cdots) = \tilde{T}(\hat{\nabla}_\mu^{ab}(x)\phi^b(x),\ \cdots,\ F_{\rho\sigma}^d(z)\cdots) \tag{3.8}$$

because the extra terms for the potential are cancelled by the commutators that are produced by the time differentiation. This is analogous to electrodynamics.[1] Repeating the calculation of Ref. 1 gives the field equation (1.3) for $\tilde{T}$ products,

$$[\tilde{\nabla}_\mu^{ac}(x)\,\tilde{\nabla}_\nu^{cb}(x) - \tilde{\nabla}_\nu^{ac}(x)\,\tilde{\nabla}_\mu^{cb}(x)]\,\tilde{T}(\phi^b(x)\cdots,\ F_{\rho\sigma}^d(z)\cdots) = -g\epsilon_{acb}\,\tilde{T}(F_{\mu\nu}^c(x),\ \phi^b(x)\cdots,\ F_{\rho\sigma}^d(z)\cdots). \tag{3.9}$$

That (3.9) is true also for gauge covariant derivations of fields strength variables is seen by a straightforward but a little more tedious algebraic calculation. We also can verify (the details of the calculations are omitted in the following) that Eq. (1.4) holds for $\tilde{T}$ products without extra terms,

$$\tilde{\nabla}_{[\rho}^{ab}(z)\,\tilde{T}(F_{\sigma\lambda]}^b(z),\ F_{\mu\nu}^c(z')) = \tilde{T}(\hat{\nabla}_{[\rho}^{ab}(z)\,\hat{F}_{\sigma\lambda]}^b(z),\ F_{\mu\nu}^c(z')) = 0. \tag{3.10}$$

Here $[\rho\sigma\lambda]$ means cyclic permutation of indices. We have assumed (1.4) to hold also in the quantized theory. Equation (3.10) is true for an arbitrary $\tilde{T}$ product.

The field equation (2.2) is more difficult for $\tilde{T}$-products. For the $\tilde{T}$-product of two field strength operators we find

$$\tilde{\nabla}_{ab}^\mu(z)\,\tilde{T}(F_{\mu\nu}^b(z),\ F_{\rho\sigma}^d(z')) = \tilde{T}(\hat{\nabla}_{ab}^\mu(z)\,\hat{F}_{\mu\nu}^b(z),\ F_{\rho\sigma}^d(z')) + i[g_{\nu\rho}\nabla_\sigma^{ad}(z) - g_{\nu\sigma}\nabla_\rho^{ad}(z)]\,\delta(z-z') - igg_\nu^i\Gamma_i^{ac}(\mathbf{z},\mathbf{z}')\,\delta(z^0 - z'^0)\circ\epsilon_{cde}F_{\rho\sigma}^e(z') + ig\epsilon_{abc}\,\tilde{T}(g_\nu^i\Gamma_i^{cb}(\mathbf{z},\mathbf{z})\,\delta(z^0 - z^0),\ F_{\mu\nu}^d(z')). \tag{3.11}'$$

If we anticipate that in future formulas Γ_i will always appear in the combination

$$\Gamma_\nu^{ab}(x,y) := g_\nu^i\Gamma_i^{ab}(\mathbf{x},\mathbf{y})\,\delta(x^0 - y^0), \tag{3.12}$$

which is a particular solution of

$$\nabla_{ab}^\nu(x)\,\Gamma_\nu^{bc}(x,y) = \delta_{ac}\,\delta(x-y), \tag{3.13}$$

we can write (3.11)′ in the more compact form

$$\tilde{\nabla}_{ab}^\mu(z)\,\tilde{T}(F_{\mu\nu}^b(z),\ F_{\rho\sigma}^d(z')) = \tilde{T}(\hat{\nabla}_{ab}^\mu(z)\,\hat{F}_{\mu\nu}^b(z),\ F_{\rho\sigma}^d(z')) + i[g_{\nu\rho}\nabla_\sigma^{ad}(z) - g_{\nu\sigma}\nabla_\rho^{ad}(z)]\,\delta(z-z') - ig\,\Gamma_\nu^{ac}(z,z')\circ\epsilon_{cde}\,F_{\rho\sigma}^e(z') + g\,\tilde{T}(\mathrm{tr}[T^a\Gamma_\nu(z,z)],\ F_{\mu\nu}^d(z')).$$

The generalization of this to an arbitrary $\tilde{T}$-product is given by

$$\tilde{\nabla}_{ab}(z)\,\tilde{T}(\phi^e(x)\cdots,\ F_{\mu\nu}^b(z)\cdots F_{\rho\sigma}^d(z')\cdots) = \tilde{T}(\phi^e(x)\cdots,\ \hat{\nabla}_{ab}^\mu(z)\,\hat{F}_{\mu\nu}^b(z)\cdots F_{\rho\sigma}^d(z')\cdots) + i[g_{\nu\rho}\tilde{\nabla}_\sigma^{ad}(z) - g_{\nu\sigma}\tilde{\nabla}_\rho^{ad}(z)]\,\delta(z-z')\,\tilde{T}(\phi^e(x)\cdots,\ \cdots) - ig\,\tilde{T}(\phi^e(x)\cdots,\ \Gamma_\nu^{ac}(z,z')\circ\epsilon_{cdf}F_{\rho\sigma}^f(z')\cdots) + \cdots - ig\,\tilde{T}(\cdots,\ \Gamma_\nu^{ac}(z,x)\epsilon_{cef}\phi^f(x)\cdots F_{\rho\sigma}^d(z')\cdots) + \cdots + g\,\tilde{T}(\phi^e(x)\cdots,\ \mathrm{tr}[T^a\Gamma_\nu(z,z)]\cdots F_{\rho\sigma}^d(z')\cdots), \tag{3.11}$$

as can be seen by an easy but lengthy calculation. We observe that in addition to the first, the second, and the fourth term, which appear in the corresponding equations in electrodynamics, there are two additional terms. The third term shows that the field $F_{\mu\nu}^a$ itself carries isospin. The last term will later be seen to give rise to the additional terms in the perturbation series found by Feynman,[4] De Witt,[5] and Faddeev and Popov.[6] In our formalism the trace term arises from the fact,

that $F_{\mu\nu}^{a}$ obeys a field equation with covariant derivatives. To include the potential term of this derivative in $\tilde{T}$ ordering, we have to add for the pairing $A_c^{\mu}(z)$, $F_{\mu\nu}^{b}(z)$ just this trace term. It is not cancelled by a contribution from the commutators that are produced in the course of differentiation.

Next we want to consider equations for $\tilde{T}$ products that contain more than one o-component of the potential. To obtain the extra term that has to be added for two potential operators, we regard the equation

$$\begin{aligned}&\partial_\mu T(A_\nu^a(x),\ A_\rho^d(y)) - \partial_\nu T(A_\mu^a(x),\ A_\rho^d(y))\\ &\quad - g\epsilon_{abc}\, T(A_\mu^b(x), A_\nu^c(x),\ A_\rho^d(y))\\ &\qquad = T(F_{\mu\nu}^a(x),\ A_\rho^d(y)) + (g_\mu^0 g_\nu^i - g_\mu^i g_\nu^0)\\ &\qquad\quad \times g_\rho^0 [A_i^a(x),\ A_0^d(y)]\,\delta(x^0 - y^0).\end{aligned} \tag{3.14}$$

This can only be calculated if we make special assumptions on the gauge (which is not necessary up to this point). If we restrict ourselves to linear gauges as discussed in Sec. 2, we can use (2.26a) to evaluate the commutator. Because of (3.4) we have to define

$$\begin{aligned}\tilde{T}(A_\mu^a(x),\ A_\rho^d(y)) &= T(A_\mu^a(x),\ A_\rho^d(y))\\ &\quad + ig_\mu^0 g_\rho^0 \int d\tau(\mathbf{z})\, \Gamma_\tau^{ca}(\mathbf{z},\mathbf{x})\, \Gamma_\tau^{cd}(\mathbf{z},\mathbf{y})\, \delta(x^0 - y^0).\end{aligned} \tag{3.15}$$

This is in agreement with the relation

$$\begin{aligned}&\partial_\mu \tilde{T}(A_\nu^a(x),\ A_\rho^d(y)) - \partial_\nu \tilde{T}(A_\mu^a(x),\ A_\rho^d(y)) - g\epsilon_{abc}\\ &\quad \times \tilde{T}(A_\mu^b(x),\ A_\nu^c(x),\ A_\rho^d(y))\\ &\qquad = \tilde{T}(F_{\mu\nu}^a(x),\ A_\rho^d(y)).\end{aligned} \tag{3.16}$$

We again indicate the generalization to $\tilde{T}$ products of several potential operators:

$$\begin{aligned}\tilde{T}(A_\mu^a(x)\cdots A_\rho^d(y)\cdots) &= \tilde{T}(\hat{A}_\mu^a(x)\cdots A_\rho^d(y)\cdots)\\ &\quad + ig_\mu^0 g_\rho^0 \tilde{T}\Big(\int d\tau(\mathbf{z})\, \hat{\Gamma}_\tau^{ca}(\mathbf{z},\mathbf{x})\, \hat{\Gamma}_\tau^{cd}(\mathbf{z},\mathbf{y})\\ &\quad \times \delta(x^0 - y^0)\cdots\Big) + \cdots.\end{aligned} \tag{3.17}$$

Now we can calculate the analog of the field equation (3.11) for $\tilde{T}$ products of potential operators. If there is only one potential, we get

$$\begin{aligned}\tilde{\nabla}_{ab}^{\mu}(z)\, \tilde{T}(F_{\mu\nu}^b(z),\ A_\rho^{a'}(z')) &= \tilde{T}(\tilde{\nabla}_{ab}^{\mu}(z)\, \hat{F}_{\mu\nu}^b(z),\ A_\rho^{a'}(z'))\\ &\quad + i[\delta_{aa'} g_{\nu\rho}\delta(z - z') + \Gamma_\nu^{ac}(z, z')\\ &\quad \times (\delta_{ca'} \overleftarrow{\partial}_\rho' + g\epsilon_{cda'} A_\rho^d(z'))]\\ &\quad + g\tilde{T}(\mathrm{tr}[T^a \Gamma_\nu(z,z)],\ A_\rho^{a'}(z')),\end{aligned} \tag{3.18}$$

where Γ_ν is defined by (3.12). Again the additional trace term appears. The generalization of this formula to an arbitrary $\tilde{T}$ product is given by

$$\begin{aligned}&\tilde{\nabla}_{ac}^{\mu}(z)\, \tilde{T}(\phi^b(x)\cdots, F_{\mu\nu}^c(z)\cdots A_\rho^{a'}(z')\cdots)\\ &\quad = i(\delta_{aa'} g_{\nu\rho}\,\delta(z - z')\, \tilde{T}(\phi^b(x)\cdots,\ \cdots) + \tilde{\nabla}_\rho^{a'c}(z')\\ &\qquad \times \tilde{T}(\phi^b(x)\cdots, \Gamma_\nu^{ac}(z,z')\cdots)) + \cdots\\ &\qquad + g\,\tilde{T}(\phi^b(x)\cdots,\ \mathrm{tr}[T^a\Gamma_\nu(z,z)]\cdots A_\rho^{a'}(z')\cdots)\\ &\qquad - ig\,\tilde{T}(\cdots, \Gamma_\nu^{ac}(z,x)\,\epsilon_{cbd}\,\phi^d(x)\cdots A_\rho^{a'}(z')\cdots) + \cdots\\ &\qquad + \tilde{T}(\phi^b(x)\cdots,\ \hat{\nabla}_{ac}^{\mu}(z)\,\hat{F}_{\mu\nu}^c(z)\cdots A_\rho^{a'}(z')\cdots).\end{aligned} \tag{3.19}$$

In proving this, one must be careful to keep track of all additional terms that arise from differentiation. The term $\Gamma_\nu^{aa'}(z,z')\overleftarrow{\partial}_\rho'$ is first obtained inside a T product and the derivative has to be taken out of the T product. This gives additional equal time-commutators that can be calculated by means of (2.26b) and add up with the other terms to perfectly defined $\tilde{T}$ products. Doing so, one has to take into account that the extra terms (3.15) of $\tilde{T}$ products are operators. When differentiating on the left-hand side of (3.19) we get commutators of F_{0i}^a with the extra terms required by $\tilde{T}$ ordering. These prove to be necessary to cancel with other terms from the equal time commutators.

If we use field equation (2.2a), the last term can alternatively be written in the form

$$\begin{aligned}&\tilde{T}(\phi^b(x)\cdots,\ \hat{\nabla}_{ac}^{\mu}(z)\,\hat{F}_{\mu\nu}^c(z)\cdots A_\rho^{a'}(z')\cdots)\\ &\quad = g\,\tilde{T}(\phi^b(x)\cdots,\ [\hat{\nabla}_\nu^{cd}(z)\,\phi^d(z)]\,\epsilon_{cae}\phi^e(z)\cdots A_\rho^{a'}(z')\cdots).\end{aligned} \tag{3.20}$$

This may by expressed as the covariant derivative of a $\tilde{T}$-ordered product in view of Eq. (3.7),

$$\begin{aligned}&\tilde{T}(\phi^b(x)\cdots,\ j_\nu^a(z)\cdots A_\rho^{a'}(z')\cdots)\\ &\quad = g\tilde{\nabla}_\nu^{cd}(z)\,\tilde{T}(\phi^b(x)\cdots,\ [\phi^d(z)\epsilon_{cae}\phi^e(w)]\cdots\\ &\qquad \times A_\rho^{a'}(z')\cdots)\big|_{z=w}.\end{aligned} \tag{3.21}$$

(3.19) and (3.21) are covariant field equations for $\tilde{T}$-ordered products of potentials.

Finally we derive the field equations obeyed by the scalar field. We have to calculate

$$\tilde{\nabla}_{bc}^{\mu}(x)\,\tilde{\nabla}_\mu^{cd}(x)\,\tilde{T}(\phi^d(x)\cdots\phi^{b'}(x')\cdots, A_\mu^a(z)\cdots). \tag{3.22}$$

Here the only problem is the evaluation of the time derivative. With the help of the commutation relations (2.26c) and (2.3) we obtain

$$\begin{aligned}&\tilde{\nabla}_{bc}^0(x)\,\tilde{\nabla}_0^{cd}(x)\,\tilde{T}(\phi^d(x)\cdots\phi^{b'}(x')\cdots,\ A_\mu^a(z)\cdots)\\ &\quad = \tilde{\nabla}_{bc}^0(x)\,\tilde{T}(\hat{\nabla}_0^{cd}(x)\,\phi^d(x)\cdots\phi^{b'}(x')\cdots,\ A_\mu^a(z)\cdots)\\ &\quad = \tilde{T}(\hat{\nabla}_{bc}^0(x)\,\hat{\nabla}_0^{cd}(x)\,\phi^d(x)\cdots\phi^{b'}(x')\cdots, A_\mu^a(z)\cdots)\\ &\qquad - i\,\delta_{bb'}\,\delta(x - x')\,\tilde{T}(\cdots,\ A_\mu^a(z)\cdots) + \cdots.\end{aligned} \tag{3.23}$$

The commutator (2.23c) and its analog with ϕ^b replaced by $(\nabla_0\phi)^b$ are needed to cancel the additional terms from including the potential terms of the covariant derivatives in $\tilde{T}$ ordering. The last term is due to (2.3). We then have the field equation

$$\begin{aligned}&\tilde{\nabla}_{bc}^{\mu}(x)\,\tilde{\nabla}_\mu^{cd}(x)\,\tilde{T}(\phi^d(x)\cdots\phi^{b'}(x')\cdots,\ A_\mu^a(z)\cdots)\\ &\quad = -m^2\tilde{T}(\phi^b(x)\cdots\phi^{b'}(x')\cdots,\ A_\mu^a(z)\cdots)\\ &\qquad - i\,\delta_{bb'}\,\delta(x - x')\,\tilde{T}(\cdots, A_\mu^a(z)\cdots) + \cdots,\end{aligned} \tag{3.24}$$

where we made use of Eq. (2.2b).

We want to remark that the scalar field is, of course, not essential for our treatment. If we drop all terms that refer to scalar variables, we get the equations for a pure Yang–Mills field including all the complications we are interested in.

4. GREEN'S FUNCTIONS

The (disconnected and unrenormalized) Green's functions are defined as the vacuum expectation values of $\tilde{T}$-ordered products

$$G^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots) = \langle 0 \,|\, \tilde{T}(\phi^b(x)\cdots, A^a_\mu(z)\cdots)\,|\,0\rangle. \tag{4.1}$$

Differential equations for these Green's functions follow from the field equations (3.9), (3.10), (3.19), and (3.24). It is convenient to use the condensed notation introduced by Mandelstam.[3] We shall describe it briefly.

Consider the set of functions

$$F^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots).$$

These functions can be regarded as bilinear forms over $V^*\times V$, where V is a suitable vector space and V^* its dual. More explicitly we write

$$F^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots) =: (e^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots)\,|\,F), \tag{4.2}$$

where $(\ |\ )$ is the canonical bilinear form over $V^*\times V$, F is an element of V, and $e^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots)$ is a basis element of V^*. In this scheme the set of Green's functions (4.1) is determined by a specific vector $G\in V$. G has to obey vector equations which we shall now consider.

We define operators $\tilde{\phi}^b(x)$, $\tilde{A}^a_\mu(z)$ by their action on the basis vectors of V^* as follows:

$$e^{b'\cdots,\,a\cdots}_{\mu\cdots}(x'\cdots, z\cdots)\,\tilde{\phi}^b(x) = e^{b'\cdots b,\,a\cdots}_{\mu\cdots}(x'\cdots x, z\cdots), \tag{4.3a}$$

$$e^{b\cdots,\,a'\cdots}_{\mu\cdots}(x\cdots, z'\cdots)\,\tilde{A}^a_\nu(z) = e^{b\cdots,\,a'\cdots\,a}_{\mu\cdots\,\nu}(x\cdots, z'\cdots z). \tag{4.3b}$$

The tilde distinguishes these operators from quantum mechanical operators. Their action on vectors $F\in V$ is defined by transpostion, i.e.,

$$(e^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots)\,\tilde{\phi}^{b'}(x')\,|\,F) =: (e^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots)\,|\,\tilde{\phi}^{b'}(x')F). \tag{4.4}$$

Another set of operators $\eta^b(x)$, $Z^a_\mu(z)$ is needed to produce the source terms of the field equations. It is defined by the following commutation rules:

$$[\eta^b(x), \tilde{\phi}^{b'}(x')] = i\,\delta_{bb'}\,\delta(x-x'), \quad [\eta^b(x), \tilde{A}^a_\mu(z)] = 0, \tag{4.5a}$$

$$[Z^a_\mu(z), \tilde{A}^{a'}_\nu(z')] = -i\{\delta_{aa'}\,g_{\mu\nu}\,\delta(z-z') + \tilde{\Gamma}^{ac}_\mu(z,z')\,[\delta_{ca'}\overleftarrow{\partial}'_\nu + g\,\epsilon_{cda'}\,\tilde{A}^d_\nu(z')]\}, \tag{4.5b}$$

$$[Z^a_\mu(z), \tilde{\phi}^b(x)] = ig\,\tilde{\Gamma}^{ac}_\mu(z,x)\,\epsilon_{cbd}\,\tilde{\phi}^d(x). \tag{4.5c}$$

Here for linear gauges $\tilde{\Gamma}^{ab}_\mu(x,y)$ is defined by the equation

$$\tilde{\Gamma}^{ab}_\mu(x,z) = g^i_\mu \int d\tau(\mathbf{w})\,\gamma^{ac}_i(\mathbf{x},\mathbf{w}) \times[(1\partial^I - ig\mathbf{T}\,\tilde{\mathbf{A}}^I)\gamma_I]^{-1}\ (\mathbf{w},\mathbf{z})^{cb}\,\delta(x^0-z^0), \tag{4.6}$$

as may be seen from (2.22) and (3.12). In addition to (4.5) we have to require

$$e_0\eta^b(x) = 0, \quad e_0 Z^a_\mu(z) = 0, \tag{4.7}$$

where e_0 is a basis element of the one-dimensional subspace of V^*,

$$(e_0\,|\,F)\in\mathbb{C}.$$

We also postulate for the Green's vector G,

$$(e_0\,|\,G) = 1. \tag{4.8}$$

We now formulate the field equations. First we define the gauge-covariant derivation of basis elements

$$\tilde{\nabla}^{bc}_\mu(x)\,e^{c\cdots,\,a\cdots}_{\nu\cdots}(x\cdots, z\cdots) := \frac{\partial}{\partial x^\mu}\,e^{b\cdots,\,a\cdots}_{\nu\cdots}(x\cdots, z\cdots) - g\,\epsilon_{bdc}\,e^{c\cdots,\,a\cdots}_{\nu\cdots}(x\cdots, z\cdots)\tilde{A}^d_\mu(x). \tag{4.9}$$

Because of (4.3a) this is equivalent to

$$\tilde{\nabla}^{bc}_\mu(x)\,\tilde{\phi}^c(x) = [\delta_{bc}\partial_\mu - g\,\epsilon_{bdc}\tilde{A}^d_\mu(x)]\,\tilde{\phi}^c(x). \tag{4.10}$$

The field equation (3.9) is then represented by

$$[\tilde{\nabla}^{bc}_\mu(x)\,\tilde{\nabla}^{cd}_\nu(x) - \tilde{\nabla}^{bc}_\nu(x)\,\tilde{\nabla}^{cd}_\mu(x)]\,\tilde{\phi}^d(x) = -g\epsilon_{bcd}\,\tilde{F}^c_{\mu\nu}(x)\phi^d(x), \tag{4.11}$$

where (3.5) or (3.16) imply

$$\tilde{F}^a_{\mu\nu}(x) = \partial_\mu\tilde{A}^a(x) - \partial_\nu\tilde{A}^a_\mu(x) - g\,\epsilon_{acd}\tilde{A}^c_\mu(x)\tilde{A}^d_\nu(x). \tag{4.12}$$

This shows that (3.10) is represented by an operator identity on V^*, i.e.,

$$\nabla^{ac}_{[\mu}(z)\,\tilde{F}^c_{\nu\rho]}(z) = 0. \tag{4.13}$$

The relations for the Green's functions, implied by (3.19) and (3.24) are more complicated. They give rise to the following equations for the Green's vector:

$$[\tilde{\nabla}^\mu_{bc}(x)\,\tilde{\nabla}^{cd}_\mu(x)\,\tilde{\phi}^d(x) + m^2\tilde{\phi}^b(x) - \eta^b(x)]G = 0, \tag{4.14a}$$

$$\{\tilde{\nabla}^\mu_{ac}(z)\,\tilde{F}^c_{\mu\nu}(z) - g\,\mathrm{tr}[T^a\tilde{\Gamma}_\nu(z,z)] - \tilde{j}^a_\nu(z) - Z^a_\nu(z)\}G = 0. \tag{4.14b}$$

$\tilde{\nabla}_\mu$ and $\tilde{F}_{\mu\nu}$ have been defined above and, in agreement with (3.20), $\tilde{j}^a_\nu(z)$ is given by

$$\tilde{j}^a_\nu(z) = g[\tilde{\nabla}^{cd}_\nu(z)\,\tilde{\phi}^d(z)]\,\epsilon_{cae}\,\tilde{\phi}^e(z). \tag{4.15}$$

The field equations for the Green's functions are recovered, if we multiply (4.14) by the basis elements

$$e^{b\cdots,\,a\cdots}_{\mu\cdots}(x\cdots, z\cdots) = e_0\tilde{\phi}^b(x)\cdots\tilde{A}^a_\mu(z)\cdots, \tag{4.16}$$

[here the multiplication is defined by the bilinear form $(\ |\ )$], and shift the operators η^b, Z^a_μ to the left by means of (4.5) until (4.7) can be applied.

We now consider (4.14b) in more detail. First we write it in the form

$$\{\Box \tilde{A}^a_\nu(z) - \partial^\mu \partial_\nu \tilde{A}^a_\mu(z) - \tilde{J}^a_\nu(z) - g\ \mathrm{tr}[T^a \Gamma_\nu(z,z)] - Z^a_\nu(z)\} G = 0, \tag{4.17}$$

where we have separated the linear term from $\tilde{\nabla}^\mu_{ac} \tilde{F}^c_\mu$, i.e.,

$$\tilde{\nabla}^\mu_{ac}(z)\, \tilde{F}_{\mu\nu}{}^c(z) = \Box \tilde{A}^a_\nu(z) - \partial^\mu \partial_\nu \tilde{A}^a_\mu(z) - \tilde{k}^a_\nu(z), \tag{4.18}$$

and have defined

$$\tilde{J}^a_\nu(z) = \tilde{k}^a_\nu(z) + \tilde{j}^a_\nu(z). \tag{4.19}$$

From (4.17) we deduce the identity

$$\{\partial^\nu \tilde{J}^a_\nu(z) + g\, \partial^\nu \mathrm{tr}[T^a \Gamma_\nu(z,z)] + \partial^\nu Z^a_\nu(z)\} G = 0. \tag{4.20}$$

We also can derive the following identities:

$$\{\tilde{\nabla}^\mu_{ac}(z)\, \tilde{j}^c_\mu(z) + g\, \tilde{\nabla}^\mu_{ac}(z)\, \mathrm{tr}[T^c \Gamma_\mu(z,z)] + \tilde{\nabla}^\mu_{ac}(z)\, Z^c_\mu(z)\} G = 0, \tag{4.21a}$$

$$[\tilde{\nabla}^\mu_{ac}(z)\, \tilde{j}^c_\mu(z) - g\eta^c(z)\epsilon_{cad}\tilde{\phi}^d(z)]\, G = 0, \tag{4.21b}$$

$$\{\tilde{\nabla}^\mu_{ac}(z)\, Z^c_\mu(z) + g\, \tilde{\nabla}^\mu_{ac}(z)\, \mathrm{tr}[T^c \tilde{\Gamma}_\mu(z,z)] + g\, \eta^c(z)\epsilon_{cad}\tilde{\phi}^d(z)\} = 0. \tag{4.21c}$$

Equation (4.21a) follows from (4.14b) and (4.12), Eq. (4.21b) from (4.14a). To prove the operator identity (4.21c) we have to show that the left hand side commutes with $\tilde{A}^a_\mu$ and $\tilde{\phi}^b$ and that it "annihilates" e_0. This follows from (4.5), (4.7), and the equations

$$\begin{aligned} g\tilde{\nabla}^\mu_{ac}(z)\, \mathrm{tr}[T^c \tilde{\Gamma}_\mu(z,z)] &= g\, \mathrm{tr}(T^a)\, \delta(z-z) \\ &\quad + ig\epsilon_{adc}\, \tilde{\nabla}^\mu_{de}(w)\, \tilde{\Gamma}^{ce}_\mu(z,w)\big|_{w=z} \\ &= ig\epsilon_{adc}\, \tilde{\nabla}^\mu_{de}(w)\, \tilde{\Gamma}^{ce}_\mu(z,w)\big|_{w=z}, \end{aligned} \tag{4.22}$$

$$\begin{aligned} \tilde{\nabla}^\mu_{ac}(z) Z^c_\mu(z) &= Z^c_\mu(z)[\delta_{ca}\overleftarrow{\partial}^\mu + g\epsilon_{cda}\tilde{A}^\mu_d(z)] \\ &\quad - ig\epsilon_{adc}\tilde{\nabla}^\mu_{de}(w)\tilde{\Gamma}^{ce}_\mu(z,w)\big|_{w=z}. \end{aligned} \tag{4.23}$$

In (4.22) and (4.23) we have to regard $\mathrm{tr}(T^a)\, \delta(z-z) = 0 \cdot \delta(z-z)$ as zero.

To write (4.17) in a form that can easily be integrated we introduce another operator $\xi^a_\mu(z)$ that is defined by the relations

$$\begin{aligned} &[\xi^a_\mu(z),\ \tilde{A}^{a'}_\nu(z')] = -i\delta_{aa'} g_{\mu\nu}\, \delta(z-z'), \\ &[\xi^a_\mu(z),\ \tilde{\phi}^b(x)] = 0, \quad e_0 \xi^a_\mu(z) = 0. \end{aligned} \tag{4.24}$$

We can express Z^a_μ in terms of ξ^a_μ as follows,

$$\begin{aligned} Z^a_\mu(z) &= \xi^a_\mu(z) - \int d^4 w [\tilde{\nabla}^\lambda_{cd}(w)\, \xi^d_\lambda(w) \\ &\quad + g\eta^d(w)\epsilon_{dce}\tilde{\phi}^e(w)]\tilde{\Gamma}^{ac}_\mu(z,w). \end{aligned} \tag{4.25}$$

It is easily checked that this expression fulfills the commutation relations (4.5). The operator equation (4.21c) follows from (4.24) and (4.22), and (4.7) is secured by the ordering of the operators in (4.25).

Next we define

$$\gamma^{ab}_\mu(x,y) = g^i_\mu \gamma^{ab}_i(\mathbf{x},\mathbf{y})\, \delta(x^0 - y^0). \tag{4.26}$$

Hence

$$\partial^\mu \gamma^{ab}_\mu(x,y) = \delta_{ab}\delta(x-y), \tag{4.27a}$$

$$\tilde{\Gamma}^{ab}_\mu(x,y) = \int d^4 w \gamma^{ac}_\mu(x,w)(\tilde{\nabla}^\lambda \gamma_\lambda)^{-1}(w,y)^{cb}. \tag{4.27b}$$

From (4.27) we conclude

$$\tilde{\Gamma}^{ac}_\mu(z,w) = \int d^4 w' \gamma^{ad}_\mu(z,w')\, \partial^\lambda \tilde{\Gamma}^{dc}_\lambda(w',w), \tag{4.28}$$

where we stress that γ_μ is a c-number function. We now substitute (4.25) into Eq. (4.17). (4.20) and the representation (4.28) then can be used to find the following form of the field equation:

$$\begin{aligned} &[\Box \tilde{A}^a_\mu(x) - \partial^\nu \partial_\mu \tilde{A}^a_\nu(x)] G \\ &= \int d^4 w\ \delta_{ab} g^\lambda_\mu \left(\delta(z-w) + \frac{\partial}{\partial w_\lambda}\ \gamma^{ab}_\mu(z,w)\right) \\ &\quad \times \{\tilde{J}^b_\lambda(w) + g\ \mathrm{tr}[T^b \Gamma_\lambda(w,w)] + \zeta^b_\lambda(w)\} G. \end{aligned} \tag{4.29}$$

Integration of (4.29) with Feynman boundary conditions yields

$$\begin{aligned} \tilde{A}^a_\mu(z) G &= \int d^4 w \int d^4 w' D_F(z-w) \\ &\quad \times \left(\delta_{ac} g^\lambda_\mu \delta(w-w') + \frac{\partial}{\partial w'_\lambda}\ \gamma^{ac}_\mu(w,w')\right) \times \{\tilde{J}^c_\lambda(w') \\ &\quad + g\, \mathrm{tr}[T^c \tilde{\Gamma}_\lambda(w',w')] + \zeta^c_\lambda(w')\} G + \partial_\mu C^a(z) G, \end{aligned} \tag{4.30}$$

where

$$\Box D_F(z-w) = \delta(z-w). \tag{4.31}$$

The undetermined gradient term can be fixed by the gauge condition (2.20). The latter can be written by means of (4.26) as

$$\int d^4 w \tilde{A}^\lambda_c(w) \gamma^{ca}_\lambda(w,z) = B^a(z). \tag{4.32}$$

To simplify, we set $B^a(z) = 0$ and obtain

$$\begin{aligned} \tilde{A}^a_\mu(z) G &= \int d^4 z' D^{ab}_{\mu\nu}(z,z'|\gamma)\, \{\tilde{J}^\nu_b(z') \\ &\quad + g\, \mathrm{tr}[T^b \tilde{\Gamma}^\nu(z',z')] + \zeta^\nu_b(z')\} G, \end{aligned} \tag{4.33}$$

where $D^{ab}_{\mu\nu}(z,z'|\gamma)$ is the gauge field propagator in the gauge defined by γ^{ab}_ν and is given explicitly by

$$\begin{aligned} D^{ab}_{\mu\nu}(z,z'|\gamma) = \int d^4 w \int d^4 w' &\left(\delta_{ac} g^\sigma_\mu \delta(z-w) \right. \\ &\left. + \frac{\partial}{\partial z^\mu}\ \gamma^\sigma_{ca}(w,z)\right) \\ &\times D_F(w-w') \\ &\times \left(\delta_{cb} g_{\sigma\nu}\, \delta(w'-z') + \frac{\partial}{\partial z'^\nu}\ \gamma^{cb}_\sigma(w',z')\right). \end{aligned} \tag{4.34}$$

The integration of the scalar field equations is simple. If we write (4.14a) as

$$[(\Box + m^2)\tilde{\phi}^b(x) - \tilde{H}^b(x) - \eta^b(x)]\, G = 0, \tag{4.35}$$

where $\tilde{H}^b(x)$ denotes the terms of (4.14a) that contain the $\tilde{A}^a_\mu$ operators, we get the integrated form

$$\tilde{\phi}^b(x) G = \int d^4 x' \Delta_F(x-x')[\tilde{H}^b(x') + \eta^b(x')] G. \tag{4.36}$$

Here, of course, Δ_F is the solution under Feynman boundary conditions of the equation

$$(\Box + m^2)\Delta_F(x - x') = \delta(x - x'). \tag{4.37}$$

From the integrated equations (4.33), (4.36) we can recover the Feynman rules for our system. These rules coincide with the "naive" Feynman rules except for the contribution of the trace term in (4.33). To obtain a more explicit version of the latter contribution we expand the inverse operator in the representation (4.27b) into a power series,

$$\begin{aligned}(\tilde{\nabla}^\lambda\gamma_\lambda)^{-1}(x,y)^{ab} &= (I - igT^c\tilde{A}^\lambda_c\gamma_\lambda)^{-1}(x,y)^{ab} \\ &= \left[\delta_{ab}\delta(x-y) + \sum_{n=0}^{\infty}\int d^4w_1 \cdots \int d^4w_n g\epsilon_{acd}\tilde{A}^\lambda_c(x)\right. \\ &\quad \left.\times\gamma^{de}_\lambda(x,w_1)\times\cdots\times g\,\epsilon_{fgh}\tilde{A}^\rho_g(w_n)\,\gamma^{hb}_\rho(w_n,y)\right].\end{aligned} \tag{4.38}$$

The corresponding expansion of the trace term reads

$$\begin{aligned}&g\ \mathrm{tr}[T^a\tilde{\Gamma}_\mu(z,z)] \\ &= -i\left[\sum_{n=0}^{\infty}\int d^4w_1 \cdots \int d^4w_n g\,\epsilon_{bac}\gamma^{cd}_\mu(z,w_1)g\,\epsilon_{def}\tilde{A}^\lambda_e(w_1)\right. \\ &\quad \left.\times\gamma^{fg}_\lambda(w_1,w_2)\times\cdots\times g\,\epsilon_{hrs}\tilde{A}^\rho_r(w_n)\,\gamma^{sb}_\rho(w_n,z)\right].\end{aligned} \tag{4.39}$$

It is seen that the correct Feynman rules require the inclusion of additional terms, which can be visualized as closed loops built from asymmetrical "propagators" $\gamma^{ca}_\nu(z,z')$ and vertices $g\epsilon_{abc}$. Here the index a is contracted with the second, the index c with the first isospin index of a γ_ν "propagator" while b is contracted with the isospin index of a propagator $iD^{db}_{\mu\nu}(x,z)$. The Lorentz index of the latter is contracted with the Lorentz index of the propagator γ_ν. In addition these loops get an overall factor of -1.

These terms have been found by Feynman,[4] De Witt,[5] Faddeev and Popov,[6] Mandelstam,[3] and others. They are usually referred to as loops of "ghost particles."

The axial gauge defined in Sec. 2 is particularly interesting. We see from Eqs. (2.29), (2.30) that in this case

$$\begin{aligned}\mathrm{tr}[T^a\tilde{\Gamma}_\mu(z,z)_A] &= \mathrm{tr}[T^a\gamma_\mu(z,z)_A] \\ &= -\mathrm{tr}(T^a)g^i_\mu n_i\int_{-\infty}^{0} ds\ \delta(\mathbf{z}-\mathbf{z}-\mathbf{n}s)\ \delta(z_0 - z^0) \\ &= 0.\end{aligned} \tag{4.40}$$

Hence the ghost particles disappear in this gauge.

Finally we consider the generalization of our formulas to four-dimensional linear gauges. As our results are perfectly covariant nothing forbids us to leave our starting point: the Coulomb-like gauges with γ_μ defined by (4.26). If we admit arbitrary solutions of Eq. (4.27a) to construct the kernel $\tilde{\Gamma}_\mu$ according to (4.27b), we obtain general four-dimensional linear gauges. Examples are the Lorentz invariant solution

$$\gamma^{ab}_\mu(x,y) = \delta_{ab}\partial_\mu D_F(x-y) \tag{4.41}$$

of (4.27a), which corresponds to the Landau gauge, and the four-dimensional generalization

$$\gamma^{ab}_\mu(x,y) = -\delta_{ab}\int_{-\infty}^{0} d\xi_\mu\ \delta(x-y-\xi), \tag{4.42}$$

of the spatial path gauge considered in Sec. 2.

We end with the remark that our approach is completely analogous to electrodynamics.[1] The covariant definition of time ordering automatically gives the correct Eqs. (4.14) for the Green's functions. Due to the kernel Γ_μ these equations are gauge dependent. Mandelstam[3] eliminated this reference to a particular gauge by the introduction of path dependent variables. We prefer to formulate the theory in terms of the local concept of covariant differentiation. In this framework it is not possible to remove the gauge dependence of the equations.

ACKNOWLEDGMENTS

I am greatly indebted to Professor K. Meetz for numerous discussions and critical comments. Also I want to acknowledge financial support from the Studienstiftung des deutschen Volkes.

[1]K. Meetz, J. Math. Phys. **16**, 1765 (1975).
[2]S. Mandelstam, Ann. Phys. (N.Y.) **19**, 1 (1962).
[3]S. Mandelstam, Phys. Rev. **175**, 1580 (1968).
[4]R. P. Feynman, Acta Phys. Polon. **24**, 697 (1963).
[5]B. S. De Witt, Phys. Rev. **162**, 1195 (1967).
[6]L. D. Faddeev and V. N. Popov, Phys. Lett. B 25, 29 (1967).
[7]R. E. Peierls, Proc. R. Soc. Lond. 213, 143 (1952).
[8]J. Schwinger, Phys. Rev. **125**, 1043 (1961).
[9]R. L. Arnowitt and S. I. Fickler, Phys. Rev. **127**, 1821 (1962).
[10]J. Schwinger, Phys. Rev. **130**, 402 (1963).

Lie theory and separation of variables. 9. Orthogonal R-separable coordinate systems for the wave equation $\psi_{tt}-\Delta_2\psi=0$

E. G. Kalnins*, †

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, Québec, Canada

W. Miller, Jr.‡

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 12 February 1975)

A list of orthogonal coordinate systems which permit R-separation of the wave equation $\psi_{tt}-\Delta_2\psi=0$ is presented. All such coordinate systems whose coordinate curves are cyclides or their degenerate forms are given. In each case the coordinates and separation equations are computed. The two basis operators associated with each coordinate system are also presented as symmetric second order operators in the enveloping algebra of the conformal group O(3,2).

INTRODUCTION

In this article we complement the contents of our previous article[1] (hereafter referred to as I) by giving a detailed treatment of the orthogonal coordinate systems for which the two-dimensional wave equation

$$\partial_{tt}\psi=\Delta_2\psi \qquad (*)$$

admits an R-separable solution.[2] We recall that an R-separable solution of (*) can be written in the form $\exp[Q(\mu,\rho,\nu)]A(\mu)B(\rho)C(\nu)$. Here μ,ρ,ν are curvilinear coordinates and Q is a function such that either

$$\frac{\partial^2 Q}{\partial\lambda\partial\lambda'}\neq 0, \quad \lambda\neq\lambda', \quad \lambda,\lambda'=\mu,\rho,\nu,$$

for at least two distinct pairs λ,λ' or $Q=0$. The latter case is the familiar one of separation of variables. In searching for R-separable solutions of (*) we restrict our attention in this article to orthogonal curvilinear coordinate systems. These are systems of coordinates μ,ρ,ν such that the differential form

$$ds^2=dt^2-dx^2-dy^2 \qquad (0.1)$$

can be written

$$ds^2=F\,d\mu^2+G\,d\rho^2+H\,d\nu^2, \qquad (0.2)$$

with F, G, and H real functions of μ,ρ,ν. In a subsequent article we shall give a systematic treatment of the nonorthogonal systems for which (*) admits a separation of variables.

The methods necessary for systematically finding all such orthogonal R-separable coordinate systems have been developed in some detail in the book by Bôcher.[3] These methods can be readily adapted to the problem of interest in this article. There are however a number of new developments occurring in the case of (*). These developments stem from the fact that (*) is inherently more complicated than Laplace's equation

$$(\partial_{xx}+\partial_{yy}+\partial_{zz})\psi=0, \qquad (0.3)$$

which Bôcher treated in detail. The contents of the article are arranged as follows.

In Sec. I we give the basic ideas necessary to construct the coordinates which allow an R-separation of (*). This involves a treatment of pentaspherical space, relevant properties of cyclides, and the method of finding the pentaspherical coordinates (and hence the coordinates t,x,y) in terms of the various curvilinear coordinates. Enough detail is presented in this section so as to make the article reasonably self-contained. In Sec. II the connection between the wave equation (*) and pentaspherical coordinates is discussed.

Section III contains the classification of orthogonal R-separable coordinates of (*). In addition the separation equations are given and identified as much as possible. We also give the two symmetric second order operators whose eigenvalues are the separation constants. These operators are expressed in terms of the symmetry group of (*) discussed in detail in I.

The best-known coordinate systems which permit separation of variables in the wave, Laplace, and Helmholtz equations have the property that the coordinate surfaces are orthogonal families of confocal quadrics

$$\frac{x^2}{\lambda-a_1}+\frac{y^2}{\lambda-a_2}+\frac{z^2}{\lambda-a_3}=1, \quad a_i \text{ const} \qquad (0.4)$$

or their limits.[4] Thus the coordinate surfaces are ellipsoids, hyperboloids, spheres, planes, etc. The Helmholtz equation separates only in coordinate systems of this type, but the wave and Laplace equations admit more general separable systems. This fact is related to the greater symmetry of the latter differential equations. Indeed, the wave equation admits an inversion symmetry which transforms the coordinates x,y,t to $x/\mathbf{x}\cdot\mathbf{x}$, $y/\mathbf{x}\cdot\mathbf{x}$, $t/\mathbf{x}\cdot\mathbf{x}$, where $\mathbf{x}\cdot\mathbf{x}=t^2-x^2-y^2$. Under inversion and space–time translations the orthogonal coordinate surfaces (0.4) are transformed into orthogonal surfaces, each of the form

$$\alpha(t^2-x^2-y^2)^2+ax^2+by^2+ct^2+dx+ey+ft+h=0. \qquad (0.5)$$

The fourth-order surfaces (0.5) are cyclides[3,5,6] and the coordinate surfaces are orthogonal families of confocal cyclides. The set of all cyclides is invariant

under the conformal symmetry group of the wave equation. Moreover, one can show by explicit construction that certain confocal families of cyclides define orthogonal coordinate systems which permit separation of variables in the wave equation. No separable systems other than these are known. Two families of confocal cyclides define equivalent coordinate systems if one can be obtained from the other by a transformation belonging to the conformal symmetry group SO(3, 2) of the wave equation. Certain special families of cyclides can be mapped to the form (0.5) with $\alpha = 0$ by a conformal symmetry, and these families lead to the special coordinate surfaces (0.4) and their limits.

To determine all distinct cyclidic separable coordinate systems, we clearly need to classify the distinct equivalence classes of cyclides under the action of the conformal group.

However, as shown explicitly in I, the action of this group on x, y, t (Minkowski) space is rather complicated. To simplify the computation of equivalence classes, one sets up a correspondence between three-dimensional Minkowski space and five-dimensional pentaspherical space as defined in Sec. I. In pentaspherical space the general cyclide takes the simple form (1.9) and the action of the conformal group SO(3, 2) reduces to matrix multiplication. Thus the classification of cyclides into SO(3, 2) symmetry classes can be carried out in a straightforward manner, and the results mapped back to Minkowski space to yield R-separable coordinate systems for the wave equation.

I. PENTASPHERICAL COORDINATES AND ORTHOGONAL FAMILIES OF CONFOCAL CYCLIDES

In this section we will outline the use of pentaspherical coordinates in classifying orthogonal families of confocal cyclides. Such orthogonal families, each provide an R-separable coordinate system for (*). The results presented here summarize those aspects of the work of Bôcher that are relevant for this article. Further details can be found in Bôcher's book and also the book by Coolidge.[5]

Any set of objects that can be put into one to one correspondence with sets of five homogeneous coordinates $x_1 : x_2 : x_3 : x_4 : x_5$ not all simultaneously zero but connected by the relation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 0 \tag{1.1}$$

are called points in pentaspherical space. It is clear that in general the quantities x_i are complex numbers. For our purposes the subset of pentaspherical coordinates of interest for the wave equation (*) can be obtained from the coordinates t, x, y as follows. Instead of considering the usual Cartesian coordinates t, x, y in three-dimensional Minkowski space, consider the Cartesian coordinates defined by

$$Z = t, \quad X = ix, \quad Y = iy. \tag{1.2}$$

The correspondence between a point (t, x, y) in Minkowski space and a point in five-dimensional space is then achieved as follows. The stereographic projection of the Cartesian coordinates with respect to the four-dimensional unit sphere embeds the point (Z, X, Y) in a four-dimensional space. The homogeneous or projective coordinates of the corresponding four-vector are

$$\begin{aligned} &y_1 = r^2 - p^2 - q^2 + s^2, \quad y_2 = r^2 - p^2 - q^2 - s^2, \\ &y_3 = 2ips, \quad y_4 = 2iqs, \quad y_5 = 2rs, \end{aligned} \tag{1.3}$$

where the coordinates t, x, y are given by

$$t = r/s, \quad x = p/s, \quad y = q/s. \tag{1.4}$$

If we adopt entirely real coordinates by writing $z_i = y_i$, $i = 1, 2, 5$, and $z_i = -iy_i$, $i = 3, 4$, we see that these coordinates satisfy

$$z_1^2 - z_2^2 + z_3^2 + z_4^2 - z_5^2 = 0 \quad (z_i \text{ all real}). \tag{1.5}$$

The subset of pentaspherical space of interest then consists of those points whose pentaspherical coordinates are

$$\begin{aligned} &x_1 = i(r^2 - p^2 - q^2 + s^2), \quad x_2 = r^2 - p^2 - q^2 - s^2, \\ &x_3 = 2ips, \quad x_4 = 2iqs, \quad x_5 = 2rs. \end{aligned} \tag{1.6}$$

In this work we are concerned only with these points in pentaspherical space which correspond to the real coordinates z_i satisfying (1.5) (i.e., having the same signature as this equation). An alternative equation to (1.6) can be obtained via the substitutions $p \to -ip$, $q \to -iq$, $r \to -ir$. From the form of (1.1) it can be seen that to transform one set of pentaspherical coordinates x_i into another set x_i' via a linear transformation

$$x_i' = V_{ij} x_j, \tag{1.7}$$

which preserves

$$\Omega = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$$

is only possible if $V = (V_{ij})$ is an orthogonal matrix:

$$VV^T = 1 \quad [V^T = (V_{ji}), \ V_{ij} \in C]. \tag{1.8}$$

In particular for the case of interest here the orthogonal transformations V corresponding to points in pentaspherical space of the form (1.6) are isomorphic to elements of the group O(3, 2). This is the symmetry group of (*).

A cyclide is defined to be the locus of points x_i in pentaspherical space lying on the quadric surface

$$\Phi = \sum_{i,j=1}^{5} a_{ij} x_i x_i = 0 \tag{1.9}$$

with $a_{ij} = a_{ji}$ and $\det(a_{ij}) \neq 0$. The problem of classifying types of cyclides under the group of orthogonal transformations V as in (1.7) and (1.8) is then the problem of classifying the intersections of two quadric forms in five-dimensional projective space, where one form is required to be equivalent to Ω, (1.5). This is performed by the method of elementary divisors applied to the two quadratic forms.[6] If we take the quadratic forms to be Φ as in (1.9) and $\Omega = \sum_{i,j=1}^{5} b_{ij} x_i x_j$, each class of quadratic forms Φ, Ω is then specified by the corresponding invariant factors. The invariant factors form a complete set of invariants for each class of pairs Ω, Φ. This means that if Ω', Φ' have the same invariant factors as Ω, Φ, the two systems are related by a linear substitution

$$x_i' = c_{ij} x_j, \quad \det(c_{ij}) \neq 0.$$

The invariant factors of a given pair of quadratic forms are obtained as follows. Suppose $D=\det|\lambda a_{ij}-b_{ij}|$ contains the factor $(\lambda-u)^{l_0}$. A second index l_1 is defined to be the highest power of $(\lambda-u)$ which divides all the first minors of D. Proceeding in this manner we obtain the terminating set of indices $e_1=l_0-l_1$, $e_2=l_1-l_2,\ldots,$ $e_r=l_{r-1}$. The powers $(\lambda-u)^{e_1}$, $(\lambda-u)^{e_2},\ldots,(\lambda-u)^{e_r}$ are called the invariant factors to the base $\lambda-u$ of the determinant D of the family of forms. All possible invariant factors of D then determine a complete set of invariants. The standard notation for the inequivalent classes of pairs Ω, Φ of quadratic forms is to display the indices e_i for each of the roots of $D=0$ within a square bracket. Those indices belonging to the same base or root of $D=0$ are enclosed in conventional brackets. As an example consider the invariant factors $(\lambda-a)^2$, $(\lambda-b)$, $(\lambda-c)$, $(\lambda-d)$ the corresponding notation is [2111]. If the invariant factors are $(\lambda-a)^2$, $(\lambda-a)$, $(\lambda-c)$, $(\lambda-d)$, then there is more than one invariant factor to the base a. Such a cyclide is then called a degenerate form of the corresponding cyclide in which there is only one invariant factor to each different base. For this second example we have a degenerate case of the cyclide [2111] and write this as [(21)11]. If the set of invariant factors are $(\lambda-a)^2$, $(\lambda-b)$, $(\lambda-b)$, $(\lambda-c)$, then the notation would be [2(11)1] and so on. The list of pairs of quadratic forms in five variables which are inequivalent are (this does not include the singular cases, which we do not need here, see, for instance, Bromwich[6]):

1. [11111] $\Omega=x_1^2+x_2^2+x_3^2+x_4^2+x_5^2,$
$\Phi=\lambda_1x_1^2+\lambda_2x_2^2+\lambda_3x_3^2+\lambda_4x_4^2+\lambda_5x_5^2;$ (1.10)

2. [2111] $\Omega=2x_1x_2+x_3^2+x_4^2+x_5^2,$
$\Phi=2\lambda_1x_1x_2+x_1^2+\lambda_3x_3^2+\lambda_4x_4^2+\lambda_5x_5^2;$ (1.11)

3. [311] $\Omega=2x_1x_3+x_2^2+x_4^2+x_5^2,$
$\Phi=\lambda_1(2x_1x_3+x_2^2)+2x_1x_2+\lambda_4x_4^2+\lambda_5x_5^2;$ (1.12)

4. [221] $\Omega=2x_1x_2+2x_3x_4+x_5^2,$
$\Phi=2\lambda_1x_1x_2+x_1^2+2\lambda_3x_3x_4+x_3^2+\lambda_5x_5^2;$ (1.13)

5. [41] $\Omega=2x_1x_4+2x_2x_3+x_5^2,$
$\Phi=2\lambda_1(x_1x_4+x_2x_3)+2x_1x_3+x_2^2+\lambda_5x_5^2;$ (1.14)

6. [32] $\Omega=2x_1x_3+2x_4x_5+x_2^2,$
$\Phi=\lambda_1(2x_1x_3+x_2^2)+2x_1x_2+2\lambda_4x_4x_5+x_4^2;$ (1.15)

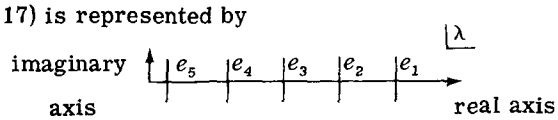
7. [5] $\Omega=2x_1x_5+2x_2x_4+x_3^2,$
$\Phi=\lambda_1(2x_1x_5+2x_2x_4+x_3^2)+2x_1x_4+2x_2x_3.$ (1.16)

The pairs of forms for a degenerate cyclide can be obtained from these formulas, e.g., the quadratic forms Ω, Φ corresponding to the configuration [(11)111] are obtained from (1.10) by putting $\lambda_1=\lambda_2$, and so on. Each type of cyclide is then associated with one of the seven types listed or one of the corresponding degenerate forms. The corresponding equations defining the cyclide are $\Omega=0$ and $\Phi=0$. The types of cyclides of particular interest here are those belonging to a confocal family. The most general such family is associated with the configuration [11111] and is given by the pair of equations,

$$\Omega=x_1^2+x_2^2+x_3^2+x_4^2+x_5^2=0,$$
$$\Phi=\frac{x_1^2}{\lambda-e_1}+\frac{x_2^2}{\lambda-e_2}+\frac{x_3^2}{\lambda-e_3}+\frac{x_4^2}{\lambda-e_4}+\frac{x_5^2}{\lambda-e_5}=0, \quad (1.17)$$

where λ is the parameter specifying the family. For the subset of pentaspherical space of interest to us Eqs. (1.17) may correspond to a number of different real nondegenerate coordinate curves in t,x,y space. These possibilities are,

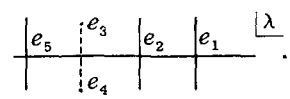
(i) The coordinates x_i are in fact the pentaspherical coordinates and are given by (1.6) [or the substitution $p\to -ip$, $q\to -iq$, $r\to -ir$ applied to (1.6)]. To give a real curve all the e_i must then be real. Bôcher introduces a diagramatic notation for such a confocal family of cyclides as follows. In the complex λ plane (1.17) is represented by



(ii) Two of the quantities e_i are mutually complex conjugate, say e_3, e_4. The corresponding choice of variables for x_i is

$$x_1=i(r^2-p^2-q^2+s^2),\quad x_2=r^2-p^2-q^2-s^2,$$
$$x_3=\sqrt{2}(r+ip)s,\quad x_4=\sqrt{2}(r-ip)s,\quad x_5=2iqs. \quad (1.18)$$

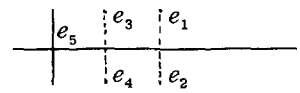
Another associated choice is obtained by taking $p\to -ip$, $q\to -iq$, $r\to -ir$ in these formulas. The notation for such a family of cyclides is $[\hat{1}1111]$ and the corresponding diagrammatic representation is



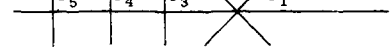
(iii) Two pairs of the quantities e_i are mutually complex conjugate, say e_1, e_2 and e_3, e_4. The corresponding choice of variables for x_i is

$$x_1=\sqrt{i}\,(r^2-p^2-q^2+is^2),\quad x_2=\sqrt{-i}\,(r^2-p^2-q^2-is^2) \quad (1.19)$$
$$x_3=\sqrt{2}(r+ip)s,\quad x_4=\sqrt{2}(r-ip)s,\quad x_4=2iqs.$$

Another associated choice is obtained by taking $p\to -ip$, $q\to -iq$, $r\to -ir$. The notation for such a family of cyclides is $[\hat{1}\hat{1}111]$ with the corresponding diagrammatic representation



The equations for a family of cyclides corresponding to the configuration [(11)111] are readily obtained from Eqs. (1.17) by putting $e_1=e_2$. The corresponding diagrammatic representation of this configuration is



The equations of the remaining configurations 2—7 are obtained as limiting cases of the general configuration (1.17). This leads to equations which are more convenient than those found in Eqs. (1.11)—(1.16). The method is illustrated here for the [2111] configuration and is explained in detail in Bôcher's book. As an illustration of the procedure we subject (1.17) to the

transformation $x_i \to \sqrt{a_i}\, x_i$ (a_i real) and take

$$e_2 = e_1 + \epsilon, \quad x_2 = x_1 + \epsilon x_2, \tag{1.20}$$

where ϵ is a first order quantity. Then by choosing a_i such that

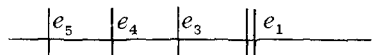
$$a_1 + a_2 = 0, \quad a_2\epsilon = 1, \quad a_3 = a_4 = a_5 = 1, \tag{1.21}$$

Eqs. (1.17) become

$$\Omega = 2x_1x_2 + x_3^2 + x_4^2 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^2} + \frac{2x_1x_2}{\lambda - e_1} + \frac{x_3^2}{\lambda - e_3} + \frac{x_4^2}{\lambda - e_4} + \frac{x_5^2}{\lambda - e_5} = 0. \tag{1.22}$$

These are then the equations of cyclides of type [2111]. The coordinates x_i in (1.22) have two interpretations:

(i) The e_i are all real. The corresponding diagrammatic representation is

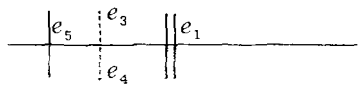


Here the two close parallel lines at e_1 signify the invariant factor index 2 in the [2111] configuration. The choice of variables x_i in this case is

$$x_1 = -2s^2, \quad x_2 = r^2 - p^2 - q^2, \tag{1.23}$$
$$x_3 = 2ips, \quad x_4 = 2iqs, \quad x_5 = 2rs.$$

The variables x_i are in this case a complex linear combination of the pentaspherical coordinates given in (1.6). An associated set of variables is given by the transformation $p \to -ip$, $q \to -iq$, $r \to -ir$.

(ii) Two of the quantities e_i, say e_3, e_4, are mutually complex conjugate. This corresponds to the configuration $[2\hat{1}1]$ and has the diagrammatic representation



The choice of variables x_i is given by

$$x_1 = -2s^2, \quad x_2 = r^2 - p^2 - q^2, \tag{1.24}$$
$$x_3 = \sqrt{2}(r + ip)s, \quad x_4 = \sqrt{2}(r - ip)s, \quad x_5 = 2iqs.$$

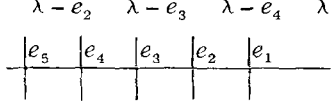
An associated set of variables is given by the transformation $p \to -ip$, $q \to -iq$, $r \to -ir$.

As we have mentioned, the expressions for all confocal families of cyclides can be derived from the general system (1.17) by methods similar to those illustrated here to pass to the configuration [2111]. We now list the equations for these families of curves and their associated diagrams. In the case of the configuration [221] we give the coordinates x_i in terms of the homogeneous coordinates p, q, r and s.

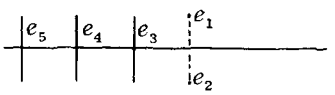
1. [11111], $[\hat{1}1111]$, and $[\hat{1}\hat{1}111]$

$$\Omega = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{\lambda - e_1} + \frac{x_2^2}{\lambda - e_2} + \frac{x_3^2}{\lambda - e_3} + \frac{x_4^2}{\lambda - e_4} + \frac{x_5^2}{\lambda - e_5} = 0: \tag{1.25}$$

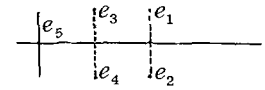
(i) [11111]



(ii) $[\hat{1}1111]$



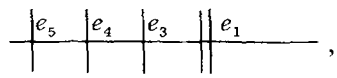
(iii) $[\hat{1}\hat{1}111]$



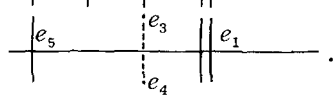
2. [2111] and $[2\hat{1}\hat{1}1]$:

$$\Omega = 2x_1x_2 + x_3^2 + x_4^2 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^2} + \frac{2x_1x_2}{\lambda - e_1} + \frac{x_3^2}{\lambda - e_3} + \frac{x_4^2}{\lambda - e_4} + \frac{x_5^2}{\lambda - e_5} = 0: \tag{1.26}$$

(i) [2111]



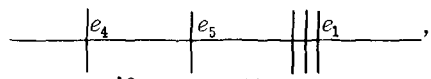
(ii) $[2\hat{1}\hat{1}1]$



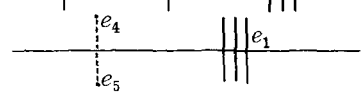
3. [311] and $[3\hat{1}\hat{1}]$

$$\Omega = 2x_1x_3 + x_2^2 + x_4^2 + x_5^2 = 0, \tag{1.27}$$
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^3} + \frac{2x_1x_2}{(\lambda - e_1)^2} + \frac{2x_1x_3 + x_2^2}{\lambda - e_1} + \frac{x_4^2}{\lambda - e_4} + \frac{x_5^2}{\lambda - e_5} = 0:$$

(i) [311]



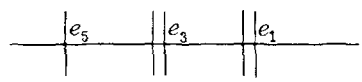
(ii) $[3\hat{1}\hat{1}]$



4. [221]

$$\Omega = 2x_1x_2 + 2x_3x_4 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^2} + \frac{2x_1x_2}{\lambda - e_1} + \frac{x_3^2}{(\lambda - e_3)^2} + \frac{2x_3x_4}{\lambda - e_3} + \frac{x_5^2}{\lambda - e_5} = 0: \tag{1.28}$$

(i) [221]



The corresponding expressions for the coordinates x_i in this case are

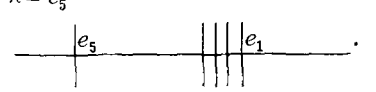
$$x_1 = -2s^2, \quad x_2 = r^2 - p^2 - q^2,$$
$$x_3 = \sqrt{2}(r - p)s, \quad x_4 = \sqrt{2}(r + p)s, \quad x_5 = 2iqs. \tag{1.29}$$

The associated set of coordinates being given as usual by $p \to -ip$, $q \to -iq$, $r \to -ir$. From Eqs. (1.10)–(1.16) it is seen that Ω is always one of the types found in systems corresponding to the configurations [11111], [2111], or [221]. The correspondence between the x_i's in this list with p, q, r, and s has now been determined in all cases.

5. [41]

$$\Omega = 2x_1x_4 + 2x_2x_3 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^4} + \frac{2x_1x_2}{(\lambda - e_1)^3} + \frac{2x_1x_3 + x_2^2}{(\lambda - e_1)^2} + \frac{2x_1x_4 + 2x_2x_3}{\lambda - e_1} + \frac{x_5^2}{\lambda - e_5} = 0: \tag{1.30}$$

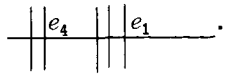
(i) [41]



6. [32]

$$\Omega = 2x_1x_3 + 2x_4x_5 + x_2^2 = 0, \tag{1.31}$$

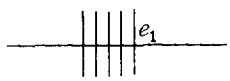
$$\Phi = \frac{x_1^2}{(\lambda - e_1)^3} + \frac{2x_1x_2}{(\lambda - e_1)^2} + \frac{2x_1x_3 + x_2^2}{\lambda - e_1} + \frac{x_4^2}{(\lambda - e_4)^2} + \frac{2x_4x_5}{\lambda - e_4} = 0:$$

(i) [32] 

7. [5]

$$\Omega = 2x_1x_5 + 2x_2x_4 + x_3^2 = 0, \tag{1.32}$$

$$\Phi = \frac{x_1^2}{(\lambda - e_1)^5} + \frac{2x_1x_2}{(\lambda - e_1)^4} + \frac{x_2^2 + 2x_1x_3}{(\lambda - e_1)^3} + \frac{2x_2x_3 + 2x_1x_4}{(\lambda - e_1)^2} + \frac{2x_1x_5 + 2x_2x_4 + x_3^2}{\lambda - e_1} = 0:$$

(i) [5] 

In the expression for Φ in this last case the final term is identically zero as it is proportional to Ω.

As was mentioned earlier, the coordinate curves for the cases in which brackets are inserted inside the square brackets can be obtained from this list by the appropriate substitution, e.g., [(32)] corresponds to curves (1.31) with $e_1 = e_4$.

Any two confocal families of the same type and configuration are equivalent under the action of linear transformations of the x_i which preserve the form Ω if their parameters e_i', λ' and e_i, λ are related by the equations

$$e_i = \frac{\alpha e_i' + \beta}{\gamma e_i' + \delta}, \quad \lambda = \frac{\alpha\lambda' + \beta}{\gamma\lambda' + \delta}, \quad \alpha, \beta, \gamma, \delta \in \mathbf{R}, \tag{1.33}$$

with $\alpha\delta - \beta\gamma \neq 0$. This equivalence is with respect to transformations which are isomorphic to the orthogonal transformations V which in our case are elements of O(3, 2).

We now turn our attention to the problem of relating the coordinates x_i in Eqs. (1.25)–(1.28), (1.30)–(1.32) to the parameters which specify an orthogonal family of such surfaces. These latter quantities are the curvilinear coordinates whose coordinate curves are mutually orthogonal at the common point of intersection. The problem of the ranges of variation of the parameters and the number of inequivalent types of parametrization for the real subset (1.6) are the subject of Sec. III. Here we just give the form of the coordinates x_i corresponding to each of the cases 1–7 outlined above when the coordinate curves are all of this type. The corresponding curvilinear coordinates are denoted by $\lambda = \mu, \rho, \nu$. For a coordinate system generated by cyclides of the type [11111] the coordinate curves are given by the equations

$$\Omega = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = 0,$$
$$\Phi = \frac{x_1^2}{\lambda - e_1} + \frac{x_2^2}{\lambda - e_2} + \frac{x_3^2}{\lambda - e_3} + \frac{x_4^2}{\lambda - e_4} + \frac{x_5^2}{\lambda - e_5} = 0, \tag{1.34}$$

with $\lambda = \mu, \rho$ or ν. The corresponding expression for the coordinates x_i is:

1. [11111]

$$\sigma x_i^2 = \phi(e_i)/f'(e_i), \quad i = 1, \ldots, 5, \tag{1.35}$$

where

$$-1/\sigma = e_1x_1^2 + e_2x_2^2 + e_3x_3^2 + e_4x_4^2 + e_5x_5^2$$

and

$$f(\lambda) = (\lambda - e_1)(\lambda - e_2)(\lambda - e_3)(\lambda - e_4)(\lambda - e_5),$$
$$\phi(\lambda) = (\mu - \lambda)(\nu - \lambda)(\rho - \lambda).$$

The coordinates in Minkowski space can be found from these expressions via the relations

$$t = \frac{-x_5}{x_2 + ix_1}, \quad x = \frac{ix_3}{x_2 + ix_1}, \quad y = \frac{ix_4}{x_2 + ix_1} \tag{1.36}$$

if the x_i are given as in (1.6). We will see in Sec. III that this need not always be the case, and we may be required to permute the expressions on the right-hand side of equations (1.6) so as to correspond to the correct signature as in (1.5). We now give the expressions for the coordinates x_i for the remaining six types of families of cyclides. These can be deduced by the same methods as used to deduce the form of the cyclides [2111] from the general case [11111]. We again refer to Bôcher's book for details. (Bôcher has given the formulas required to pass the configurations [2111] and [311]. The authors have extended this to include all remaining cases. Only the results are presented here.)

2. [2111]

$$\sigma x_1^2 = \frac{(\mu - e_1)(\nu - e_1)(\rho - e_1)}{(e_3 - e_1)(e_4 - e_1)(e_5 - e_1)},$$
$$2\sigma x_1x_2 = \frac{\partial}{\partial e_1}\left[\frac{(\mu - e_1)(\nu - e_1)(\rho - e_1)}{(e_3 - e_1)(e_4 - e_1)(e_5 - e_1)}\right],$$
$$\sigma x_3^2 = \frac{(\mu - e_3)(\nu - e_3)(\rho - e_3)}{(e_1 - e_3)^2(e_4 - e_3)(e_5 - e_3)}, \tag{1.37}$$
$$\sigma x_4^2 = \frac{(\mu - e_4)(\nu - e_4)(\rho - e_4)}{(e_1 - e_4)^2(e_3 - e_4)(e_5 - e_4)},$$
$$\sigma x_5^2 = \frac{(\mu - e_5)(\nu - e_5)(\rho - e_5)}{(e_1 - e_5)^2(e_3 - e_5)(e_4 - e_5)},$$

where

$$-1/\sigma = 2e_1x_1x_2 + x_1^2 + e_3x_3^2 + e_4x_4^2 + e_5x_5^2,$$

$$\sigma x_1^2 = \frac{(\mu - e_1)(\nu - e_1)(\rho - e_1)}{(e_4 - e_1)(e_5 - e_1)},$$

$$2\sigma x_1x_2 = \frac{\partial}{\partial e_1}\left[\frac{(\mu - e_1)(\nu - e_1)(\rho - e_1)}{(e_4 - e_1)(e_5 - e_1)}\right], \tag{1.38}$$

$$\sigma(2x_1x_3 + x_2^2) = \frac{1}{2}\frac{\partial^2}{\partial e_1^2}\left[\frac{(\mu - e_1)(\nu - e_1)(\rho - e_1)}{(e_4 - e_1)(e_5 - e_1)}\right],$$

$$\sigma x_4^2 = \frac{(\mu - e_4)(\nu - e_4)(\rho - e_4)}{(e_1 - e_4)^3(e_5 - e_4)},$$

$$\sigma x_5^2=\frac{(\mu-e_5)(\nu-e_5)(\rho-e_5)}{(e_1-e_5)^3(e_4-e_5)},$$

where

$$\frac{-1}{\sigma}=e_1(2x_1x_3+x_2^2)+2x_1x_2+e_4x_4^2+e_5x_5^2.$$

4. [221]

$$\begin{aligned}
&\sigma x_1^2=\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_3-e_1)^2(e_1-e_5)},\\
&2\sigma x_1x_2=-\frac{\partial}{\partial e_1}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_3-e_1)^2(e_1-e_5)}\right],\\
&\sigma x_3^2=\frac{(\mu-e_3)(\nu-e_3)(\rho-e_3)}{(e_3-e_1)^2(e_3-e_5)},\\
&2\sigma x_3x_4=-\frac{\partial}{\partial e_3}\left[\frac{(\mu-e_3)(\nu-e_3)(\rho-e_3)}{(e_3-e_1)^2(e_5-e_4)}\right],\\
&\sigma x_5^2=\frac{(\mu-e_5)(\nu-e_5)(\rho-e_5)}{(e_1-e_5)^2(e_3-e_5)^2},
\end{aligned}\tag{1.39}$$

where

$$-1/\sigma=2e_1x_1x_2+x_1^2+2e_3x_3x_4+x_3^2+e_5x_5^2.$$

5. [41]

$$\begin{aligned}
&\sigma x_1^2=\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_1-e_5)},\\
&2\sigma x_1x_2=\frac{\partial}{\partial e_1}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_1-e_5)}\right],\\
&\sigma(2x_1x_3+x_2^2)=\frac{1}{2}\frac{\partial^2}{\partial e_1^2}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_1-e_5)}\right],\\
&\sigma(2x_1x_4+2x_2x_3)=\frac{1}{6}\frac{\partial^3}{\partial e_1^3}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_1-e_5)}\right],\\
&\sigma x_5^2=\frac{(\mu-e_5)(\nu-e_5)(\rho-e_5)}{(e_5-e_1)^4},
\end{aligned}\tag{1.40}$$

where

$$-1/\sigma=2e_1(x_2x_3+x_1x_4)+2x_1x_3+x_2^2+e_5x_5^2.$$

6. [32]

$$\begin{aligned}
&\sigma x_1^2=\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_4-e_1)^2},\\
&2\sigma x_1x_2=\frac{\partial}{\partial e_1}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_4-e_1)^2}\right],\\
&\sigma(2x_1x_3+x_2^2)=\frac{1}{2}\frac{\partial^2}{\partial e_1^2}\left[\frac{(\mu-e_1)(\nu-e_1)(\rho-e_1)}{(e_4-e_1)^2}\right],\\
&\sigma x_4^2=\frac{(\mu-e_4)(\nu-e_4)(\rho-e_4)}{(e_4-e_1)^3},\\
&2\sigma x_4x_5=\frac{\partial}{\partial e_4}\left[\frac{(\mu-e_4)(\nu-e_4)(\rho-e_4)}{(e_4-e_1)^3}\right],
\end{aligned}\tag{1.41}$$

where

$$-1/\sigma=e_1(2x_1x_3+x_2^2)+2x_1x_2+2e_4x_4x_5+x_4^2.$$

7. [5]

$$\begin{aligned}
&\sigma x_1^2=(\mu-e_1)(\nu-e_1)(\rho-e_1),\\
&2\sigma x_1x_2=\frac{\partial}{\partial e_1}[(\mu-e_1)(\nu-e_1)(\rho-e_1)],\\
&\sigma(2x_1x_3+x_2^2)=\frac{1}{2}\frac{\partial^2}{\partial e_1^2}[(\mu-e_1)(\nu-e_1)(\rho-e_1)],\\
&2\sigma(x_2x_3+x_1x_4)=-1,
\end{aligned}\tag{1.42}$$

where σ is given by the last equation in this case.

These are the basic formulas for the pentaspherical coordinates expressed in terms of the curvilinear coordinates μ, ρ, and ν for all nondegenerate cyclides. The expressions for the coordinates in the case of a degenerate configuration are directly derivable from these formulas. The explicit methods for doing this will be discussed in Sec. III, where we evaluate all the possible inequivalent systems 1—7 and the associated degenerate forms. Finally in this section we give the formula expressing the line element ds^2 in terms of the curvilinear coordinates μ, ρ, ν and the pentaspherical coordinates x_i,

$$ds^2=\frac{1}{4\sigma s^2}\left(\frac{(\mu-\nu)(\mu-\rho)}{f(\mu)}d\mu^2+\frac{(\nu-\mu)(\nu-\rho)}{f(\nu)}d\nu^2+\frac{(\rho-\mu)(\rho-\nu)}{f(\rho)}d\rho^2\right)\tag{1.43}$$

with $f(\lambda)=\Pi_{i=1}^5(\lambda-e_i)$ as in (1.35). In each case σ is the quantity in the above list given for each configuration. The quantity s is the homogeneous coordinate that was introduced in (1.3) and can be expressed in terms of the x_i depending on the configuration in question. This formula is basic to the classification of coordinate systems which are inequivalent under the action of the underlying transformation group O(3, 2).

We summarize what has been done to this point. We have given the equations required to pass to a subspace of pentaspherical space having definite real signature as in (1.5). The associated group of transformations which preserve this subspace is isomorphic to O(3, 2) the local symmetry group of (*). The corresponding second order curves or cyclides in these coordinates can then be classified into equivalence classes under the action of this group of transformations. Those curves of special interest are the families of confocal cyclides and the coordinate systems to which they correspond. An important feature here is that all families of confocal orthogonal cyclides can be obtained as specified limits of the most general case corresponding to the configuration [11111].

II. THE TWO-DIMENSIONAL WAVE EQUATION AND R-SEPARABLE COORDINATES GENERATED FROM ORTHOGONAL FAMILIES OF CONFOCAL CYCLIDES

In this section we summarize the results that enable (*) to have an R-separable solution. For more details we refer to Bocher's book[3] and also to Morse and Feshbach.[4] The central result with which we will be concerned is the form of the equation (*) when written in terms of the cyclidic coordinates discussed in the previous section. Of central interest is the case of cyclidic coordinates corresponding to the configuration [11111]. The result is the following. If ψ is a solution of $\partial_{tt}\psi=\Delta_2\psi$ and if we write

$$\psi = \sqrt{2}\sigma^{1/4} s\phi(\mu, \nu, \rho), \tag{2.1}$$

where μ, ν, ρ are cyclidic coordinates of the type [11111] and

$$-1/\sigma = e_1 x_1^2 + e_2 x_2^2 + e_3 x_3^2 + e_4 x_4^2 + e_5 x_5^2$$

with S as in (1.3), then ψ satisfies the differential equation

$$(\rho-\nu)\frac{\partial^2\phi}{\partial u^2} + (\mu-\rho)\frac{\partial^2\phi}{\partial v^2} + (\nu-\mu)\frac{\partial^2\phi}{\partial w^2} + (\mu-\nu)(\nu-\rho)(\rho-\mu)\left(\tfrac{5}{4}(\mu+\nu+\rho) - \tfrac{3}{4}\sum_{i=1}^{5} e_i\right)\phi = 0, \tag{2.2}$$

where

$$\frac{\partial}{\partial u} = 2\sqrt{f(\mu)}\frac{\partial}{\partial \mu}, \quad \frac{\partial}{\partial v} = 2\sqrt{f(\nu)}\frac{\partial}{\partial \nu},$$

and

$$\frac{\partial}{\partial w} = \sqrt{f(\rho)}\frac{\partial}{\partial \rho}.$$

Here $f(\lambda)$ is as usual given by

$$f(\lambda) = (\lambda - e_1)(\lambda - e_2)(\lambda - e_3)(\lambda - e_4)(\lambda - e_5).$$

Equation (2.2) admits a separable solution

$$\phi = E_1(\mu)E_2(\nu)E_3(\rho) \tag{2.3}$$

with each of the separated functions satisfying an equation of the form

$$\sqrt{f(\lambda)}\frac{d}{d\lambda}\sqrt{f(\lambda)}\frac{dE_i}{d\lambda} + \left[\frac{5}{16}\lambda^3 - \frac{3}{16}\left(\sum_{i=1}^{5} e_i\right)\lambda^2 - \frac{A}{4}\lambda - \frac{B}{4}\right]E_i = 0. \tag{2.4}$$

With this result all the separation equations for the coordinate systems given in the previous section can be obtained by taking appropriate limits in the above equations. Equation (2.4) is an equation of the Lamé type with six elementary singularities.[7] The quantities A and B are separation constants.

III. CLASSIFICATION OF ORTHOGONAL R-SEPARABLE COORDINATE SYSTEMS FOR THE WAVE EQUATION

In this section a systematic treatment is given of the orthogonal R-separable coordinates of (*), which can be constructed as limiting cases of general cyclidic coordinates with configurations [11111], [$\hat{1}$1111], and [$\hat{1}\hat{1}$111]. For each coordinate system we give the expression for the corresponding pentaspherical coordinates x_i and the Cartesian coordinates t, x, and y. The operators whose eigenvalues are the separation constants are also given in each case and expressed in terms of the generators of the symmetry group of (*) which were derived in I. We also say what we can about the solutions of the separated equations. Our procedure is the following. For the completely cyclidic coordinates listed in Eqs. (1.35)—(1.42) we must choose ranges for the curvilinear coordinates in such a way that the differential form (1.43) when expressed as in (0.2) must satisfy

$$\operatorname{sgn}F = \operatorname{sgn}G = -\operatorname{sgn}H. \tag{3.1}$$

This ensures that the space is three-dimensional Minkowski space. The classification of all such parametrizations into equivalence classes under the relation (1.33) then gives the inequivalent coordinate systems we need. For the general configuration [11111] the equivalence relation (1.33) allows us to interchange all the e_i in the way specified by these formulas. However, for the remaining configurations such as [2111] only the three unit indices can change under the relation (1.33) when classifying equivalence classes of this type. In addition for each class of coordinate systems we choose a standardized representative which has a simple form. In most cases this will involve taking one of the indices e_i to be ∞.

The method of connecting the operators whose eigenvalues are the separation constants with the generators of the symmetry group $G = O(3, 2)$ is achieved by noting that the generators Γ_{ij} as defined in I are related to the generators of the underlying O(3, 2) group which preserves the pentaspherical space identity (1.1) with the choice of coordinates (1.6). The relations are

$$\begin{aligned} &\Gamma_{13} = L_{14}, \quad \Gamma_{12} = L_{15}, \quad \Gamma_{23} = L_{34}, \\ &\Gamma_{45} = L_{52}, \quad \Gamma_{15} = -iL_{12}, \quad \Gamma_{14} = -iL_{15}, \\ &\Gamma_{52} = iL_{23}, \quad \Gamma_{53} = iL_{24}, \quad \Gamma_{24} = -iL_{35}, \end{aligned} \tag{3.2}$$

where

$$L_{ij} = x_j\partial_i - x_i\partial_j, \quad i, j = 1, 2, \ldots, 5,$$

with the x_j as in (1.6). By means of the relations (see I)

$$\begin{aligned} &M_{12} = \Gamma_{23}, \quad M_{01} = \Gamma_{42}, \quad M_{02} = \Gamma_{43}, \quad D = \Gamma_{15}, \\ &P_0 = \Gamma_{14} + \Gamma_{45}, \quad K_0 = \Gamma_{14} - \Gamma_{45}, \quad P_1 = \Gamma_{12} + \Gamma_{25}, \\ &K_1 = \Gamma_{12} - \Gamma_{25}, P_2 = \Gamma_{13} + \Gamma_{35}, \quad K_2 = \Gamma_{13} - \Gamma_{35}, \end{aligned} \tag{3.3}$$

the operators whose eigenvalues are the separation constants in a given R-separable coordinate system can be expressed as second order symmetric operators in the generators of the O(3, 2) symmetry group of (*). In the subsequent classification of R-separable orthogonal solutions of (*) we will have occasion to introduce a number of modifications of Bôcher's diagramatic notation as well as some of the limiting procedures of interest for the various degenerate configurations being considered.

A further comment is in order here. In order to give all the coordinate systems that are potentially of interest, we give in the subsequent listing, with the exception of systems of the type [11111], all the separable systems of (*) which are inequivalent under the underlying $E(2, 1)$ group. This gives a more thorough treatment of these coordinate systems already considered in an earlier article.[8] In the concluding remarks we indicate which of these systems, which are not equivalent under $E(2, 1)$, are equivalent under the symmetry group O(3, 2) of (*).

We now proceed to the classification of the coordinate systems of interest.

A. The configurations [11111], [$\hat{1}$1111], [$\hat{1}\hat{1}$111] and their degenerate forms

1. *The configurations [11111], [$\hat{1}$1111] and [$\hat{1}\hat{1}$111]*

Here we give those configurations of the form [11111]

which are inequivalent under the procedure outlined in the introductory paragraphs of this section. For configurations of this type we can transform the quantities e_i via (1.33) to be

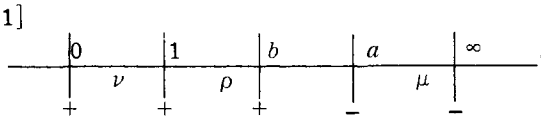
$$e_1=\infty,\quad e_2=a,\quad e_3=b,\quad e_4=1,\quad e_5=0.$$

In addition to Bôcher's diagrammatic notation for such a configuration, as given in Sec. I, we put the sign of the expression σx_i^2 at the bottom of the vertical line in the diagram of the [11111] configuration. From the formulas (1.6) the arrangement of these signs indicates how the choice of pentaspherical coordinates should be made. This involves a permutation of the quantities on the right-hand side of (1.6). In each of the inequivalent parametrizations for the configuration [11111] a specific choice of the x_i is made to within a permutation of those x_i whose squares have the same sign. This is sufficient for our purposes as all coordinate systems that are related by such permutations will be equivalent and related by a group transformation. The two additional operators $\hat{A}$, $\hat{B}$ whose eigenvalues are A and B, respectively, as in (2.4), have the form

$$\begin{aligned}\hat{A}&=\frac{(\nu+\rho)}{(\mu-\rho)(\mu-\nu)}\frac{\partial^2}{\partial u^2}+\frac{(\mu+\rho)}{(\nu-\rho)(\nu-\mu)}\frac{\partial^2}{\partial v^2}\\&\quad+\frac{(\mu+\nu)}{(\rho-\nu)(\rho-\mu)}\frac{\partial^2}{\partial w^2},\\ \hat{B}&=\frac{\nu\rho}{(\mu-\rho)(\mu-\nu)}\frac{\partial^2}{\partial u^2}+\frac{\mu\rho}{(\nu-\rho)(\nu-\mu)}\frac{\partial^2}{\partial v^2}\\&\quad+\frac{\mu\nu}{(\rho-\nu)(\rho-\mu)}\frac{\partial^2}{\partial w^2}\end{aligned}\tag{3.4}$$

when acting on the functions $\phi(\mu,\nu,\rho)$ as in (2.1). The part of the solution of (*) that gives the R-separation (called hereafter the modulation factor following Morse and Feshbach) is from (2.1), $\sqrt{2}\,\sigma^{1/4}s$. Corresponding to the configuration [11111] being considered in this subsection we have the following inequivalent possibilities.

(a) [11111]



For such a configuration the pentaspherical coordinates are

$$\begin{aligned}&\sigma x_1^2=-1,\quad \sigma x_2^2=-\frac{(\mu-a)(\nu-a)(\rho-a)}{(a-b)(a-1)a},\\&\sigma x_3^2=-\frac{(\mu-b)(\nu-b)(\rho-b)}{(b-a)(b-1)b},\\&\sigma x_4^2=-\frac{(\mu-1)(\nu-1)(\rho-1)}{(a-1)(b-1)},\quad \sigma x_5^2=\frac{\mu\nu\rho}{ab}.\end{aligned}\tag{3.5}$$

The coordinates in three-dimensional Minkowski space are given by the formulas

$$\begin{aligned}t&=\frac{-x_2}{(x_1+ix_5)}=\frac{1}{R}\left(\frac{(\mu-a)(\nu-a)(\rho-a)}{(a-b)(a-1)a}\right)^{1/2},\\x&=\frac{ix_4}{(x_1+ix_5)}=\frac{1}{R}\left(\frac{(\mu-1)(\rho-1)(1-\nu)}{(a-1)(b-1)}\right)^{1/2},\\y&=\frac{ix_3}{(x_1+ix_5)}=\frac{1}{R}\left(\frac{(\mu-b)(\rho-b)(\nu-b)}{(a-b)(b-1)b}\right)^{1/2},\end{aligned}\tag{3.6}$$

where $R=i(1+\sqrt{\mu\nu\rho/ab})$. The modulation factor is $\sqrt{2}\sigma^{1/4}s=(1+\sqrt{\mu\nu\rho/ab})^{1/2}$.

The operators $\hat{A}$, $\hat{B}$ defining the eigenvalues of the separation constants (we refer to these as basis operators subsequently) are

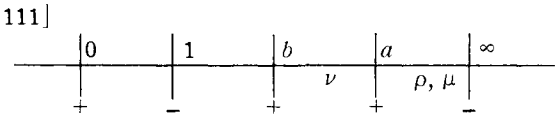
$$\begin{aligned}\hat{A}&=bM_{01}^2-M_{02}^2-aM_{12}^2-\tfrac{1}{4}(a+1)(P_2+K_2)^2\\&\quad+\tfrac{1}{4}(b+1)(P_0+K_0)^2+\tfrac{1}{4}(a+b)(P_1+K_1)^2,\\4\hat{B}&=b(P_0+K_0)^2-a(P_2+K_2)^2+ab(P_1+K_1)^2\end{aligned}\tag{3.7}$$

and the separation equations have the form

$$\sqrt{h(\lambda)}\frac{d}{d\lambda}\sqrt{h(\lambda)}\frac{dE_i}{d\lambda}-(\tfrac{3}{16}\lambda^2+A\lambda+B)E_i=0\tag{3.8}$$

with $h(\lambda)=\lambda(\lambda-1)(\lambda-b)(\lambda-a)$ and $\lambda=\mu,\nu$ or ρ for $i=1$, 2, or 3 respectively just as in (2.3). Equation (3.10) is a standard form of an equation with five elementary singularities (see, for instance, Ince, Ref. 7, p. 500). It should be noted here that the form of the pentaspherical coordinates (1.6) when subjected to the transformation $p\to ip$, $q\to -iq$, $r\to -ir$ gives no new information, i.e., exactly the same coordinate system results.

(b) [11111]



The pentaspherical coordinates are as in (3.6) with the three-dimensional Minkowski space coordinates given by

$$\begin{aligned}t&=-x_4/(x_1+ix_5),\quad x=ix_2/(x_1+ix_5),\\y&=ix_3/(x_1+ix_5).\end{aligned}\tag{3.9}$$

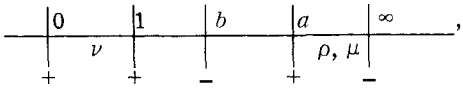
The modulation factor $\sqrt{2}\,\sigma^{1/4}S$ is the same as in (a). The basis operators are

$$\begin{aligned}\hat{A}&=aM_{02}^2+bM_{01}^2-M_{12}^2+\tfrac{1}{4}(a+1)(P_2+K_2)^2\\&\quad+\tfrac{1}{4}(b+1)(P_0-K_0)^2+\tfrac{1}{4}(a+b)(P_1-K_1)^2,\\-4\hat{B}&=b(P_1+K_1)^2+a(P_2-K_2)^2+ab(P_0+K_0)^2,\end{aligned}\tag{3.10}$$

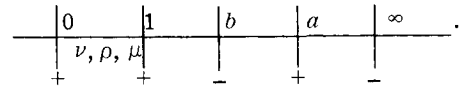
and the separation equations have the form (3.10).

(c) [11111]

(i)



(ii)



The pentaspherical coordinates are as in (3.6) with the three space coordinates given by

$$\begin{aligned}t&=-x_3/(x_1+ix_5),\quad x=ix_4/(x_1+ix_5),\\y&=ix_2/(x_1+ix_5).\end{aligned}\tag{3.11}$$

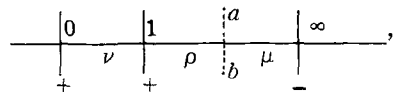
The modulation factor is the same as in (a). The basis operators are

$$\begin{aligned}\hat{A}&=aM_{01}^2-bM_{12}^2-M_{02}^2-\tfrac{1}{4}(a+1)(P_0+K_0)^2\\&\quad+\tfrac{1}{4}(b+1)(P_2+K_2)^2-\tfrac{1}{4}(a+b)(P_1+K_1)^2,\\4\hat{B}&=b(P_2+K_2)^2-a(P_0+K_0)^2-ab(P_1+K_1)^2,\end{aligned}\tag{3.12}$$

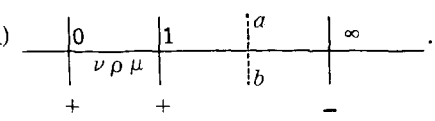
and the separation equations have the form (3.10).

(d) $[\hat{1}\hat{1}111]$

(i)



(ii)



The pentaspherical coordinates are as in (3.6) with $a = \alpha + i\beta$, $b = \alpha - i\beta$, $\alpha, \beta \in \mathbb{R}$. The three space coordinates are given by

$$t = \frac{(x_2 + x_3)}{\sqrt{2}(x_1 + ix_5)}, \quad x = \frac{i(x_2 - x_3)}{\sqrt{2}(x_1 + ix_5)}, \quad y = \frac{ix_4}{(x_1 + ix_5)} \tag{3.13}$$

The basis operators are

$$\begin{aligned} 2\hat{A} &= \alpha(M_{01}^2 - M_{12}^2) + \beta(M_{01}M_{12} + M_{12}M_{01}) \\ &\quad + (\alpha+1)[(P_2+K_2)^2 - (P_0+K_0)^2] + \alpha(P_1+K_1)^2 \\ &\quad - (\beta+1)[(P_2+K_2)(P_0+K_0) + (P_0+K_0)(P_2+K_2)], \end{aligned} \tag{3.14}$$

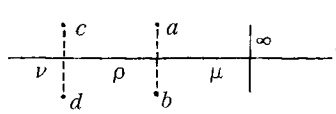
$$\begin{aligned} 4\hat{B} &= \alpha[(P_0+K_0)^2 - (P_2+K_2)^2] + (\alpha^2+\beta^2)(P_1+K_1)^2 \\ &\quad + \beta[(P_0+K_0)(P_2+K_2) + (P_2+K_2)(P_0+K_0)]. \end{aligned}$$

The modulation factor is

$$\sqrt{2}\sigma^{1/4}s = \{1 + [\mu\nu\rho/(\alpha^2+\beta^2)]^{1/2}\}^{1/2},$$

and the separation equations have the form (3.10).

(e) $[\hat{1}\hat{1}\hat{1}11]$



The pentaspherical coordinates are given by

$$\begin{aligned} \sigma x_1^2 &= -1, \quad \sigma x_2^2 = -\frac{(\mu-a)(\nu-a)(\rho-a)}{(a-b)(a-c)(a-d)}, \\ \sigma x_3^2 &= -\frac{(\mu-b)(\nu-b)(\rho-b)}{(b-a)(b-c)(b-d)}, \\ \sigma x_4^2 &= -\frac{(\mu-c)(\nu-c)(\rho-c)}{(c-a)(c-b)(c-d)}, \quad \sigma x_5^2 = -\frac{(\mu-d)(\nu-d)(\rho-d)}{(d-c)(d-a)(d-b)}, \end{aligned} \tag{3.15}$$

where $a = \alpha + i\beta$, $b = \alpha - i\beta$, $c = \gamma + i\delta$, $d = \gamma - i\delta$ with $\alpha, \beta, \gamma, \delta \in \mathbb{R}$. The three space coordinates are given by

$$\begin{aligned} t &= -(x_4 + x_5)/[(x_2 + x_3) + i\sqrt{2}x_1], \\ x &= (x_3 - x_2)/[i(x_2 + x_3) - \sqrt{2}x_1], \\ y &= (x_5 - x_4)/[i(x_2 + x_3) - \sqrt{2}x_1], \end{aligned} \tag{3.16}$$

and the modulation factor is

$$\sqrt{2}\sigma^{1/4}s = [(-x_2 - x_3)/\sqrt{2} - ix_1]^{1/2}.$$

The basis operators are

$$\begin{aligned} \hat{A} &= 2\beta[(P_0 - K_0)M_{01} + M_{01}(P_0 - K)] + \beta[(P_0 - K_0)(P_2 - K_2) \\ &\quad + (P_2 - K_2)(P_0 - K_0)] - 4\gamma(M_{12}M_{01} + M_{01}M_{12}) \\ &\quad - 2\gamma[M_{12}(P_2 - K_2) + (P_2 - K_2)M_{12}] \\ &\quad + 2(\gamma - \alpha)[\tfrac{1}{4}(P_1 - K_1)^2 - M_{02}^2], \end{aligned}$$

$$\begin{aligned} \hat{B} &= 2[(\alpha+\delta)^2 + (\beta+\delta)^2][M_{12}^2 + \tfrac{1}{4}(P_0 - K_0)^2 \\ &\quad + M_{01}^2 + \tfrac{1}{4}(P_2 - K_2)^2 + \tfrac{1}{2}M_{01}(P_2 - K_2) + \tfrac{1}{2}(P_2 - K_2)M_{01}] \\ &\quad - 2\beta\delta[(P_0 - K_0)M_{12} + M_{12}(P_0 - K_0)] \\ &\quad + 4\alpha\gamma[(P_1 - K_1)^2 - 4M_{02}^2] + [1/(\alpha+\gamma)] \\ &\quad [\{2(\alpha+\gamma)(\delta\gamma - \alpha\beta) + (\beta - \delta)(\alpha^2 + \gamma^2 - \beta^2 - \delta^2)\} \\ &\quad \times\{[(M_{12} + \tfrac{1}{2}(P_0 - K_0)][\tfrac{1}{2}(K_2 - P_2) - M_{01}] \\ &\quad + [\tfrac{1}{2}(K_2 - P_2) - M_{01}][M_{12} + \tfrac{1}{2}(P_0 - K_0)]\} \\ &\quad + \{2(\alpha+\gamma)(\delta\gamma + \alpha\beta) - (\beta+\delta)(\alpha^2 + \gamma^2 - \beta^2 - \delta^2)\} \\ &\quad \times\{[\tfrac{1}{2}(K_0 - P_0) - M_{12}][\tfrac{1}{2}(P_2 - K_2) + M_{01}] + [\tfrac{1}{2}(P_2 - K_2) + M_{01}] \\ &\quad \times[\tfrac{1}{2}(K_0 - P_0) - M_{12}]\} + 2\alpha(\gamma^2 - \delta^2)[\tfrac{1}{4}(P_1 - K_1)^2 - M_{02}^2]. \end{aligned} \tag{3.17}$$

2. The configurations [(11)111], [11(11)1]

Here we must digress briefly to explain how the pentaspherical coordinates for the configuration [(11)111] can be obtained from the formulas (1.35) for the general configuration. To find the pentaspherical coordinates for the configuration [(11)111] for which say $e_1 = e_2$, we proceed as follows, putting

$$e_1 = e_2 + \epsilon, \quad \lambda = e_2 + \epsilon\lambda',$$

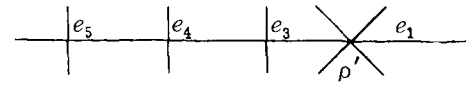
where for definiteness we take $\lambda = \rho$. The resulting expression for the pentaspherical coordinates is

$$\begin{aligned} \sigma x_1^2 &= \frac{(\mu - e_1)(\nu - e_1)}{(e_3 - e_1)(e_4 - e_1)(e_5 - e_1)}(1 - \rho'), \\ \sigma x_2^2 &= \frac{(\mu - e_1)(\nu - e_1)}{(e_3 - e_1)(e_4 - e_1)(e_5 - e_1)}\rho', \\ \sigma x_3^2 &= \frac{(\mu - e_3)(\nu - e_3)}{(e_1 - e_3)(e_4 - e_3)(e_5 - e_3)}, \\ \sigma x_4^2 &= \frac{(\mu - e_4)(\nu - e_4)}{(e_1 - e_4)(e_3 - e_4)(e_5 - e_4)}, \\ \sigma x_5^2 &= \frac{(\mu - e_5)(\nu - e_5)}{(e_1 - e_5)(e_3 - e_5)(e_4 - e_5)}. \end{aligned} \tag{3.18}$$

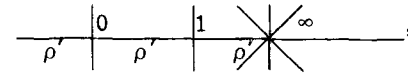
The coordinate curves corresponding to the new curvilinear coordinate ρ' are

$$x_1^2/\rho' + x^2(\rho' - 1) = 0. \tag{3.19}$$

This defines a family of real curves for $0 < \rho' < 1$ if $\mathrm{sgn}(x_1^2/x_2^2) = -1$. Otherwise for a real curve we must have $\mathrm{sgn}(x_1^2/x_2^2) = 1$. The diagrammatic notation for the family of degenerate cyclides specified by the curvilinear coordinates μ and ν is



The method of obtaining other degenerate forms corresponding to a configuration [(11)111] is to generalize the procedure outlined here to the case of two adjacent parameters e_i, e_{i+1} becoming equal. The diagram representing the curve (3.19) is



where ρ' may be in one of the regions indicated according as the relative sign of x_1^2 and x_2^2 is ± 1, as we have discussed above. The separation equations for the func-

tion $\psi(\mu,\nu,\rho')$ are given by (2.4) with $e_1=e_2$ and $\lambda=\mu,\nu$. For ρ' we obtain

$$(e_2-e_3)(e_2-e_4)(e_2-e_5)\sqrt{|\rho'(\rho'-1)|}\frac{d}{d\rho'}\sqrt{|\rho'(\rho'-1)|}\frac{dE_3}{d\rho'}$$

$$=[\tfrac{1}{16}e_2^2(e_2+3e_3+3e_4+3e_5)+\tfrac{1}{4}Ae_2+\tfrac{1}{4}B]E_3.$$

For all the classes of inequivalent coordinate systems of the type [(11)111] the quantities e_i will be standardized to be 0, 1, a, and ∞. This greatly simplifies all the calculations. For instance, in the example we have presented here this standardization can be achieved by taking

$$e_1=\infty,\quad e_3=a,\quad e_4=1,\quad e_5=0.$$

The resulting standardized form then gives the following expressions for the pentaspherical coordinates:

$$\sigma x_1^2=\rho'-1,\quad \sigma x_2^2=-\rho',\quad \sigma x_3^2=\frac{(\mu-a)(\nu-a)}{a(a-1)},$$
$$\sigma x_4^2=\frac{(\mu-1)(\nu-1)}{(1-a)},\quad \sigma x_5^2=\frac{\mu\nu}{a}. \tag{3.20}$$

The separation equations are

$$\sqrt{p(\lambda)}\frac{d}{d\lambda}\sqrt{p(\lambda)}\frac{dE_i}{d\lambda}-(A\lambda+B)E_i=0. \tag{3.21}$$

$\lambda=\mu,\nu$ and $i=1,2$ respectively, and

$$\sqrt{|\rho'(\rho'-1)|}\frac{d}{d\rho'}\sqrt{|\rho'(\rho'-1)|}\frac{dE_3}{d\rho'}-(\tfrac{1}{4}+A)E_3=0.$$

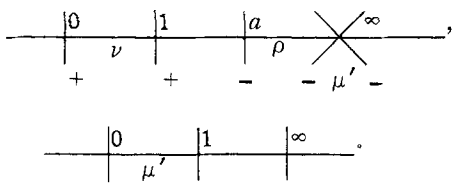
Here $p(\lambda)=\lambda(\lambda-1)(\lambda-a)$. Equation (3.21) is a form of Lamé's equation (see, for instance, Ince, Ref. 7, p. 502). The basis operators $\hat{A}$, $\hat{B}$ whose eigenvalues are the separation constants A and B respectively are in this case

$$\hat{A}=-\tfrac{1}{4}-\sqrt{|\rho'(\rho'-1)|}\frac{d}{d\rho'}\sqrt{|\rho'(\rho'-1)|}\frac{d}{d\rho'} \tag{3.22}$$
$$\hat{B}=\frac{1}{(\nu-\mu)}\left(\nu\sqrt{p(\mu)}\frac{\partial}{\partial\mu}\sqrt{p(\mu)}\frac{\partial}{\partial\mu}-\mu\sqrt{p(\nu)}\frac{\partial}{\partial\nu}\sqrt{p(\nu)}\frac{\partial}{\partial\nu}\right).$$

acting on ϕ.

We now proceed to the evaluation of the inequivalent types of coordinate systems of type [(11)111] and $[\hat{1}\hat{1}(11)1]$.

(a) [(11)111]



The pentaspherical coordinates are obtained from (3.19) subjected to the transformation

$$\rho'\to\mu',\quad \nu\to\nu,\quad \mu\to\rho.$$

The three space Minkowski coordinates are given by

$$t=-x_5/(ix_3+x_4)=(1/R)\sqrt{\nu\rho/a},$$
$$x=ix_2/(ix_3+x_4)=(1/R)\cos\phi, \tag{3.23}$$
$$y=ix_1/(ix_3+x_4)=(1/R)\sin\phi,$$

where

$$R=\left[\left(\frac{(\mu-a)(\nu-a)}{a(1-a)}\right)^{1/2}+\left(\frac{(\mu-1)(\nu-1)}{(a-1)}\right)^{1/2}\right]$$

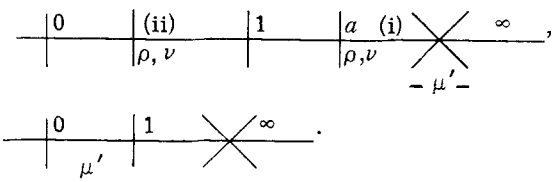
and we have put $\nu'=\sin^2\phi$. The modulation factor is

$$\sqrt{2}\,\sigma^{1/4}S=R^{1/2} \tag{3.24}$$

and the basis operators are

$$4\hat{B}=(P_0+K_0)^2-a(P_0-K_0)^2,\quad \hat{A}=-\tfrac{1}{4}-M_{12}^2. \tag{3.25}$$

(c)† [(11)111]



The pentaspherical coordinates are as in (a). The three space coordinates are given by

$$t=x_3/(ix_4+x_5),\quad x=ix_1/(ix_4+x_5),$$
$$y=ix_2/(ix_4+x_5) \tag{3.26}$$

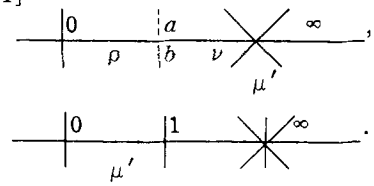
The modulation factor is

$$\sqrt{2}\,\sigma^{1/4}S=\left[\left(\frac{\mu\nu}{a}\right)^{1/2}+\left(\frac{(\mu-1)(\nu-1)}{(a-1)}\right)^{1/2}\right]^{1/2} \tag{3.27}$$

and the basis operators are,

$$4\hat{B}=-(P_0-K_0)^2+4aD^2,\quad \hat{A}=-\tfrac{1}{4}-M_{12}^2. \tag{3.28}$$

(d) $[\hat{1}\hat{1}(11)1]$



The pentaspherical coordinates are given by

$$\sigma x_1^2=\mu'-1,\quad \sigma x_2^2=-\mu',\quad \sigma x_3^2=\frac{(\rho-a)(\nu-a)}{a(a-b)},$$
$$\sigma x_4^2=\frac{(\rho-b)(\nu-b)}{b(b-a)},\quad \sigma x_5^2=\frac{\rho\nu}{ab}. \tag{3.29}$$

The three space coordinates are given by

$$t=\sqrt{2}\,x_5/R,\quad x=i\sqrt{2}\,x_1/R,\quad y=i\sqrt{2}\,x_2/R, \tag{3.30}$$

where $R=i(x_3-x_4)-(x_3+x_4)$. The modulation factor is

$$\sqrt{2}\,\sigma^{1/4}S=\sqrt{i}\left\{2\mathrm{Re}\left[\left(\frac{i(\rho-a)(\nu-a)}{a(a-b)}\right)^{1/2}\right]\right\}^{1/2} \tag{3.31}$$

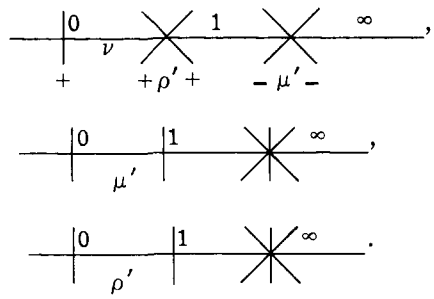
and the basis operators are

$$\hat{B}=\alpha(P_0K_0+K_0P_0)+2\beta(P_0^2-K_0^2),$$
$$\hat{A}=-\tfrac{1}{4}-M_{12}^2. \tag{3.32}$$

Here as usual $a=\alpha+i\beta$, $b=\alpha-i\beta$, $\alpha,\beta\in\mathbb{R}$. The separation equations have the form (3.20) and (3.21) with $p(\lambda)=\lambda(\lambda-a)(\lambda-b)$ and a is replaced in (3.21) by ab.

3. The configuration [(11)(11)1]

There is only one such coordinate system of interest here. The diagrams of this system are



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \mu' - 1, \quad \sigma x_2^2 = -\mu', \quad \sigma x_3^2 = (1-\nu)(1-\rho'),$$
$$\sigma x_4^2 = (1-\nu)\rho', \quad \sigma x_5^2 = \nu. \tag{3.33}$$

The three space coordinates are given by

$$t = ix_2/(ix_1 + x_5), \quad x = x_4/(ix_1 + x_5),$$
$$y = x_3/(ix_1 + x_5), \tag{3.34}$$

and the modulation factor is

$$\sqrt{2}\,\sigma^{1/4} S = (\sqrt{1-\mu'} + \sqrt{\nu})^{1/2}. \tag{3.35}$$

If we write $\mu' = \cos^2\phi$, $\nu = \cos^2\psi$, $\rho' = \cos^2\theta$, then

$$t = \cos\phi/(\sin\phi + \cos\psi),$$
$$x = \sin\psi\cos\theta/(\sin\phi + \cos\psi), \tag{3.36}$$
$$y = \sin\psi\sin\theta/(\sin\phi + \cos\psi).$$

The separation equations for this system of coordinates are given by

$$\sqrt{p(\mu)}\frac{d}{d\mu}\sqrt{p(\mu)}\frac{dE_1}{d\mu} - (A\mu + B)E_1 = 0, \tag{3.37}$$

where $\sqrt{p(\mu)} = \mu(\mu - 1)$.

Equation (3.37) is a form of the Legendres equation with spherical harmonic solution,

$$\frac{d^2E_2}{d\theta^2} + BE_2 = 0, \quad \frac{d^2E_3}{d\phi^2} + AE_3 = 0. \tag{3.38}$$

The basis defining operators are

$$4\hat{A} = (P_0 - K_0)^2, \quad \hat{B} = M_{12}^2. \tag{3.39}$$

This completes the classification of inequivalent coordinate systems of type [11111] and its degenerate forms. These are the only coordinate systems which will prove to be strictly R separable in the classification presented in this article.

B. THE CONFIGURATIONS [2111], [$2\hat{1}\hat{1}1$] and their degenerate forms

1. The configurations [2111] and [$2\hat{1}\hat{1}1$]

Here we give the configurations of the form [2111] and [$2\hat{1}\hat{1}1$], which are inequivalent under the equivalence relation discussed in the begining of this section. By applying a transformation of the type (1.33) to the indices e_i for the configuration [2111] it is always possible to choose these numbers in the standard form

$$e_1 = \infty, \quad e_3 = a, \quad e_4 = 1, \quad e_5 = 0 \tag{3.40}$$

with e_1 the number associated with the invariant factor index 2. The two operators whose eigenvalues are the separation constants are given as in (3.5) with $f(\lambda) = \lambda(\lambda-1)(\lambda-a)$.

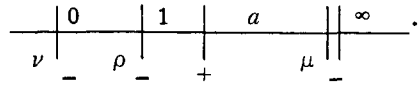
The separation equations with the choice of e_i given above are

$$\sqrt{f(\lambda)}\frac{d}{d\lambda}\sqrt{f(\lambda)}\frac{dE_i}{d\lambda} - (A\lambda + B)E_i = 0 \tag{3.41}$$

with $f(\lambda) = \lambda(\lambda-1)(\lambda-a)$ and $\lambda = \mu, \nu$, or ρ for $i = 1, 2$, or 3 respectively. This is Lamé's equation with four elementary singularities. For the configuration [$2\hat{1}11$] the separation equations are as in (3.41) with $f(\lambda) = \lambda(\lambda-a)(\lambda-b)$, where $a = \alpha + i\beta$ and $b = \alpha - i\beta$, $\alpha, \beta \in \mathbb{R}$. We now give the inequivalent coordinate systems.

For the choice of e_i given in (3.40) there is no modulation factor in the R-separated solutions. The solutions of (*) of the type [2111] are therefore separable.

(a) [2111]



The pentaspherical coordinates for this configuration are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = \mu + \nu + \rho + a + 1,$$
$$\sigma x_3^2 = \frac{(\mu-a)(\nu-a)(\rho-a)}{a(a-1)}, \tag{3.42}$$
$$\sigma x_4^2 = \frac{(\mu-1)(\nu-1)(\rho-1)}{(1-a)}, \quad \sigma x_5^2 = \frac{\mu\nu\rho}{a}.$$

The three space variables are given by

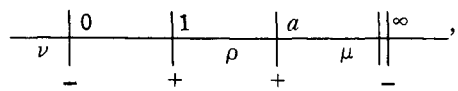
$$t^2 = -x_3^2/x_1^2 = (\mu-a)(\nu-a)(\rho-a)/a(a-1),$$
$$x^2 = x_4^2/x_1^2 = (\mu-1)(\nu-1)(\rho-1)/(a-1), \tag{3.43}$$
$$y^2 = x_5^2/x_1^2 = -\mu\nu\rho/a,$$

and the basis operators are

$$\hat{A} = P_0^2 - (a+1)P_2^2 - aP_1^2 + M_{12}^2 - M_{01}^2 - M_{02}^2,$$
$$\hat{B} = aP_2^2 + aM_{12}^2 - M_{02}^2. \tag{3.44}$$

For the remaining inequivalent systems of type [2111] we give the corresponding diagrams and the transformation which relates the three space coordinates given in (3.43) to each system. The expressions for the operators $\hat{A}$, $\hat{B}$ can be obtained from (3.44) via this substitution. In each case the pentaspherical coordinates are given by (3.42).

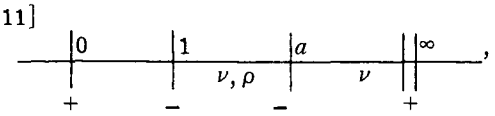
(b) [2111]



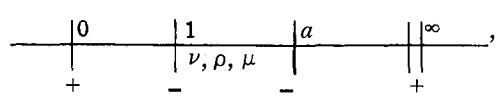
(a) − (b), $t \to x$, $x \to iy$, $y \to t$.

(c) [2111]

(i)

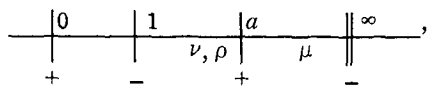


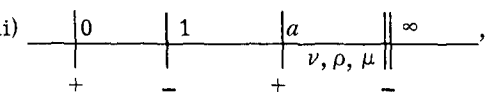
(ii)



(a) → (c), $t \to ix$, $x \to y$, $y \to it$.

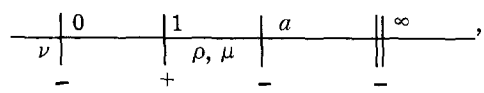
(d) [2111]

(i)


(iii)


(a) → (d), $t \to x$, $x \to t$, $y \to iy$.

(e) [2111]

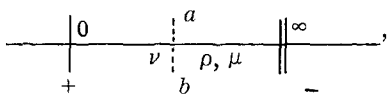


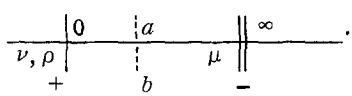
(a) → (e), $t \to ix$, $x \to it$, $y \to y$.

In all the above systems the choice of pentaspherical coordinates is made in two ways. If the net signature of the terms σx_i^2 from (1.37) for $i = 1, 2, 3, 4, 5$ is plus, then the form of the x_i's is as in (1.23). If the net signature is minus, then the required form of the x_i is obtained from (1.23) via the transformation

$$p \to -ip, \quad q \to -iq, \quad r \to -ir.$$

(f) [$2\hat{1}11$]

(i)


(ii)


The pentaspherical coordinates are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = \mu + \nu + \rho + a + b$$

$$\sigma x_3^2 = \frac{(\mu - a)(\nu - a)(\rho - a)}{a(a - b)},$$

$$\sigma x_4^2 = \frac{(\mu - b)(\nu - b)(\rho - b)}{b(b - a)}, \quad \sigma x_5^2 = \frac{\mu\nu\rho}{ab}, \tag{3.45}$$

where $a = \alpha + i\beta$, $b = \alpha - i\beta$, $\alpha\beta \in \mathbb{R}$. The three space coordinates are given by

$$t + ix = i\sqrt{2}\, x_3/x_1, \quad y = ix_5/x_1; \tag{3.46}$$

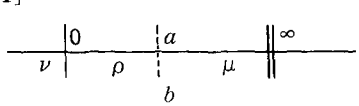
this follows from the use of formula (1.24) relating the x_i's to p, q, r, and s. [More exactly the coordinates obtained from (1.24) via the transformation $p \to -ip$, $q \to -iq$, $r \to -ir$.] The basis operators are

$$\hat{A} = \alpha(P_0^2 - P_1^2 - P_2^2) + 2\beta P_0 P_1 + M_{12}^2 - M_{01}^2 - M_{02}^2,$$

$$\hat{B} = -(\alpha^2 + \beta^2)P_2^2 + \alpha(M_{12}^2 - M_{02}^2) - \beta(M_{12}M_{02} + M_{02}M_{12}). \tag{3.47}$$

The term $\alpha(P_0^2 - P_1^2 - P_2^2)$ is included in the above expression for $\hat{A}$ so as to correspond to the correct operator derived from equations (3.5).

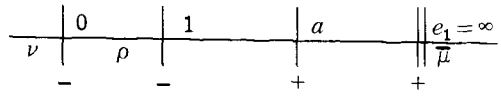
(g) [$2\hat{1}11$]



(f) → (g), $t \to x$, $x \to t$, $y \to iy$.

2. Configurations having a radial coordinate in three space derivable from configurations of the form [2111] and [$2\hat{1}11$]

Such coordinates can be derived in a straightforward manner, which we illustrate in detail for the first system of this section.

(a) (i)


For this diagram write $\mu = e_1 + \bar{\mu}$ in formulas (1.37) and then take $e_1 \to \infty$. The resulting pentaspherical coordinates have the form

$$\sigma x_1^2 = \bar{\mu}, \quad 2\sigma x_1 x_2 = 1, \quad \sigma x_3^2 = \frac{(\nu - a)(\rho - a)}{a(a - 1)},$$

$$\sigma x_4^2 = -\frac{(\nu - 1)(\rho - 1)}{(a - 1)}, \quad \sigma x_5^2 = \frac{\nu\rho}{a}, \tag{3.48}$$

and the three space coordinates are given by

$$t = \frac{x_3}{x_1} = r\left(\frac{(\nu - a)(\rho - a)}{a(a - 1)}\right)^{1/2},$$

$$x = \frac{ix_4}{x_1} = r\left(\frac{(\nu - 1)(\rho - 1)}{(a - 1)}\right)^{1/2}, \quad y = \frac{ix_5}{x_1} = r\left(\frac{-\nu\rho}{a}\right)^{1/2}, \tag{3.49}$$

where $r^2 = 1/\bar{\mu}$. The basis operators are given by

$$\hat{A} = \frac{4}{(\rho - \nu)}\left(\sqrt{P(\nu)}\frac{d}{d\nu}\sqrt{P(\nu)}\frac{d}{d\nu} - \sqrt{P(\rho)}\frac{d}{d\rho}\sqrt{P(\rho)}\frac{d}{d\rho}\right)$$

$$= \tfrac{1}{4} - D^2 \tag{3.50}$$

and

$$\hat{B} = \frac{4}{(\rho - \nu)}\left(\rho\sqrt{P(\nu)}\frac{d}{d\nu}\sqrt{P(\nu)}\frac{d}{d\nu} - \nu\sqrt{P(\rho)}\frac{d}{d\rho}\sqrt{P(\rho)}\frac{d}{d\rho}\right)$$

$$= M_{02}^2 - aM_{12}^2,$$

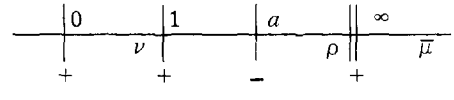
where $P(\lambda) = \lambda(\lambda - 1)(\lambda - a)$. The separation equations have the form

$$\sqrt{P(\lambda)}\frac{d}{d\lambda}\sqrt{P(\lambda)}\frac{dE_i}{d\lambda} + (A\lambda + B)E_i = 0, \tag{3.51}$$

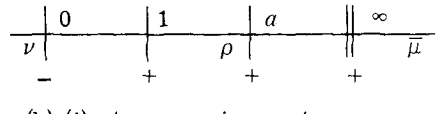
$\lambda = \nu, \rho$ and $i = 1, 2$, respectively, and $A = j(j + 1)$,

$$r^2\frac{d^2E_3}{dr^2} + 2r\frac{dE_3}{dr} - j(j + 1)E_3 = 0. \tag{3.52}$$

Equation (3.51) is a form of the Lamé equation and the solutions of (3.52) are r^j and r^{-j-1}.

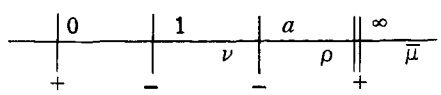
(ii)


(i) → (ii), $t \to it$, $x \to ix$, $y \to iy$.

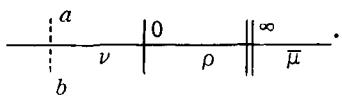
(b) (i)


(a) (i) – (b) (i), $t \to x$, $x \to iy$, $y \to t$,

$\hat{B} = M_{01}^2 + aM_{02}^2$.

(ii) 

(b) (i) − (b) (ii), $t \to it$, $x \to ix$, $y \to iy$.

(c) (i) 

Here as usual $a = \alpha + i\beta$, $b = \alpha - i\beta$, $\alpha, \beta \in \mathbb{R}$. The pentaspherical coordinates are given by

$$\sigma x_1^2 = \bar{\mu}, \quad 2\sigma x_1 x_2 = 1, \tag{3.53}$$
$$\sigma x_3^2 = \frac{(\nu - a)(\rho - a)}{a(a - b)}, \quad \sigma x_4^2 = \frac{(\nu - b)(\rho - b)}{b(b - a)}, \quad \sigma x_5^2 = \frac{\nu\rho}{ab},$$

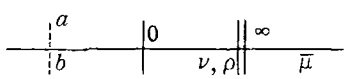
and the three space coordinates are given by

$$(t + ix) = \frac{x_3}{x_1} = \sqrt{2}\, r \frac{(\nu - a)(\rho - a)}{a(a - b)}, \quad y = \frac{ix_5}{x_1} = r\frac{\nu\rho}{ab} \tag{3.54}$$

with $r^2 = 1/\bar{\mu}$. The basis operators are as in (3.50) with $P(\lambda) = \lambda(\lambda - a)(\lambda - b)$. In particular

$$\hat{B} = \alpha(M_{12}^2 - M_{02}^2) - \beta(M_{02}M_{12} + M_{12}M_{02}), \tag{3.55}$$

and the separation equations are as in (3.51), (3.52) with appropriate changes in $P(\lambda)$ as above.

(ii) 

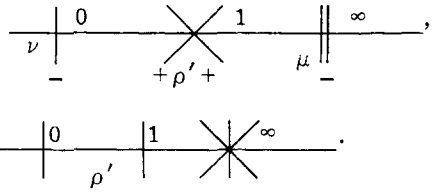
(c) (i) − (c) (ii), $t \to it$, $x \to ix$, $y \to iy$.

The three space parametrizations corresponding to (a), (b), and (c) in this subsection are recognized as the three possible Lamé bases for the group O(2, 1). These bases have been discussed by the authors[9] and Macfadyen and Winternitz.[10] The results presented in this subsection give the parametrization of these bases inside and outside the cone $t^2 - x^2 - y^2 = 0$.

3. Degenerate systems of the type [21(11)]

The coordinate systems of this type are chosen in such a way that the parameters e_i are $e_1 = \infty$, with the remaining free parameters 1 and 0

(a) [21(11)]



The pentaspherical coordinates are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = -\mu - \nu + 1, \tag{3.56}$$
$$\sigma x_3^2 = -(\mu - 1)(\nu - 1)(1 - \rho'),$$
$$\sigma x_4^2 = -(\mu - 1)(\nu - 1)\rho', \quad \sigma x_5^2 = \mu\nu.$$

The three space coordinates are given by

$$t = ix_5/x_1, \quad x = x_3/x_1, \quad y = x_4/x_1. \tag{3.57}$$

With $\mu = \cosh^2 A$, $\nu = -\sinh^2 B$, $\rho' = \sin^2\phi$, the three space coordinates assume the form

$$t = \cosh A \sinh B, \quad x = \sinh A \cosh B \cos\phi, \quad y = \sinh A \cosh B \sin\phi. \tag{3.58}$$

The separation equations have the form

$$\frac{1}{\cosh A}\frac{d}{dA}\cosh A \frac{dE_1}{dA} + \left(\frac{m^2}{\cosh^2 A} + K\right) E_1 = 0, \tag{3.59}$$

$$\frac{1}{\sinh B}\frac{d}{dB}\frac{dE_2}{dB} + \left(\frac{-m^2}{\sinh^2 B} + K\right) E_2 = 0, \tag{3.60}$$

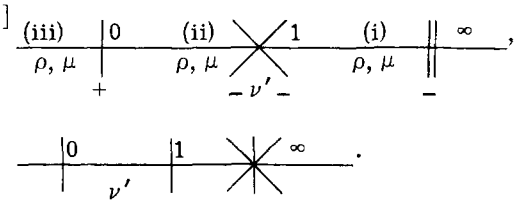
and

$$\frac{d^2E_3}{d\phi^2} + m^2 E_3 = 0, \tag{3.61}$$

where (*) has the solution $E_1(A)E_2(B)E_3(\phi)$. The basis operators $\hat{A}$ and $\hat{B}$ whose eigenvalues are $-m^2 - K$ and $-m^2$ respectively are

$$\hat{A} = M_{01}^2 + M_{02}^2 + p_1^2 + p_2^2 = P_0^2 - \tfrac{1}{2}(P_0K_0 + K_0P_0 + 1), \tag{3.62}$$
$$\hat{B} = M_{12}^2.$$

The separation equations (3.59), and (3.60) can be identified with Legendré's equation. The linearly independent solutions of (3.60) are $P_j^m(\cosh B)$, $Q_j^m(\cosh B)$, where $K = -j(j + 1)$. The solutions of (3.59) can be obtained from those of (3.60) by putting $B \to A + i\pi/2$.

(b) [21(11)] 

There are three cases to consider here as indicated in the above diagram. We put

(i) $\mu = \cosh^2 A$, $\rho = \cosh^2 B$,

(ii) $\mu = \cos^2\alpha$, $\rho = \cos^2\beta$,

(iii) $\mu = -\sinh^2 A$, $\rho = -\sinh^2\beta$,

with $\nu' = \cos^2\phi$ in all cases. The resulting three space variables are in these cases:

(i) $t = \cosh A \cosh B$, $x = \sinh A \sinh B \cos\phi$, $y = \sinh A \sinh B \sin\phi$,

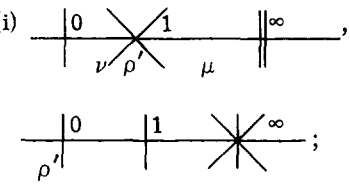
(ii) $t = \cos\alpha \cos\beta$, $x = \sin\alpha \sin\beta \cos\phi$, $y = \sin\alpha \sin\beta \sin\phi$, (3.63)

(iii) $t = \sinh A \sinh B$, $x = \cosh A \cosh B \cos\phi$, $y = \cosh A \cosh B \sin\phi$.

The basis defining operators are

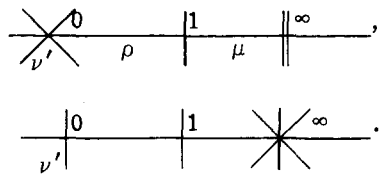
$$\hat{A} = M_{01}^2 + M_{02}^2 - P_1^2 - P_2^2 = -P_0^2 - \tfrac{1}{2}(P_0K_0 + K_0P_0 + 1), \tag{3.64}$$
$$\hat{B} = M_{12}^2.$$

(c) [21(11)]

(i) 

by putting $\mu=\cosh^2A$, $\nu=\cos^2\alpha$, $\rho'=-\sinh^2B$, this system gives the three space coordinates

$$t=\sin\alpha\sinh A\sinh B,\quad x=\cos\alpha\cosh A,\quad y=\sin\alpha\sinh A\cosh B. \tag{3.65}$$

(ii) 

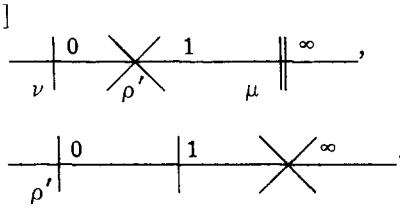
With $\mu=\cosh^2A$, $\rho=\cos^2\alpha$, and $\nu'=-\sinh^2B$ the three space coordinates are

$$t=\sin\alpha\cosh A\sinh B,\quad x=\cos\alpha\sinh A,\quad y=\sin\alpha\cosh A\cosh B. \tag{3.66}$$

The basis defining operators for these coordinate systems are

$$\hat A=M_{01}^2-M_{12}^2+P_0^2-P_2^2,\quad \hat B=M_{02}^2. \tag{3.67}$$

(d) [21(11)]



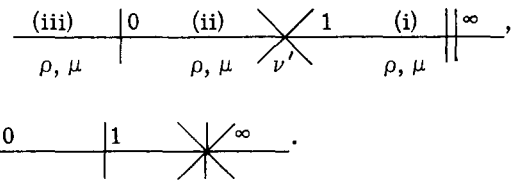
With $\mu=\cosh^2A$, $\nu=-\sinh^2B$ and $\rho'=-\sinh^2C$ the three space coordinates become

$$t=\sinh A\cosh B\cosh C,\quad x=\cosh A\sinh B,\quad y=\sinh A\cosh B\sinh C. \tag{3.68}$$

The basis defining operators of this system are

$$\hat A=M_{01}^2-M_{12}^2+P_0^2-P_2^2,\quad \hat B=M_{02}^2. \tag{3.69}$$

(e) [21(11)]



There are three possible cases to consider here:

(i) $\mu=\cosh^2A$, $\rho=\cosh^2B$,

(ii) $\mu=\cos^2\alpha$, $\rho=\cos^2\beta$,

(iii) $\mu=-\sinh^2A$, $\rho=-\sinh^2B$,

where in all cases $\nu'=-\sinh^2C$. The resulting coordinate systems in three space are

(i) $t=\sinh A\sinh B\cosh C$, $x=\cosh A\cosh B$, $y=\sinh A\sinh B\sinh C$,

(ii) $t=\cos\alpha\cos\beta\cosh C$, $x=\sin\alpha\sin\beta$, $y=\cos\alpha\cos\beta\sinh C$, (3.70)

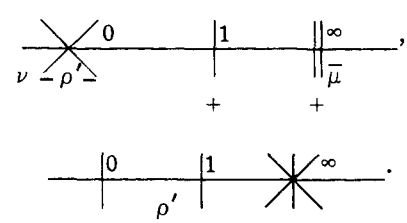
(iii) $t=\cosh A\cosh B\cosh C$, $x=\sinh A\sinh B$, $y=\cosh A\cosh B\sinh C$

and the basis operators are

$$\hat A=M_{01}^2-M_{12}^2-P_0^2+P_2^2,\quad \hat B=M_{02}^2. \tag{3.71}$$

4. Coordinate systems containing a radial coordinate in three space and derivable from the configuration [21(11)]

These systems are derivable in exactly the same manner as those of subsection 2.

(a) (i) 

The pentaspherical coordinates are given by

$$\sigma x_1^2=\bar\mu,\quad 2\sigma x_1x_2=-1,\quad \sigma x_3^2=(1-\nu),\quad \sigma x_4^2=\nu(1-\rho'),\quad \sigma x_5^2=\nu\rho', \tag{3.72}$$

With $\nu=-\sinh^2A$ and $\rho'=\sin^2\alpha$, $\bar\mu=1/r^2$ these formulas give the three space coordinates

$$t=x_3/x_1=r\cosh A,\quad x=ix_4/x_1=r\sinh A\cos\alpha,\quad y=-ix_5/x_1=r\sinh A\sin\alpha. \tag{3.73}$$

These are just the familiar polar coordinates inside the cone $t^2-x^2-y^2=0$. The basis operators are

$$\hat A=\tfrac14-D^2,\quad \hat B=M_{12}^2. \tag{3.74}$$

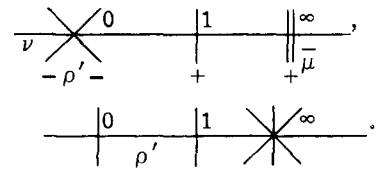
The separation equations are

$$\frac{d^2E_1}{dr^2}+\frac1r\frac{dE_1}{dr}-\frac{j(j+1)}{r^2}E_1=0,$$

$$\frac{1}{\sinh A}\frac{d}{dA}\sinh A\frac{dE_2}{dA}-\left(\frac{m^2}{\sinh^2A}+j(j+1)\right)E_2=0, \tag{3.75}$$

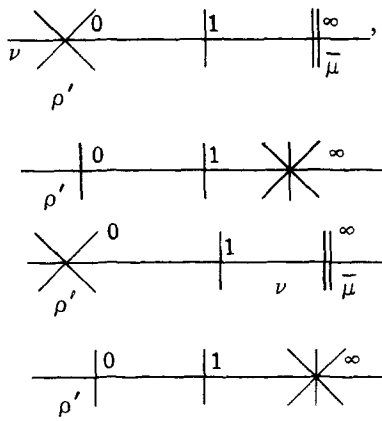
$$\frac{d^2E_3}{d\phi^2}+m^2E_3=0,$$

where $E_1(r)E_2(A)E_3(\phi)$ is a solution of (*). The second of these equations is just a form of the Legendre equation with solutions $P_j^m(\cosh A)$, $Q_j^m(\cosh A)$. The other two equations have the elementary solutions $E_1=r^j$, r^{-j-1}, and $E_3=\exp(\pm im\phi)$

(ii) 

The three space coordinates for this second configuration are obtained from (3.73) via the transformation $\cosh A\rightleftharpoons\sinh A$.

(b) (i)



The system given by diagram (i) yields the three space coordinates

$$t = r\sinh A \cosh B, \quad x = r\sinh A \sinh B, \quad y = r\cosh A, \tag{3.76}$$

where

$$\nu = -\sinh^2 A, \quad \rho' = \sinh^2 B, \text{ and } \bar{\mu} = 1/r^2.$$

The defining operators are

$$\hat{A} = \tfrac{1}{4} - D^2, \quad \hat{B} = M_{01}^2. \tag{3.77}$$

The coordinates (3.76) are the familiar hyperbolic coordinates inside the cone $t^2 - x^2 - y^2 = 0$. For diagram (ii) the only change is in the three space variables subjected to the transformation $\sinh A \rightleftharpoons \cosh A$.

5. Coordinate systems corresponding to the configuration [(21) 11]

To obtain the expression for the pentaspherical coordinates corresponding to [(21)11] requires the substitution

$$e_3 = e_1 + \epsilon, \quad \lambda = e_1 + \epsilon + \epsilon^2\lambda' \tag{3.78}$$

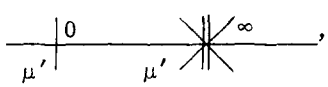
into (1.37). Here ϵ is a first order quantity and for definiteness we may take $\lambda = \mu$. The resulting expression for the pentaspherical coordinates is

$$\begin{aligned}
\sigma x_1^2 &= -\frac{(\rho - e_1)(\nu - e_1)}{(e_4 - e_1)(e_5 - e_1)}, \\
2\sigma x_1 x_2 &= -\frac{\partial}{\partial e_1}\left[\frac{(\rho - e_1)(\nu - e_1)}{(e_4 - e_5)(e_5 - e_1)}\right] - \frac{(\rho - e_1)(\nu - e_1)}{(e_4 - e_1)(e_5 - e_1)}\mu', \\
\sigma x_3^2 &= \frac{(\rho - e_1)(\nu - e_1)}{(e_4 - e_1)(e_5 - e_1)}\mu', \quad \sigma x_4^2 = \frac{(\rho - e_4)(\nu - e_4)}{(e_1 - e_4)^2(e_5 - e_4)} \\
\sigma x_5^2 &= \frac{(\rho - e_5)(\nu - e_5)}{(e_1 - e_5)^2(e_4 - e_5)}.
\end{aligned} \tag{3.79}$$

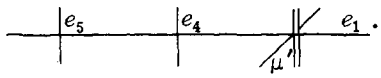
The resulting coordinate curve for the coordinate μ' is

$$x_1^2 + x_3^2/\mu' = 0. \tag{3.80}$$

The diagram corresponding to such a curve is



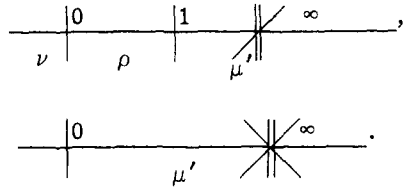
while the diagram representing the coordinate curves of the curvilinear coordinates ρ and ν is



The inequivalent classes of coordinates of this type are now given. In each case the e_i can be standardized as usual to be

$$e_1 = \infty, \quad e_4 = 1, \quad e_5 = 0. \tag{3.81}$$

(a) [(21)11]



From formulas (3.79) the pentaspherical coordinates for the coordinate system are given by

$$\begin{aligned}
&\sigma x_1^2 = 1, \quad 2\sigma x_1 x_2 = 1 - \nu - \rho - \mu', \\
&\sigma x_3^2 = \mu', \quad \sigma x_4^2 = (1 - \rho)(\nu - 1), \quad \sigma x_5^2 = \nu\rho.
\end{aligned} \tag{3.82}$$

With $\nu = -\sinh^2 A$ and $\rho = \sin^2\alpha$ this gives the three space coordinates

$$\begin{aligned}
&t = ix_3/x_1 = K, \quad x = x_4/x_1 = \cosh A \cos\alpha, \\
&y = x_5/x_1 = \sinh A \sin\alpha.
\end{aligned} \tag{3.83}$$

Here $K = \sqrt{\mu'}$. The separation equations have the form

$$\frac{d^2E_1}{dA^2} + (-\tau^2\sinh^2 A + V)E_1 = 0, \tag{3.84}$$

$$\frac{d^2E_2}{d\alpha^2} + (-\tau^2\sin^2\alpha - V)E_2 = 0, \tag{3.85}$$

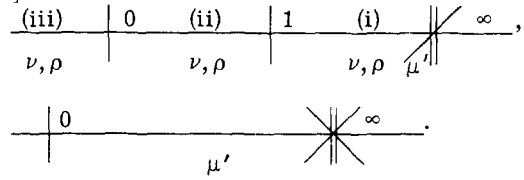
$$\frac{d^2E_3}{dK^2} + \tau^2 E_3 = 0, \tag{3.86}$$

and the basis operators $\hat{A}$, $\hat{B}$ whose eigenvalues are the separation constants V, $-\tau^2$ respectively are

$$\hat{A} = M_{12}^2 - P_2^2, \quad \hat{B} = P_0^2. \tag{3.87}$$

Equations (3.84) and (3.85) are easily seen to be forms of Mathieu's equation. Here as usual $\psi = E_1(A)E_2(\alpha)E_3(K)$ is a solution of (*).

(b) [(21)11]



There are three cases to consider here. If we choose

(i) $\rho = \cosh^2 A, \quad \nu = \cosh^2 B,$

(ii) $\rho = \cos^2\alpha, \quad \nu = \cos^2\beta,$

(iii) $\rho = \sinh^2 A, \quad \nu = -\sinh^2 B,$

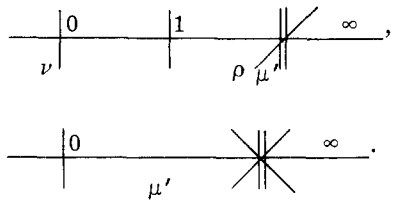
with $\mu' = K^2$ in all cases, then the resulting three space coordinates are

$$\begin{aligned}
&\text{(i)} \quad t = \sinh A \sinh B, \quad x = K, \quad y = \cosh A \cosh B, \\
&\text{(ii)} \quad t = \sin\alpha \sin\beta, \quad x = K, \quad y = \cos\alpha \cos\beta, \\
&\text{(iii)} \quad t = \cosh A \cosh B, \quad x = K, \quad y = \sinh A \sinh B.
\end{aligned} \tag{3.88}$$

The basis operators in this case are

$$\hat{A} = M_{02}^2 - P_0^2, \quad \hat{B} = P_1^2. \tag{3.89}$$

(c) [(21)11]



If we put $\rho = \cosh^2 A$, $\nu = -\sinh^2 B$, and $\mu' = K^2$, then the three space coordinates are

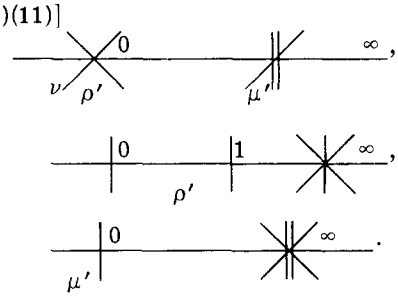
$$t = \sinh A \cosh B, \quad x = K, \quad y = \cosh A \sinh B \tag{3.90}$$

with basis operators

$$\hat{A} = M_{02}^2 + P_0^2, \quad \hat{B} = P_1^2. \tag{3.91}$$

6. Coordinate systems corresponding to the configuration [(21)(11)] and [(21) 11]

(a) [(21)(11)]



The pentaspherical coordinates are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = \mu' - \nu,$$
$$\sigma x_3^2 = -\mu', \quad \sigma x_4^2 = \nu(1-\rho'), \quad \sigma x_5^2 = \nu\rho'. \tag{3.92}$$

Set

$$\nu = -r^2, \quad \rho' = \sin^2\phi, \text{ and } \mu' = -K^2.$$

The corresponding three space coordinates are given by

$$t = x_3/x_1 = K, \quad x = ix_4/x_1 = r\cos\phi,$$
$$y = ix_5/x_1 = r\sin\phi. \tag{3.93}$$

The separation equations are

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dE_1}{dr}\right) - \left(\frac{m^2}{r^2} + S^2\right)E_1 = 0, \tag{3.94}$$

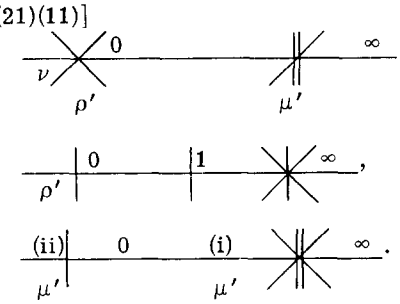
$$\frac{d^2E_2}{d\phi^2} + m^2E_2 = 0, \quad \frac{d^2E_3}{dK^2} + S^2E_3 = 0. \tag{3.95}$$

The corresponding basis defining operators $\hat{A}$, $\hat{B}$ with eigenvalues $-m^2$, $-S^2$ respectively are

$$\hat{A} = M_{12}^2, \quad \hat{B} = P_0^2. \tag{3.96}$$

Equation (3.94) is a form of Bessel's equation.

(b) [(21)(11)]



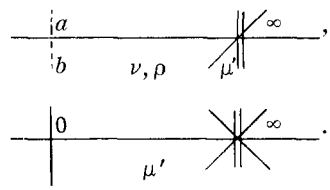
For the case (i) with $\mu' = K^2$, $\rho' = -\sinh^2 A$, and $\nu = -r^2$, the three space coordinates are

$$t = r\sinh A, \quad x = K, \quad y = r\cosh A. \tag{3.97}$$

For (ii) the three space coordinates are as in (3.97) with $\cosh A \rightleftharpoons \sinh A$. The basis operators for this system are

$$\hat{A} = M_{02}^2, \quad \hat{B} = P_1^2 \tag{3.98}$$

(c) [(21)$\widehat{11}$]



This is the only coordinate system of this type. The pentaspherical coordinates are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = b + a - \nu - \rho - \mu',$$
$$\sigma x_3^2 = \mu', \tag{3.99}$$
$$\sigma x_4^2 = \frac{(\nu-a)(\rho-a)}{(b-a)}, \quad \sigma x_5^2 = \frac{(\nu-b)(\rho-b)}{(a-b)}.$$

This corresponds to a choice of three space coordinates:

$$(x + it) = \sqrt{2}i\,x_4/x_1, \quad y = ix_3/x_1 = K, \tag{3.100}$$

where $K = \sqrt{\mu'}$. The separation equations have the form (3.51) with $P(\lambda) = (\lambda - a)(\lambda - b)$ and

$$\frac{d^2E_3}{dK^2} + BE_3 = 0. \tag{3.101}$$

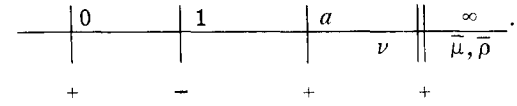
The basis operators $\hat{A}$, $\hat{B}$ are

$$\hat{A} = -M_{01}^2 + \alpha(P_1^2 - P_0^2) - \beta P_0 P_1, \quad \hat{B} = P_2^2. \tag{3.102}$$

7. Coordinate systems on the cone $t^2 - x^2 - y^2 = 0$ that can be obtained from the configuration [2111] and its degenerate forms

The method for obtaining coordinate systems on the cone is similar to that for obtaining the coordinate systems with a radial coordinate in three space. The method is illustrated for the first coordinate system of this subsection.

(a)



The pentaspherical coordinates for such a diagram are obtained from (1.37) by putting $\mu = e_1 + \bar{\mu}$, $\rho = e_1 + \bar{\rho}$ and making the substitutions $x_i \to e_i x_i$ $(i = 3, 4, 5)$ and $x_2 \to e_1^2 x_2$, $x_1 \to x_1$. Then the pentaspherical coordinates assume the form

$$\sigma x_1^2 = \bar{\mu}\bar{\rho}, \quad 2\sigma x_1 x_2 = 0, \quad \sigma x_3^2 = (\nu - a)/a(a-1),$$
$$\sigma x_4^2 = (1-\nu)/(a-1), \quad \sigma x_5^2 = \nu/a. \tag{3.103}$$

The corresponding choice of three space variables is

$$t = r\sqrt{(\nu-1)/(a-1)}, \quad x = r\sqrt{\nu/a},$$
$$y = r\sqrt{(\nu-a)/a(a-1)}, \tag{3.104}$$

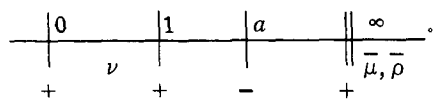
where $r^2 = 1/\bar{\mu}\bar{\rho}$. The separation equations here are given by (3.41) for the variable ν and

$$r^2\frac{d^2E_2}{dr^2}+r\frac{dE_2}{dr}-j(j+1)E_2=0. \tag{3.105}$$

The basis defining operators $\hat{A}$, $\hat{B}$ corresponding to the separation constants $j(j+1)$ and B are

$$\hat{A}=\tfrac{1}{4}-D^2,\quad \hat{B}=M_{02}^2-aM_{12}^2. \tag{3.106}$$

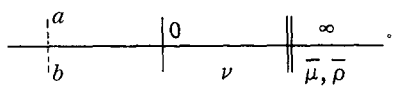
(b)



The properties of this system are obtained from those of (a) via

(a) → (b), $t\to ix$, $x\to y$, $y\to it$.

(c)



The pentaspherical coordinates are

$$\sigma x_1^2=\bar{\mu}\bar{\rho},\quad 2\sigma x_1x_2=0,\quad \sigma x_3^2=(\nu-a)/a(a-b),$$
$$\sigma x_4^2=(\nu-b)/b(b-a),\quad \sigma x_5^2=\nu/ab. \tag{3.107}$$

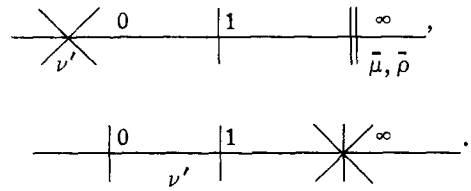
The corresponding three space coordinates are

$$(x+it)=i\sqrt{2}x_4/x_1,\quad y=ix_5/x_1. \tag{3.108}$$

The separation equations are given by (3.41) with $P(\lambda)=\lambda(\lambda-a)(\lambda-b)$ and (3.105), where as usual $r^2=1/\bar{\mu}\bar{\rho}$. The basis defining operators $\hat{A}$, $\hat{B}$ are

$$\hat{A}=\tfrac{1}{4}-D^2,\quad \hat{B}=\alpha(M_{12}^2-M_{02}^2)-\beta(M_{02}M_{12}+M_{12}M_{02}). \tag{3.109}$$

(d)



The pentaspherical coordinates are

$$\sigma x_1^2=\bar{\mu}\bar{\rho},\quad 2\sigma x_1x_2=0,\quad \sigma x_3^2=-1,$$
$$\sigma x_4^2=(1-\nu')\quad \sigma x_5^2=\nu'. \tag{3.110}$$

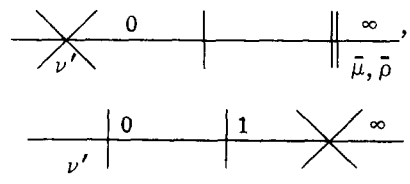
The three space coordinates on the cone are

$$t=r,\quad x=r\cos\phi,\quad y=r\sin\phi,$$

where $1/\bar{\mu}\bar{\rho}=r^2$, $\nu'=\cos^2\phi$. The separation equations are (3.105) and the third equation of (3.75). The basic defining operators are clearly

$$\hat{A}=\tfrac{1}{4}-D^2,\quad \hat{B}=M_{12}^2. \tag{3.111}$$

(e)



The three space coordinates are

$$t=r\cosh A,\quad x=r\sinh A,\quad y=\pm r$$

with the basis defining operators

$$\hat{A}=\tfrac{1}{4}-D^2,\quad \hat{B}=M_{01}^2. \tag{3.112}$$

C. The configurations [311], [3$\hat{1}\hat{1}$], and their degenerate forms

Here we give the configurations of the form [311] and [3$\hat{1}\hat{1}$] which are inequivalent with respect to the by now familiar equivalence relation. It is possible to standardize the parameters e_i such that

$$e_1=\infty,\quad e_4=1,\quad e_5=0, \tag{3.113}$$

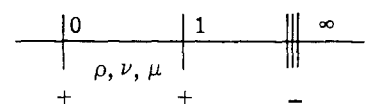
where, of course, e_1 is the parameter associated with the invariant factor index 3. The two operators whose eigenvalues are the separation constants are given as in (3.5) with $f(\lambda)=\lambda(\lambda-1)$. The separation equations with the above choice of the e_i are

$$\sqrt{f(\lambda)}\frac{d}{d\lambda}\left(\sqrt{f(\lambda)}\frac{dE_i}{d\lambda}\right)-(A\lambda+B)E_i=0 \tag{3.114}$$

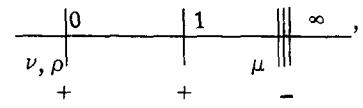
with $f(\lambda)$ as above and $E_1(\mu)E_2(\nu)E_3(\rho)$ as the separable solution of (*). For the configuration [311] the separation equations are as in (3.114) with $f(\lambda)=(\lambda-a)(\lambda-b)$, where as usual $a=b^*=\alpha+i\beta$, $\alpha,\beta\in\mathbb{R}$. We now list the inequivalent systems of this type.

(a) [311]

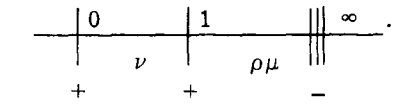
(i)



(ii)



(iii)



The pentaspherical coordinates are given by

$$\sigma x_1^2=-1,\quad 2\sigma x_1x_2=\mu+\nu+\rho,$$
$$\sigma(2x_1x_3+x_2^2)=\nu+\mu+\rho-(\mu\nu+\mu\rho+\nu\rho)-1, \tag{3.115}$$
$$\sigma x_4^2=-(\mu-1)(\nu-1)(\rho-1),\quad \sigma x_5^2=\mu\nu\rho.$$

The three space coordinates are given by

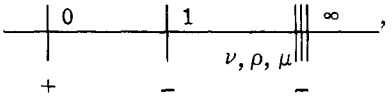
$$2t=2x_2/x_1=-\mu-\nu-\rho, \tag{3.116}$$
$$x=ix_4/x_1=\sqrt{(\mu-1)(\nu-1)(1-\rho)},\quad y=x_5/x_1=\sqrt{\mu\nu\rho}.$$

The basis operators are

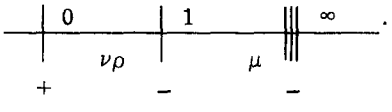
$$\hat{A}=P_0M_{20}+M_{20}P_0-P_1M_{02}-M_{02}P_1+P_0^2+P_1^2,$$
$$\hat{B}=-M_{12}^2+P_2^2+P_2M_{02}+M_{02}P_2. \tag{3.117}$$

(b) [311]

(i)



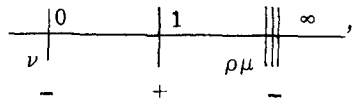
(ii)



The properties of this system can be deduced from those of (a) via the transformation

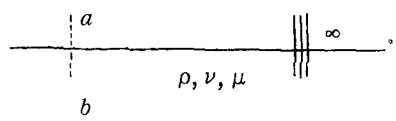
(a) → (b), $t \to y$, $x \to ix$, $y \to t$.

(c) [311]



(a) → (b), $t \to it$, $x \to ix$, $y \to iy$.

(d) [3$\hat{1}\hat{1}$]



The pentaspherical coordinates are given by

$$\begin{aligned} &\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = \mu + \nu + \rho, \\ &\sigma(2x_1x_3 + x_2^2) = \mu\rho + \nu\rho + \nu\mu - (a+b)(\mu+\nu+\rho) \\ &\qquad + a^2 + ab + b^2, \\ &\sigma x_4^2 = (\mu - a)(\rho - a)(\nu - a)/(b - a), \\ &\sigma x_5^2 = (\mu - b)(\rho - b)(\nu - b)/(a - b), \end{aligned} \tag{3.118}$$

where $a = \alpha + i\beta$, $b = \alpha - i\beta$, $\alpha, \beta \in \mathbf{R}$. The three space coordinates are given by

$$(x + it) = i\sqrt{2}\, x_4/x_1, \quad y = -x_2/x_1 \tag{3.119}$$

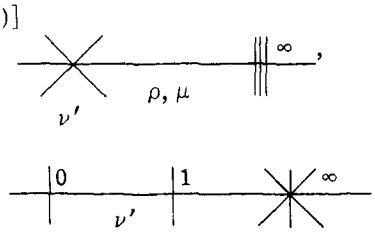
with the basis operators given by

$$\begin{aligned} \hat{A} &= M_{12}P_1 + P_1M_{12} + M_{02}P_0 + P_0M_{02} + 2\alpha P_2^2 \\ &\quad + \alpha(P_0^2 - P_1^2) - 2\beta P_0P_1, \\ \hat{B} &= \alpha(M_{12}P_1 + P_1M_{12} + M_{02}P_0 + P_0M_{02}) + \beta(M_{02}P_1 + P_1M_{02} \\ &\quad - M_{12}P_0 - P_0M_{12}) + (\alpha^2 + \beta^2)P_2^2 - M_{01}^2 \\ &\quad + (\alpha^2 - \beta^2)(P_1^2 - P_0^2) - 4\alpha\beta P_0P_1. \end{aligned} \tag{3.120}$$

1. Degenerate systems of type [3(11)]

The coordinate systems of this type are chosen such that the free parameters e_1 and e_5 are ∞ and 0 respectively.

(a) [3(11)]



The pentaspherical coordinates are given by

$$\begin{aligned} &\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = -\mu - \rho - 1, \\ &\sigma(2x_1x_3 + x_2^2) = -\mu\rho, \quad \sigma x_4^2 = \mu\rho(1 - \nu'), \\ &\sigma x_5^2 = \mu\rho\nu'. \end{aligned} \tag{3.121}$$

The three space coordinates are given by the formulas

$$t = \pm x_2/x_1, \quad x = ix_4/x_1, \quad y = ix_5/x_1. \tag{3.122}$$

Translating t by $\pm\frac{1}{2}$ and putting $\mu = \xi^2$, $\rho = \eta^2$, and $\nu' = \sin^2\alpha$. We obtain the more familiar form,

$$t = \pm\tfrac{1}{2}(\xi^2 + \eta^2), \quad x = \xi\eta\cos\alpha, \quad y = \xi\eta\sin\alpha. \tag{3.123}$$

The separation equations for this system assume the form

$$\frac{d^2E_i}{d\lambda^2} + \frac{1}{\lambda}\frac{dE_i}{d\lambda} + \left(q^2 - \frac{m^2}{\lambda^2}\right)E_i = 0, \tag{3.124}$$

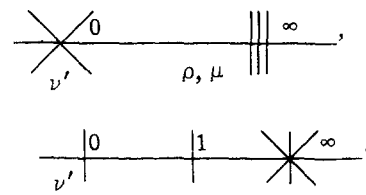
where $\lambda = \xi, \eta$ and $i = 1, 2$ respectively, and

$$\frac{d^2E_3}{d\alpha^2} + m^2E_3 = 0.$$

Equation (3.124) is Bessel's equations with linearly independent solutions $J_m(q\lambda)$, $Y_m(q\lambda)$. The solutions of the third equation are $E_3 = \exp(\pm im\alpha)$. The basis operators $\hat{A}$, $\hat{B}$ whose eigenvalues are q^2 and $-m^2$, respectively, are

$$\begin{aligned} &\hat{A} = M_{01}P_1 + P_1M_{01} + M_{02}P_2 + P_2M_{02} = P_0D + DP_0, \\ &\hat{B} = M_{12}^2. \end{aligned} \tag{3.125}$$

(b) [3(11)]



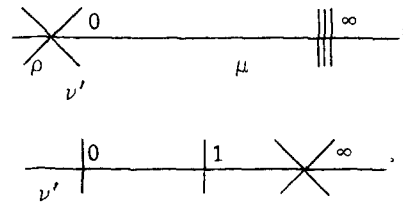
The three space coordinates in this case are

$$\begin{aligned} &y = (x_2/x_1 - \tfrac{1}{2}) = \pm\tfrac{1}{2}(\xi^2 + \eta^2), \quad t = ix_4/x_1 = \xi\eta\cosh A, \\ &x = x_5/x_1 = \xi\eta\sinh A. \end{aligned} \tag{3.126}$$

where $\mu = \xi^2$, $\rho = \eta^2$, and $\nu' = -\sinh^2 A$. The resulting basis operators are

$$\begin{aligned} &\hat{A} = M_{02}P_0 + P_0M_{02} + M_{12}P_1 + P_1M_{12}, \\ &\hat{B} = M_{01}^2. \end{aligned} \tag{3.127}$$

(c)



This system corresponds to the choice of three space coordinates

$$\begin{aligned} &t = ix_5/x_1 = \xi\eta\sinh A, \quad x = (x_2/x_1 - \tfrac{1}{2}) = \tfrac{1}{2}(\xi^2 - \eta^2) \\ &y = x_4/x_1 = \xi\eta\cosh A, \end{aligned} \tag{3.128}$$

$\mu = \xi^2$, $\rho = -\eta^2$, and $\nu' = -\sinh A$. The basis operators are

$$\hat{A} = M_{12}P_2 + P_2M_{12} + M_{01}P_0 + P_0M_{01}, \quad \hat{B} = M_{02}^2. \tag{3.129}$$

2. Degenerate systems of type [(31)1]

The formulas for the pentaspherical coordinates corresponding to the degenerate configuration [(31)1] are obtained from these of (1.38) via the substitution

$$e_4 = e_1 + \epsilon, \quad \lambda = e_1 + \epsilon + \epsilon^3\lambda', \tag{3.130}$$

where for definiteness we can take $\lambda = \nu$. The resulting expression for the pentaspherical coordinates is

$$\sigma x_1^2=\frac{(\mu-e_1)(\rho-e_1)}{(e_5-e_1)},\quad 2\sigma x_1x_2=\frac{\partial}{\partial e_1}\left(\frac{(\mu-e_1)(\rho-e_1)}{(e_5-e_1)}\right),\tag{3.131}$$

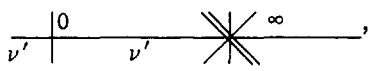
$$\sigma(2x_1x_3+x_2^2)=\frac{1}{2}\frac{\partial^2}{\partial e_1^2}\left(\frac{(\mu-e_1)(\rho-e_1)}{(e_5-e_1)}\right)+\frac{(\mu-e_1)(\rho-e_1)}{(e_5-e_1)}\nu',$$

$$\sigma x_4^2=-\frac{(\mu-e_1)(\rho-e_1)}{(e_5-e_1)}\nu',\quad \sigma x_5^2=\frac{(\mu-e_5)(\rho-e_5)}{(e_1-e_5)^3}.$$

The coordinate curve for thecoordinate ν' is

$$x_1^2+x_4^2/\nu'=0,\tag{3.132}$$

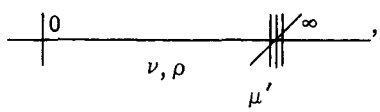
and the diagram corresponding to such a curve is



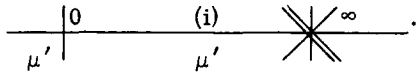
while the diagram representing the coordinate curves of the curvilinear coordinates μ, ρ is

The inequivalent classes of this type are now given. In each case e_1 and e_5 can be taken to be ∞ and 0, respectively.

(a) [(31)1]



(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2=-1,\quad 2\sigma x_1x_2=-\nu-\rho-1,$$
$$\sigma(2x_1x_3+x_2^2)=-\mu'-\nu\rho,\quad \sigma x_4^2=\mu',\quad \sigma x_5^2=\nu\rho.\tag{3.133}$$

The corresponding three space coordinates are

$$\text{(i) } t=ix_5/x_1=\xi\eta,\quad x=ix_4/x_1=k,$$
$$y=(x_2/x_1-\tfrac{1}{2})=\pm\tfrac{1}{2}(\xi^2+\eta^2),$$
$$\text{(ii) } t=\pm\tfrac{1}{2}(x_2/x_1-\tfrac{1}{2})=\pm\tfrac{1}{2}(\xi^2+\eta^2),\tag{3.134}$$
$$x=x_4/x_1=k,\quad y=ix_5/x_1=\xi\eta,$$

where $\rho=\xi^2$, $\nu=\eta^2$, and $k=(|\mu'|)^{1/2}$. The separation equations are

$$\frac{d^2E_i}{d\lambda^2}+(Q-\tau^2\lambda^2)E_i=0\tag{3.135a}$$

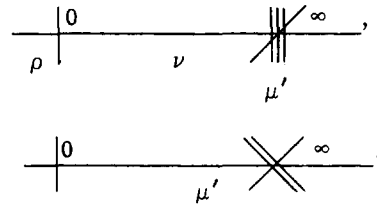
for $\lambda=\xi, \eta$ and $i=1, 2$ respectively, and

$$\frac{d^2E_3}{dk^2}+\tau^2E_3=0.\tag{3.135b}$$

Equation (3.135a) is well known to have solutions expressed as parabolic cylinder functions. The basis operators $\hat{A}$, $\hat{B}$ whose eigenvalues are the separation constants Q and $-\tau^2$, respectively, are

$$\hat{A}=M_{02}P_0+P_0M_{02},\quad \hat{B}=P_1^2.\tag{3.136}$$

(b)



In this case the three space coordinates are given by

$$t=ix_4/x_1=k,\quad x=(x_2/x_1-\tfrac{1}{2})=\tfrac{1}{2}(\xi^2-\eta^2),$$
$$y=ix_5/x_1=\xi\eta\tag{3.137}$$

with $\nu=\xi^2$, $\rho=-\eta^2$, and $k=(\mu')^{1/2}$. The basis operators are

$$\hat{A}=M_{12}P_2+P_2M_{12},\quad \hat{B}=P_0^2.\tag{3.138}$$

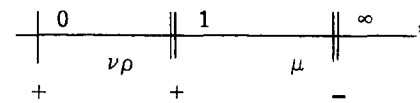
D. The configuration [221] and its degenerate forms

1. Systems of the type [221]

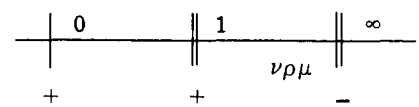
The inequivalent coordinate systems of type [221] are given in the following list. In each case the three free parameters e_1, e_3, and e_5 are standardized to be ∞, 1, and 0, respectively.

(a) [221]

(i)



(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2=-1,\quad 2\sigma x_1x_2=-\mu-\nu-\rho+2,$$
$$\sigma x_3^2=(\mu-1)(\nu-1)(\rho-1),\tag{3.139}$$

$$2\sigma x_3x_5=\mu+\nu+\rho-\mu\nu\rho-2,\quad \sigma x_5^2=\mu\nu\rho.$$

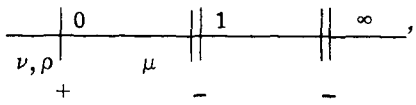
A suitable choice of three space coordinates is

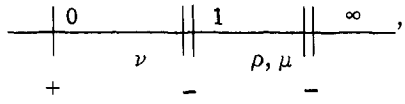
$$(t+x)^2=-x_3^2/x_1^2=(\mu-1)(\nu-1)(\rho-1),$$
$$x^2-t^2=-2x_3x_5/x_1^2=\mu+\nu+\rho-\mu\nu\rho-2,\tag{3.140}$$
$$y=\pm ix_5/x_1=\pm\sqrt{\mu\nu\rho}.$$

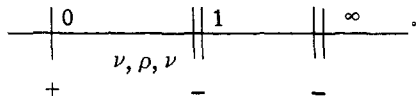
The separation equations have the form (3.51) with $p(\lambda)=\lambda(\lambda-1)^2$, i.e., the associated Legendre equation. The defining operators are

$$\hat{A}=2(P_2^2-P_0^2-P_1P_0)+M_{12}^2-M_{01}^2-M_{02}^2,$$
$$\hat{B}=P_2^2-2M_{02}^2-M_{12}M_{02}-M_{02}M_{12}.\tag{3.141}$$

(b) [221]

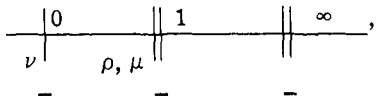
(i) 

(ii) 

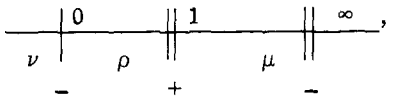
(iii) 

This system is related to (a) via the transformation

$$(a)-(b),\quad t\to ix,\quad x\to it,\quad y\to y.$$

(c) [221] 

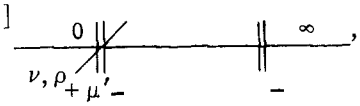
$$(a)-(c),\quad t\to it,\quad x\to ix,\quad y\to iy.$$

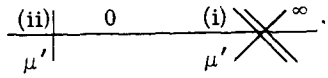
(d) [221] 

$$(a)-(d),\quad t\to x,\quad x\to t,\quad y\to iy.$$

2. Coordinate systems corresponding to the configuration [2(21)]

Here the two free parameters e_1 and e_3 may be taken as 0 and ∞ (not necessarily respectively, as will be seen).

(a) [2(21)] 



The pentaspherical coordinates are given by

$$\sigma x_1^2=-1,\quad 2\sigma x_1x_2=-\nu-\rho,\quad \sigma x_3^2=-\nu\rho,$$
$$2\sigma x_3x_4=\nu+\rho-\mu'\nu\rho,\quad \sigma x_5^2=2\nu\rho\mu'. \tag{3.142}$$

A suitable choice of three space coordinates is

$$(t-x)^2=x_3^2/x_1^2=\nu\rho,$$
$$(t^2-x^2)=2x_3x_4/x_1^2=\mu'\nu\rho-\nu-\rho, \tag{3.143}$$
$$y=\pm ix_5/\sqrt{2}\,x_1=\pm\sqrt{\mu'\nu\rho}\,.$$

The separation equations for this system are of the form (3.51) with $p(\lambda)=\lambda^3$ for the variables ν and ρ. This equation can be related to Bessel's equation. The separation equation in the variable μ' is

$$\mu'^{1/2}\frac{d}{d\mu'}\left(\mu'^{1/2}\frac{dE_3}{d\mu'}\right)+BE_3=0. \tag{3.144}$$

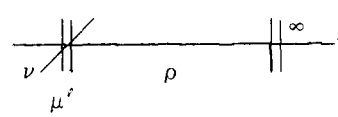
The basis operators $\hat{A}$, $\hat{B}$ are

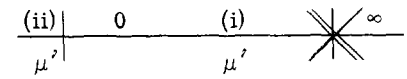
$$\hat{A}=M_{01}^2-(P_0+P_1)^2,\quad \hat{B}=(M_{12}-M_{20})^2. \tag{3.145}$$

For (ii) the results follow from (i) via the transformation

$$(i)-(ii),\quad t\to x,\quad x\to t,\quad y\to iy.$$

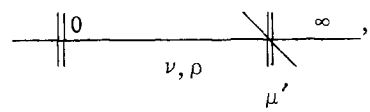
This does not change the operators $\hat{A}$ and $\hat{B}$ but gives new expressions for the three space coordinates.

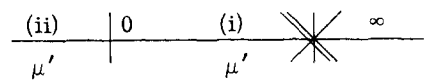
(b) [2(21)] 



(b) (i) and (b) (ii) are obtained from (a) (i) and (a) (ii), respectively, via the transformation

$$t\to it,\quad x\to ix,\quad y\to iy\,.$$

(c) [2(21)] 



The pentaspherical coordinates in this case are given by

$$\sigma x_1^2=\nu\rho,\quad 2\sigma x_1x_2=\nu+\rho,\quad \sigma x_3^2=-1,$$
$$2\sigma x_3x_4=\mu'-\nu-\rho,\quad \sigma x_5^2=\mu'. \tag{3.146}$$

For (i) a suitable choice of three space coordinates is

$$(t-x)^2=-x_1^2/x_3^2=\nu\rho, \tag{3.147}$$
$$(x^2-t^2)=-2x_1x_2/x_3^2=\nu+\rho,\quad y=\pm ix_5/x_3=\pm(\mu')^{1/2}.$$

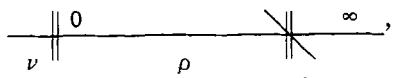
The separation equations for this system are of the form (3.51) with $p(\lambda)=\lambda^2$, $\lambda=\nu,\rho$. This is a form of Bessel's equation. For the variable μ' the equation has the form (3.144) and the basis defining operators $\hat{A}$, $\hat{B}$ are

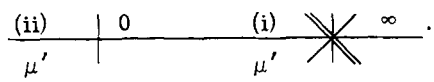
$$\hat{A}=M_{01}^2-(P_0+P_1)^2,\quad \hat{B}=P_2^2. \tag{3.148}$$

The corresponding properties for (ii) are obtained from those of (i) via the transformation

$$t \to x, \quad x \to t, \quad y \to iy.$$

(d) [2(21)]



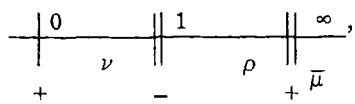


(d) (i) and (d) (ii) are obtained from (c) (i) and (c) (ii), respectively, via the transformation

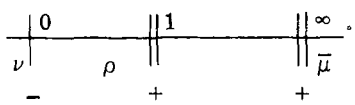
$$t \to it, \quad x \to ix, \quad y \to iy.$$

3. Coordinate systems having a radial coordinate in three space and derivable from the configurations [221] and [2(21)]

(a) (i)



(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \bar{\mu}, \quad 2\sigma x_1 x_2 = 1, \quad \sigma x_3^2 = (\nu - 1)(\rho - 1),$$
$$2\sigma x_3 x_4 = -1 - \nu\rho, \quad \sigma x_5^2 = \nu\rho. \tag{3.149}$$

For the case (i) a suitable choice of three space coordinates is given by the equations

$$(t + x)^2 = x_3^2/x_1^2 = r^2(\nu - 1)(1 - \rho),$$
$$y^2 = x_5^2/x_1^2 = r^2\nu\rho, \quad x^2 + y^2 - t^2 = 2x_2/x_1 = r^2, \tag{3.150}$$

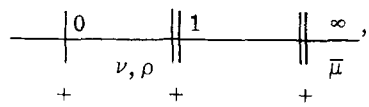
where $r^2 = 1/\bar{\mu}$. The separation equations have the form (3.51) with $p(\lambda) = (\lambda - 1)^2\lambda$ and $\lambda = \rho, \nu$ and the first of equations (3.75) for the radial coordinate. The equations in ρ and ν are associated Legendre function equations. The basis defining operators $\hat{A}$, $\hat{B}$ are

$$\hat{A} = \tfrac{1}{4} - D^2, \quad \hat{B} = 2M_{12}^2 + M_{12}M_{20} + M_{20}M_{12}. \tag{3.151}$$

System (ii) is related to (i) via the transformation

$$t \to it, \quad x \to ix, \quad y \to iy.$$

(b) (i)

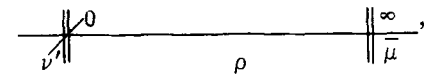


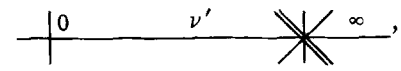
(ii)

The coordinate systems (b) (i) and (b) (ii) are related to (a) (i) and (a) (ii), respectively, via the transformation

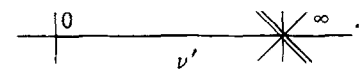
$$t \to ix, \quad x \to it, \quad y \to y.$$

(c) (i)





(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \bar{\mu}, \quad 2\sigma x_1 x_2 = 1, \quad \sigma x_3^2 = \rho,$$
$$2\sigma x_3 x_4 = -\nu'\rho - 1, \quad \sigma x_5^2 = \nu'\rho. \tag{3.152}$$

A suitable choice of three space coordinates for system (i) is

$$x - t = x_3/x_1 = re^a, \quad x^2 - t^2 = -2x_3x_4/x_1^2,$$
$$= r^2(1 + s^2e^{2a}), \quad y = x_5/x_1 = rse^a, \tag{3.153}$$

where $s = \sqrt{\nu'}$, $\sqrt{\rho} = e^a$, and $1/\bar{\mu} = r^2$. This system corresponds to horospherical coordinates on the unit hyperboloid. The separation equations are

$$\frac{d^2E_1}{da^2} + \frac{dE_1}{da} - (e^{-2a}B + A)E_1 = 0,$$
$$\frac{d^2E_3}{ds^2} + BE_3 = 0, \tag{3.154}$$

the equation for $E_2(r)$ being identical with (3.105). The basis operators $\hat{A}$, $\hat{B}$ are

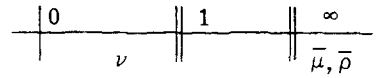
$$\hat{A} = \tfrac{1}{4} - D^2, \quad \hat{B} = (M_{01} - M_{12})^2. \tag{3.155}$$

System (ii) is related to (i) via the transformation

$$t \to it, \quad x \to ix, \quad y \to iy.$$

4. Coordinate systems on the cone $t^2 - x^2 - y^2 = 0$ obtainable from [221] and its degenerate forms

The method of obtaining such coordinates follows along the lines outlined previously. Consider the diagram



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \bar{\mu}\bar{\rho}, \quad 2\sigma x_1 x_2 = 0, \quad \sigma x_3^2 = \nu - 1,$$
$$2\sigma x_3 x_4 = -\nu, \quad \sigma x_5^2 = \nu. \tag{3.156}$$

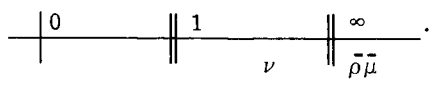
A suitable choice of three space coordinates is

$$(t + x)^2 = -x_3^2/x_1^2 = r^2(1 - \nu), \quad y^2 = x_5^2/x_1^2 = r^2\nu,$$
$$x^2 + y^2 - t^2 = 2x_2/x_1 = 0, \tag{3.157}$$

where $1/\bar{\mu}\bar{\rho} = r^2$. The separation equations are given by (3.51) with $P(\lambda) = (\lambda - 1)^2\lambda$ for the coordinate $\lambda = \nu$ and for the radial coordinate the equation is (3.105). The basis operators are

$$\hat{A} = \tfrac{1}{4} - D^2, \quad \hat{B} = 2M_{12}^2 + M_{12}M_{20} + M_{20}M_{12}. \tag{3.158}$$

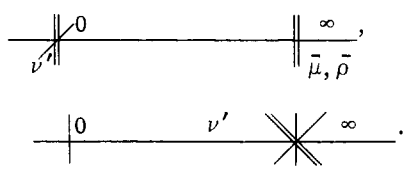
(b)



This system is related to (a) via the transformation

(a) → (b), $t \to ix$, $x \to it$, $y \to y$.

(c)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \bar{\mu}\bar{\rho}, \quad 2\sigma x_1 x_2 = 0, \quad \sigma x_3^2 = -1,$$
$$2\sigma x_3 x_4 = -\nu', \quad \sigma x_5^2 = \nu'. \tag{3.159}$$

A suitable choice of three space coordinates is

$$t - x = ix_3/x_1 = r, \quad t^2 - x^2 = -2x_3x_4/x_1^2 = r^2\nu',$$
$$y = x_5/x_1 = r(\nu')^{1/2}. \tag{3.160}$$

The separation equations have the form (3.51) with $P(\lambda) = \lambda^3$. The equation in the radial coordinate is (3.105), and the basis defining operators $\hat{A}$, $\hat{B}$ are

$$\hat{A} = \tfrac{1}{4} - D^2, \quad \hat{B} = (M_{01} - M_{12})^2. \tag{3.161}$$

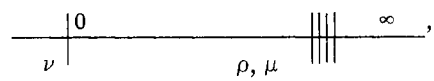
E. The configuration [41] and the degenerate form [(41)]

1. The configuration [41]

The inequivalent coordinate systems of type [41] are given in the following list. In each case the two free parameters e_1 and e_5 can be chosen to be ∞ and 0, respectively.

(a) [41]

(i)



(ii)

The corresponding pentaspherical coordinates are

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = -(\mu + \nu + \rho + 2),$$
$$\sigma(2x_1x_3 + x_2^2) = -(\nu\rho + \mu\nu + \mu\rho + \rho + \mu + \nu + 1), \tag{3.162}$$
$$2\sigma(x_1x_4 + x_2x_3) = \mu\nu\rho, \quad \sigma x_5^2 = -\mu\nu\rho.$$

A suitable choice of three space coordinates is

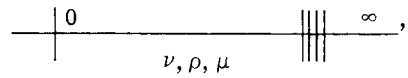
$$2(t + x) = 4x_3/x_1 = \mu\nu + \mu\rho + \nu\rho - \tfrac{1}{2}(\mu^2 + \nu^2 + \rho^2),$$
$$2(x - t) = 2(x_2/x_1 - 1) = \mu + \nu + \rho, \tag{3.163}$$
$$y^2 = -x_5^2/x_1^2 = -\mu\nu\rho.$$

Here the second equation has been subjected to a translation. This is merely a convenience. The separation equations for this coordinate system have the form (3.51) with $p(\lambda) = \lambda$. The solutions are expressible in terms of Bessel functions. The corresponding basis operators are

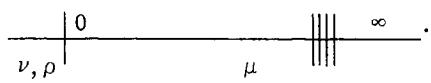
$$\hat{A} = (P_0 + P_1)M_{10} + M_{10}(P_1 + P_0) - 2P_2(M_{12} - M_{20})$$
$$- 2(M_{12} - M_{20})P_2 - (P_1 - P_0)^2, \tag{3.164}$$
$$\hat{B} = (M_{12} - M_{20})^2 - P_2(M_{12} + M_{20}) - (M_{12} + M_{20})P_2.$$

(b)

(i)



(ii)



This system is related to (a) via the transformation

(a) → (b), $t \to x$, $x \to t$, $y \to iy$.

2. The degenerate case with configuration [(41)]

The pentaspherical coordinates corresponding to such a system are obtained by making substitutions

$$e_5 = e_1 + \epsilon, \quad \lambda = e_1 + \epsilon + \epsilon^4\lambda', \tag{3.165}$$

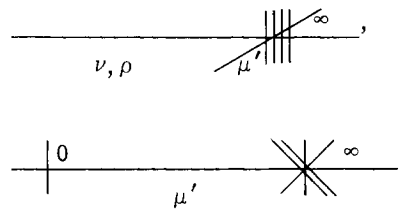
where $\lambda = \mu$, say. The resulting expression for the pentaspherical coordinates is

$$\sigma x_1^2 = -(\nu - e_1)(\rho - e_1), \quad 2\sigma x_1 x_2 = -(\nu - e_1) - (\rho - e_1),$$
$$\sigma(2x_1x_3 + x_2^2) = 1, \quad 2\sigma(x_1x_4 + x_2x_3) = -\mu'(\nu - e_1)(\rho - e_1),$$
$$\sigma x_5^2 = \mu'(\nu - e_1)(\rho - e_1). \tag{3.166}$$

If we further specialize to the case $e_1 = \infty$, these equations simplify to

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = \rho + \nu, \quad \sigma(2x_1x_3 + x_2^2) = \rho\nu,$$
$$2\sigma(x_1x_4 + x_2x_3) = -\mu', \quad \sigma x_5^2 = \mu'. \tag{3.167}$$

The one coordinate system of this type corresponds to the diagrams



A suitable choice of three space coordinates is

$$(y - t) = -4x_2/x_1 = 2(\rho + \nu), \quad (y + t) = 2ix_3/x_1,$$
$$x = ix_5/x_1 = k. \tag{3.168}$$

The separation equations have the form (3.51) for the variables ν, ρ with $p(\lambda) = \text{const}$ and (3.135b) in the variable k. The basis operators $\hat{A}$, $\hat{B}$ are

$$\hat{A} = M_{02}(P_0 + P_2) + (P_0 + P_2)M_{02} + (P_0 - P_2)^2,$$
$$\hat{B} = P_1^2. \tag{3.169}$$

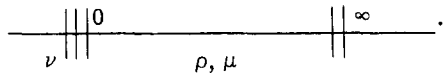
F. The configuration [32] and associated coordinate systems

1. The configuration [32]

As usual in the classification of inequivalent coordinate systems the two free parameters e_1 and e_4 can be

standardized to be 0 and ∞ (not necessarily respectively).

(a) [32]



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \mu\nu\rho, \quad 2\sigma x_1 x_2 = -(\mu\nu + \mu\rho + \nu\rho), \tag{3.170}$$
$$\sigma(2x_1x_3 + x_2^2) = -\mu - \nu - \rho, \quad \sigma x_4^2 = -1, \quad 2\sigma x_4 x_5 = \mu + \nu + \rho.$$

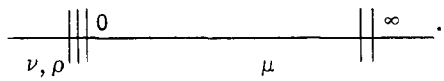
The suitable choice of three space coordinates is

$$(t-x)^2 = x_1^2/x_4^2 = -\mu\nu\rho, \quad 2y(x-t) = 2x_1x_2/x_4^2 = \mu\nu + \nu\rho + \mu\rho,$$
$$t^2 - x^2 - y^2 = -x_5/x_4 = \mu + \nu + \rho. \tag{3.171}$$

The separation equations are given by (3.51) with $p(\lambda) = \lambda^3$. The corresponding basis defining operators are

$$\hat{A} = M_{12}^2 - M_{01}^2 - M_{02}^2 - 2P_2(P_0 + P_1), \tag{3.172}$$
$$\hat{B} = M_{01}(M_{12} - M_{02}) + (M_{12} - M_{02})M_{01} - (P_0 + P_1)^2.$$

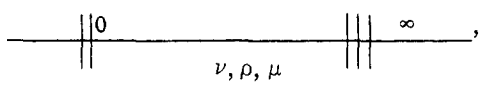
(b) [32]



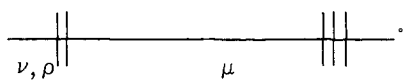
This system is related to (a) via the transformation (a) → (b), $t \to it$, $x \to ix$, $y \to iy$.

(c) [32]

(i)



(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = -(\mu + \nu + \rho),$$
$$\sigma(2x_1x_3 + x_2^2) = -\mu\nu - \mu\rho - \nu\rho, \tag{3.173}$$
$$\sigma x_4^2 = -\mu\nu\rho, \quad 2\sigma x_4 x_5 = \mu\nu + \mu\rho + \nu\rho.$$

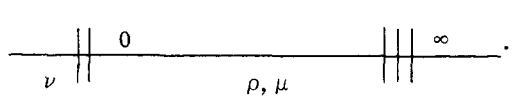
The suitable choice of three space coordinates is

$$(t-x)^2 = x_4^2/x_1^2 = \mu\nu\rho, \quad t^2 - x^2 = -2x_4x_5/x_1^2 = \mu\nu + \mu\rho + \nu\rho,$$
$$2y = 2x_2/x_1 = \mu + \nu + \rho \tag{3.174}$$

with the pentaspherical coordinates chosen as in (3.173). The separation equations are (3.41) with $f(\lambda) = \lambda^2$ and the basis operators are

$$\hat{A} = (P_0 + P_1)(M_{12} + M_{02}) + (M_{12} + M_{02})(P_0 + P_1) + 3(P_0 + P_1)^2 \tag{3.175}$$
$$\hat{B} = 2(P_1 + P_0)(M_{12} - M_{02}) + 2(M_{12} - M_{02})(P_0 + P_1) + M_{01}^2.$$

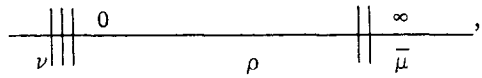
(d)



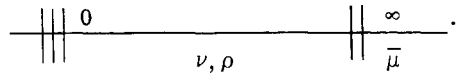
(c) → (d), $t \to ix$, $x \to it$, $y \to y$.

2. Coordinate systems of type [32] corresponding to a radial coordinate in three space

(a) (i)



(ii)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \nu\rho, \quad 2\sigma x_1 x_2 = -\nu - \rho,$$
$$\sigma(2x_1x_3 + x_2^2) = 1, \quad \sigma x_4^2 = \bar{\mu}, \quad 2\sigma x_4 x_5 = 2. \tag{3.176}$$

For (i) this corresponds to a choice of three space coordinates

$$(t-x)^2 = -x_1^2/x_4^2 = -\nu\rho r^2, \quad 2y(x-t) = 2x_1x_2/x_4^2 = (\nu + \rho)r^2,$$
$$x^2 + y^2 - t^2 = -x_5/x_4. \tag{3.177}$$

The separation equations are given by (3.51) with $p(\lambda) = \lambda^3$ for $\lambda = \nu, \rho$ and the equation in the variable r is (3.105). The basis operators $\hat{A}$, $\hat{B}$ are

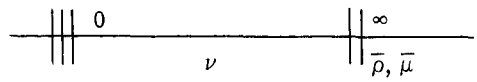
$$\hat{A} = \tfrac{1}{4} - D^2,$$
$$\hat{B} = M_{01}(M_{12} - M_{02}) + (M_{12} - M_{02})M_{01}. \tag{3.178}$$

The corresponding results for (ii) follow via the transformation $t \to it$, $x \to ix$, $y \to iy$.

3. Coordinates on the cone arising from the configuration [32]

There is one case to consider here.

(a)



The pentaspherical coordinates are given by

$$\sigma x_1^2 = \nu, \quad 2\sigma x_1 x_2 = -1, \quad \sigma(2x_1x_3 + x_2^2) = 0,$$
$$\sigma x_4^2 = \bar{\mu}\bar{\rho}, \quad 2\sigma x_4 x_5 = 0. \tag{3.179}$$

The associated choice of three space variables is

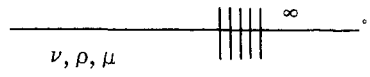
$$(t-x)^2 = x_1^2/x_4^2 = \nu r^2, \quad 2y(x-t) = 2x_1x_2/x_4^2 = -r^2,$$
$$t^2 - x^2 - y^2 = x_5/x_4 = 0. \tag{3.180}$$

The separation equations are (3.51) with $p(\lambda) = \lambda^3$ for $\lambda = \nu$ and (3.105) for the variable r. The basis operators are

$$\hat{A} = \tfrac{1}{4} - D^2,$$
$$\hat{B} = M_{01}(M_{12} - M_{02}) + (M_{12} - M_{02})M_{01}. \tag{3.181}$$

G. The configuration [5]

There is only one coordinate system for such a configuration and it has the diagram



The pentaspherical coordinates are given by

$$\sigma x_1^2 = 1, \quad 2\sigma x_1 x_2 = -(\mu + \nu + \rho), \tag{3.182}$$
$$\sigma(x_2^2 + 2x_1 x_3) = \mu\nu + \mu\rho + \nu\rho, \quad \sigma(2x_2 x_3 + 2x_1 x_4) = \mu\nu\rho.$$

This gives the three space variables

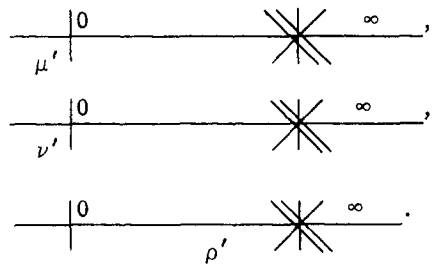
$$\begin{aligned} 2(t-x) &= -2x_2/x_1 = \mu + \nu + \rho, \\ 2(t+x) &= -2x_4/x_1 = -\tfrac{1}{2}\mu\nu\rho + \tfrac{1}{4}[\nu^2(\rho+\mu) + \rho^2(\mu+\nu)] \\ &\quad + \mu^2(\nu+\rho) - (\mu^3 + \nu^3 + \rho^3)], \\ 4y &= 4x_3/x_1 = \mu\nu + \mu\rho + \nu\rho - \tfrac{1}{2}(\mu^2 + \lambda^2 + \rho^2). \end{aligned} \tag{3.183}$$

The separation equations are (3.41) with $f(\lambda) = 1$. This gives the product of three solutions of Airy's equation. The resulting basis operators are

$$\begin{aligned} \hat{A} &= 8[2(P_0 - P_1)^2 + (P_0 + P_1)(M_{12} + M_{20}) \\ &\quad + (M_{12} + M_{20})(P_0 + P_1) - P_2 M_{10} - M_{10} P_2], \\ \hat{B} &= M_{01}(P_0 + P_1) + (P_0 + P_1)M_{01} + 4P_2(M_{12} - M_{20} - P_1 + P_0) \\ &\quad + 4(M_{12} - M_{20} - P_1 + P_0)P_2. \end{aligned} \tag{3.184}$$

H. Cartesian coordinates

The defining coordinates t, x, y can be incorporated into the scheme we have used here in the same way that Bôcher has done for the Laplace equation in three space. The diagrams for such a coordinate system are



The expressions for the pentaspherical coordinates are

$$\begin{aligned} &\sigma x_1^2 = -1, \quad 2\sigma x_1 x_2 = -\mu' - \nu' - \rho', \\ &\sigma x_3^2 = \rho', \quad \sigma x_4^2 = \nu', \quad \sigma x_5^2 = \mu', \end{aligned} \tag{3.185}$$

where the x_i are as in (1.26). The separation equations are obviously of the form

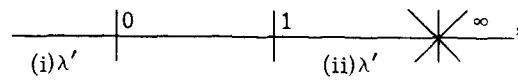
$$\frac{d^2 E_i}{d\lambda'^2} + K_i E_i = 0 \tag{3.186}$$

with basis defining operators any two of the operators P_i^2 $(i = 0, 1, 2)$.

IV. CONCLUDING REMARKS

In this paper we have made a detailed study of the orthogonal coordinate systems in three-dimensional Minkowski space for which the two-dimensional wave equation (*) admits an R-separable solution. The method for doing this is due to Bôcher and involves the use of pentaspherical coordinates. The direct relation between pentaspherical coordinates and the symmetry group of (*) was clearly demonstrated. The utility of the method over alternate ways of finding separable solutions of differential equations such as the classification of differential forms[11] is clear. Not only can the coordinates be found, but the separation equations and modulation factor can be determined from the key formulas in Sec. II.

As mentioned in the introductory comments of Sec. III, we have given a list of coordinate systems, some of which are equivalent under the action of the O(3, 2) symmetry group of (*) but not under the action of the $E(2, 1)$ subgroup. The coordinate systems corresponding to the configurations [21(11)], [(21)(11)], and [3(11)], in which the coordinate curve for a contracted variable λ' has a diagram



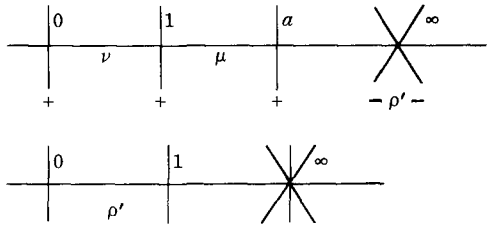
are equivalent under the O(3, 2) group action to one of the nine classes of coordinate systems which have a radial coordinate. This reflects the fact that the operator $\hat{B}$ as in (3.21) with $\lambda' = \rho'$ can always be chosen to be $-\frac{1}{4} + D^2$. Further, for the systems corresponding to the configuration [32], those which have $e_1 = \infty$, $e_4 = 0$ and $e_1 = 0$, $e_4 = \infty$ are equivalent under the action of O(3, 2) but not under the E(2, 1) subgroup. Similar comments apply to the systems with configuration [2(21)].

No attempt has been made to firmly establish that all inequivalent classes of orthogonal R separable solutions of (*) have been found. This topic will be the subject of subsequent work. Taking into account the equivalent systems as indicated in the preceding comments, we have presented 53 coordinate systems inequivalent under the O(3, 2) group action. In addition all the coordinate systems except those belonging to the configuration [11111] give separable solutions of the Helmholtz equation $\partial_{tt}\psi - \Delta_2\psi = K^2\psi$. There are 53 such systems. All the coordinate systems in Secs. 1–4, 6 of I are represented here. In particular the nine coordinate systems of the Euler–Poincaré–Darboux (EPD) equation. In subsequent articles it is our intention to look at the EPD equation in detail and to examine solutions of (*) which are R-separable but not orthogonal.

*Present address: Mathematics Department, University of Waikato, Hamilton, New Zealand.

[1]E. G. Kalnins and W. Miller Jr., J. Math. Phys. 16, xxxx (1975).
[2]P. Moon and D. E. Spencer, *Field Theory Handbook* (Springer-Verlag, Berlin, 1961).
[3]M. Bôcher, *Ueber die Reihenentwickelungen der Potentialtheorie* (Teubner, Leipzig, 1894).
[4]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part I* (McGraw-Hill, New York, 1953).
[5]J. L. Coolidge, *A Treatise on the Circle and the Sphere* (Chelsea, New York, 1971).
[6]T. J. I'A. Bromwich, *Quadratic Forms and Their Classification by Means of Invariant Factors* (Cambridge U. P., Cambridge, 1906).
[7]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[8]E. G. Kalnins, "On the separation of variables for the Laplace equation $\Delta\psi + K^2\psi = 0$ in two and three dimensional Minkowski space," SIAM J. Math. Anal. 6, 340 (1975).
[9]E. G. Kalnins and W. Miller Jr., J. Math. Phys. 15, 263 (1974).
[10]N. W. Macfadyen and P. Winternitz, J. Math. Phys. 12, 281 (1971).
[11]L. P. Eisenhart, Ann. Math. 35, 284 (1934).

†(b) [(11)111]



The pentaspherical coordinates are given as in (3.20). The three space coordinates are given by

$$t = ix_2/(x_5 + ix_1), \quad x = x_4/(x_5 + ix_1), \tag{P1}$$

$$y = x_3/(x_5 + ix_1).$$

The modulation factor is

$$\sqrt{2}\sigma^{1/4} s = (\sqrt{\nu\mu/a} + \sqrt{1-\rho'})^{1/2}.$$

The basis operators are

$$4\hat{A} = a(P_2 + K_2)^2 + (P_1 + K_1)^2, \tag{P2}$$

$$4\hat{B} = -1 - (P_0 - K_0)^2.$$

Lie theory and separation of variables. 10. Nonorthogonal R-separable solutions of the wave equation $\partial_{tt}\psi = \Delta_2\psi$

E. G. Kalnins

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, Québec, Canada

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 8 May 1975)

We classify and discuss the possible nonorthogonal coordinate systems which lead to R-separable solutions of the wave equation. Each system is associated with a pair of commuting operators in the symmetry algebra so (3,2) of this equation, one operator first order and the other first or second order. Several systems appear here for the first time.

INTRODUCTION

This paper is one of a series[1-9] investigating the relationship between the symmetry groups of the principal equations of mathematical physics and the coordinate systems for which the corresponding equations admit an R-separable solution. We recall that a solution $\psi(x_1,x_2,x_3)$ of an equation in three variables is R-separable if it can be written in the form

$$\psi(x_1,x_2,x_3)=\exp[Q(x_1,x_2,x_3)]A(x_1)B(x_2)C(x_3),$$

where e^Q contains no factors which are functions of one variable. The factor e^Q is called the modulation factor. The last two papers in this series[8,9] have dealt with a study of the wave equation in two space dimensions

$$\partial_{tt}\psi=\Delta_2\psi. \qquad (*)$$

In Paper 8 [8] of this series (hereafter referred to as I) we have given a detailed treatment of the symmetry group of (*) which is locally isomorphic to O(3,2). In that article are also discussed the principal equations contained in (*) when a generator of the Lie algebra is diagonalized. The resulting coordinate systems were called semisubgroup coordinate systems. In Paper 9 [9] (hereafter referred to as II) of this series, we complemented the contents of I with a detailed study of the orthogonal R-separable solutions of (*). This was achieved using pentaspherical space and families of confocal cyclides. The methods were principally those developed by Bôcher.[10] In this work we supplement the contents of I and II by looking for R-separable solutions of (*) which correspond to coordinates which are nonorthogonal.

If

$$ds^2=dt^2-dx^2-dy^2$$
$$=g^{ij}\,dx_i\,dx_j$$

is such that $g^{ij}\neq 0$ for at least one pair of indices $i\neq j$ and (*) admits an R-separable solution in the variables x_1,x_2,x_3, these coordinates constitute a nonorthogonal R-separable coordinate system. It is the purpose of this article to classify such coordinate systems. The contents of the paper are divided into three sections. In Sec. I we classify all coordinate systems in terms of their differential forms. This is done in detail by elementary and straightforward methods. The separation equations for each system we find are also given here. In Sec. II we give the coordinate systems in Minkowski space which correspond to the differential forms given in Sec. I. We also give the operators which specify the separation constants in each system. These are the operators associated with each system. Finally in Sec. III we look at the properties of coordinate systems which are specified by elements of an SL(2,R) subalgebra of the symmetry group of (*). This corresponds to the SL(2,R) algebra in Sec. 7 of Paper 8 of this series.

I. THE CLASSIFICATION OF SEPARABLE NONORTHOGONAL COORDINATE SYSTEMS

In this section we give a classification of the nonorthogonal coordinate systems for which (*) admits an R-separable solution. As opposed to the sophisticated methods used in II, we proceed in a straightforward manner here. These techniques have already been used previously.

We use the conditions of R-separability together with the requirement that the space be flat. The first requirement reduces to a number of special cases in which the metric g^{ij} has a prescribed form. For the space to be flat means that all the components of the Riemann curvature tensor are zero. The solution of these two constraints then gives us the list of possible nonorthogonal R-separable coordinate systems for the Laplace operator in a flat space. In each case we obtain a specific form for the metric tensor g^{ij}. Each of the nonorthogonal R-separable systems that we find corresponds to a prescribed coordinate system in Minkowski space with coordinates t,x,y. This reflects the fact that the only other candidate space satisfying the above conditions is Euclidean three-space, which does not admit nonorthogonal R-separable solutions of the Laplace operator.

A few words about our definition of R-separation are in order. More specifically we consider at first pure separation. A solution of (*) $\psi(x_1,x_2,x_3)$ in three new curvilinear variables $\mu,\nu,\rho\to x_1,x_2,x_3$ is said to be separable if $\psi=A(x_1)B(x_2)C(x_3)$ and each of the factor functions satisfies a second or first order ordinary differential equation. By a nonorthogonal coordinate system we shall mean a coordinate system for which at least one g^{ij} $(i\neq j)$ is nonzero. Here g^{ij} is the metric

tensor expressing the line element $ds^2 = g^{ij}dx_i dx_j$. For such a coordinate system the wave operator has the general form

$$\Delta = \partial_{tt} - \Delta_2 = \sum_{i \neq j} a_{ij}\partial_{ij} + \sum_i a_i \partial_i, \tag{1.1}$$

where $i,j = 1,2,3$ and at least one a_{ij} $(i \neq j)$ is nonzero. From this general form it follows that at least one of the separation equations must be of first order. The definition of separable coordinates for such a coordinate system that we adopt is that for at least one of the variables whose separation equation is first order the wave equation $\Delta\psi = 0$ can be rewritten as a function of the single variable on one side and a function of the remaining two variables on the other side so that one variable "separates." The equation in the remaining two variables separates in the same manner. (There are other more complicated ways for variables to separate which do not fall within this definition; see Sec. III. In this sense our results are not entirely complete.) In addition the coordinate functions

$$t = F(x_i), \quad x = G(x_i), \quad y = H(x_i) \quad (i = 1,2,3) \tag{1.2}$$

are real functions of the x_i only. For the case of R-separation the above definition carries over to the reduced wave equation, which results when the modulation factor e^Q is extracted. The function Q may, however, depend on the separation constants. For each coordinate system the two separation constants l_1 and l_2 are the eigenvalues of two operators $\mathcal{L}_1$ and $\mathcal{L}_2$ which are expressible as at most second order symmetric operators in the enveloping algebra of the O(3,2) symmetry group of $(*)$.

We now proceed to the solution of our problem and examine the conditions which will permit a separable solution of $(*)$. Recall that if we rewrite $(*)$ in terms of the variables x_i the equation assumes the form

$$\Delta\psi = \partial_{tt}\psi - \Delta_2\psi = a_{11}\partial_{11}\psi + a_{22}\partial_{22}\psi + a_{33}\partial_{33}\psi + a_{12}\partial_{12}\psi + a_{13}\partial_{13}\psi + a_{23}\partial_{23}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0. \tag{1.3}$$

Here Δ is the Laplacian corresponding to the contravariant metric tensor g^{ij} in the differential form:

$$ds^2 = g^{ij}dx_i dx_j. \tag{1.4}$$

The expression for Δ in terms of the metric tensor and variables x_i is

$$\Delta = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_i} \left(\sqrt{g}\ g_{ij} \frac{\partial}{\partial x_j} \right), \tag{1.5}$$

where $g = \det(g^{ij})$ and g_{ij} is the covariant metric tensor of our original contravariant tensor g^{ij}. (*Note*: In this article we prefer to write all our results in terms of the covariant variables x_1, x_2, x_3 as a matter of convenience.)

It is now the problem of separation of variables for the equation $\Delta\psi = 0$ that is our concern. From expression (1.3) we find four possibilities.

(1) *All the separation equations are first order*: From (1.3) and the fact that $a_{ij} = 2g_{ij}$ $(i \neq j)$ and $a_{ii} = g_{ii}$ we have $g_{11} = g_{22} = g_{33} = 0$. Equation (1.3) then assumes the form

$$a_{12}\partial_{12}\psi + a_{13}\partial_{13}\psi + a_{23}\partial_{23}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0. \tag{1.6}$$

If the separation equation for the variable x_3 is $\beta\, dC/dx_3 + \gamma C = 0$, then (1.6) reduces to the form

$$a_{12}\partial_{12}\phi + b_1\partial_1\phi + b_2\partial_2\phi + b_0\phi = 0,$$

where $\phi = A(x_1)B(x_2)$. The condition that this equation admit a separation is that either $a_{12} = 0$ or $b_1 = 0$, say. From the possible forms of the first order separation equations the condition $b_1 = 0$ requires $a_{13} = a_1 = 0$. In any case the covariant metric g_{ij} is singular and therefore inadmissable.

(2) *Exactly one separation equation is of second order*: If this equation is in the variable x_3, then $g_{11} = g_{22} = 0$. The resulting equation has the form

$$a_{33}\partial_{33}\psi + a_{12}\partial_{12}\psi + a_{13}\partial_{13}\psi + a_{23}\partial_{23}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0. \tag{1.7}$$

For a separable solution of (1.7) it is necessary that either $a_{13} = 0$ or $a_{23} = 0$. We cannot choose $a_{12} = 0$ as this would imply $g_{12} = 0$ and hence a singular metric tensor.

(3) *Two of the separation equations are second order*: If these equations are in the variables x_2 and x_3, then a necessary condition for the separation of (with $a_{11} = 0$ by hypothesis)

$$a_{22}\partial_{22}\psi + a_{33}\partial_{33}\psi + a_{12}\partial_{12}\psi + a_{13}\partial_{13}\psi + a_{23}\partial_{23}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0 \tag{1.8}$$

is that $a_{23} = 2g_{23} = 0$.

(4) *All the separation equations are second order*: This case is of no interest for this work as separation of variables now implies that $a_{ij} = 0$ for $i \neq j$. This is the case that has been treated in II and corresponds to orthogonal coordinates.

We now proceed to those cases of interest by taking special choices of the contravariant metric g^{ij}. We enumerate the possibilities.

(I) *R*-separable differential forms in which one nondiagonal element of the covariant metric tensor is nonzero

(A) Pure separation

The most general such form of the metric tensor is

$$g^{ij} = \begin{bmatrix} a & h & 0 \\ h & 0 & 0 \\ 0 & 0 & c \end{bmatrix}, \tag{1.9}$$

corresponding to the covariant metric tensor

$$g_{ij} = \begin{bmatrix} 0 & 1/h & 0 \\ 1/h & -a/h^2 & 0 \\ 0 & 0 & 1/c \end{bmatrix}. \tag{1.10}$$

The wave equation assumes the form

$$a_{22}\partial_{22}\psi + a_{33}\partial_{33}\psi + a_{12}\partial_{12}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0 \tag{1.11}$$

We consider first the x_1 dependence of the metric co-

efficients a, h, and c. In order that the x_1 dependence separate out in an equation of the form $T(x_1)\partial_1 A(x_1) = KA(x_1)$, the coefficients in (1.11) must satisfy the constraints

$$a_{22}=F(x_1)\hat{a}_{22},\quad a_{33}=F(x_1)\hat{a}_{33},\quad a_2=F(x_1)\hat{a}_2,$$
$$a_3=F(x_1)\hat{a}_3,\quad a_{12}=G(x_1)\hat{a}_{12},\quad a_1=G(x_1)\hat{a}_1, \tag{1.12}$$

where the functions $\hat{a}_{ij}$ and $\hat{a}_i$ depend upon x_2 and x_3 only. These conditions imply $h=[1/G(x_1)]\hat{h}(x_2,x_3)$. By suitable redefinition of x_1 we can take $G=1$. The remaining conditions imply $c=\hat{c}/F(x_1)$, $a=F(x_1)\hat{a}$. There are then two cases to consider. (i) $F(x_1)=$ constant. (ii) $F(x_1)$ not a constant. In the latter case the form of $F(x_1)$ can be found from the requirement $a_2=F(x_1)\hat{a}_2$. This means $F'(x_1)\propto F^2$. We can therefore take $F=1/x_1$ without loss of generality. The two cases to be considered are then specified by

(1) $h=h(x_2,x_3)$, $a=a(x_2,x_3)$, and $c=c(x_2,x_3)$,

(2) $h=h(x_2,x_3)$, $a=a(x_2,x_3)/x_1$, and $c=c(x_2,x_3)x_1$,

and will be considered separately.

(1) The equations which ensure that the space is flat are obtained by equating the nontrivially zero components of the Riemannian curvature tensor R_{ijkl} to zero. For the case (1) these equations are

$$2R_{1221}=a_{22}-\frac{h_3^2}{2c}-\frac{a_2h_2}{h}=0, \tag{1.13a}$$

$$2R_{1331}=a_{33}-\frac{h_3a_3}{h}-\frac{aa_2c_2}{2h^2}-\frac{a_3c_3}{2c}+\frac{ah_3^2}{2h^2}=0, \tag{1.13b}$$

$$2R_{2332}=c_{22}-\frac{h_2c_2}{h}-\frac{c_2^2}{2c}=0, \tag{1.13c}$$

$$2R_{3221}=-h_{32}+\frac{h_2h_3}{h}+\frac{h_3c_2}{2c}=0, \tag{1.13d}$$

$$2R_{3112}=a_{32}-\frac{a_2h_3}{h}-\frac{c_2a_3}{2c}=0, \tag{1.13e}$$

$$2R_{2331}=h_{33}-\frac{a_2c_2}{2h}-\frac{h_3c_3}{2c}-\frac{h_3^2}{2h}=0. \tag{1.13f}$$

For this case we consider two possibilities: $c_2\neq 0$, $c_2=0$. If $c_2\neq 0$, then equation (1.3) has the form

$$-\frac{a}{h^2}\partial_{22}\phi+\frac{1}{c}\partial_{33}\phi+\left(\frac{2l}{h}+a_2\right)\partial_2\phi+a_3\partial_3\phi-\frac{lc_2}{2hc}\phi=0, \tag{1.14}$$

where $\phi=B(x_2)C(x_3)$ and $\partial_1 A(x_1)=lA(x_1)$. Multiplying (1.14) by c, we obtain the separation condition

$$c_2/h=[f(x_2)+g(x_3)]r(x_3).$$

From (1.13c) we have

$$c_2/h=s(x_3)c^{1/2}$$

and

$$h=2f_2r/s^2.$$

Now $h\neq 0$ which implies $f\neq$ const. Accordingly we can define a new x_2 variable $x_2=f$ so that $h=h(x_3)$. From (1.13d) we then have $h_3=0$. Therefore, $h=1$ without loss of generality. The form of (1.14) now requires $a=a(x_2)$. Equation (1.13a) then implies $a=1$ or $a=0$. We also deduce that $c=t(x_2)u(x_3)$. By a suitable redefinition of x_3 we can take $c=c(x_2)$. From (1.13c) we then can take $c=x_2^2$. We finally obtain the two differential forms:

$$[1]\quad ds^2=2dx_1dx_2+x_2^2dx_3^2, \tag{1.15}$$

$$[2]\quad ds^2=dx_1^2+2dx_1dx_2+x_2^2dx_3^2. \tag{1.16}$$

If $c_2=0$, we can take $c=1$. From (1.14) we have the separation conditions $a/h=f(x_2)$, $h/c=r(x_2)s(x_3)$. From (1.13d) we have $h=t(x_2)u(x_3)$; hence $a=v(x_2)u(x_3)$. By redefinition of the variable x_2 we may take $h=u(x_3)$. From (1.13a) we then have

$$v_{22}=\alpha,\quad u_3^2=2\alpha u. \tag{1.17}$$

The general solutions of these equations are

$$v=\tfrac{1}{2}\alpha x_2^2+\beta x_2+\gamma,\quad u=(\sqrt{\alpha}/2x_3+\delta)^2, \tag{1.18}$$

where α,β,γ, and $\delta\in\mathbb{R}$, and $\alpha>0$. There are two classes of differential forms to consider:

(a) $\alpha=0$: We have the three possibilities

$$[3]\quad ds^2=2dx_1dx_2+dx_3^2, \tag{1.19}$$

$$[4]\quad ds^2=dx_1^2+2dx_1dx_2+dx_3^2, \tag{1.20}$$

$$[5]\quad ds^2=x_2dx_1^2+2dx_1dx_2+dx_3^2. \tag{1.21}$$

(b) $\alpha=1$: We have with suitable redefinitions

$$[6]\quad ds^2=x_3^2x_2^2dx_1^2+2x_3^2dx_1dx_2+dx_3^2, \tag{1.22}$$

$$[7]\quad ds^2=x_3^2(x_2^2-1)dx_1^2+2x_3^2dx_1dx_2+dx_3^2, \tag{1.23}$$

$$[8]\quad ds^2=x_3^2(x_2^2+1)dx_1^2+2x_3^2dx_1dx_2+dx_3^2. \tag{1.24}$$

This exhausts the list of separable differential forms in which the metric coefficients a, h, and c have no x_1 dependence.

(2) The equations requiring a flat space for the case of x_1 dependence have the form

$$x_1R_{1221}=\bar{R}_{1221}=0, \tag{1.25}$$
$$2x_1R_{1331}=2\bar{R}_{1331}+(\hat{a}_2\hat{c}+\hat{a}\hat{c}_2)/2h-\tfrac{1}{2}\hat{c}=0,$$
$$2R_{2332}/x_1=2\bar{R}_{2332}=0,\quad R_{3221}=\bar{R}_{3221}=0,$$
$$2x_1R_{3112}=2\bar{R}_{3112}+h_3/2\hat{c}=0,\quad R_{2331}=\bar{R}_{2331}=0.$$

Here the curvature tensor components $\bar{R}_{ijkl}$ are those in Eqs. (1.13) with $a\to\hat{a}$, $c\to\hat{c}$. Using arguments as for case (1), we find that the only forms of $\hat{a}$ and $\hat{c}$ compatible with the curvature equations are $\hat{a}=x_2$ and $\hat{c}=1$. This gives the separable differential form

$$[9]\quad ds^2=(x_2/x_1)dx_1^2+2dx_1dx_2+x_1dx_3^2. \tag{1.26}$$

(B) R-separation

As regards the possibility of an R-separable solution for coordinate systems of the type considered in this subsection, it can be shown that there are in fact no such systems. We do not reproduce the somewhat lengthy but straightforward calculations which lead to this negative result.

(II) R-separable differential forms in which two nondiagonal elements of the covariant metric tensor are nonzero and only one separation equation is second order

(A) Pure separation

The contravariant metric g^{ij} can be chosen as

$$g^{ij}=\begin{bmatrix}0 & h & 0\\ h & b^2 & bc\\ 0 & bc & c^2\end{bmatrix}. \tag{1.27}$$

The corresponding covariant metric tensor is

$$g_{ij}=\begin{bmatrix}0 & 1/h & -b/hc\\ 1/h & 0 & 0\\ -b/hc & 0 & 1/c^2\end{bmatrix}. \tag{1.28}$$

The wave equation assumes the form

$$a_{33}\partial_{33}\psi+a_{12}\partial_{12}\psi+a_{13}\partial_{13}\psi+a_1\partial_1\psi+a_2\partial_2\psi+a_3\partial_3\psi=0. \tag{1.29}$$

As before we consider first the x_1 dependence. The conditions for x_1 separation are

$$\begin{aligned}&a_{12}=F(x_1)\hat{a}_{12}, \quad a_{13}=F(x_1)\hat{a}_{13}, \quad a_1=F(x_1)\hat{a}_1,\\ &a_{33}=G(x_1)\hat{a}_{33}, \quad a_2=G(x_1)\hat{a}_2, \quad a_3=G(x_1)\hat{a}_3.\end{aligned} \tag{1.30}$$

These equations imply $h=\hat{h}(x_2,x_3)/F(x_1)$. By redefinition of x_1 we may as before take $F=1$. If G is not a constant, then the above conditions require $G'\propto G^2$, and we can take $G=1/x_1$. We again have two cases to consider:

$$\begin{aligned}&(1)\quad h=h(x_2,x_3), \quad b=b(x_2,x_3), \quad \text{and} \quad c=c(x_2,x_3),\\ &(2)\quad h=h(x_2,x_3), \quad b=x_1^{1/2}\hat{b}(x_2,x_3),\\ &\text{and} \quad c=x_1^{1/2}\hat{c}(x_2,x_3).\end{aligned} \tag{1.31}$$

(1) The curvature equations are

$$\begin{aligned}&R_{1221}=-h_3^2/4c^2=0, \quad R_{1331}\equiv 0,\\ &R_{2332}=b_3^2+bb_{33}+cc_{22}-b_{32}c-b_3c_2-c_3b_2-c_{32}b\\ &\quad+(h_2c/h)(b_3-c_2)+(c_3/c)(b_2c+bc_2-bb_3)\\ &\quad+(bh_3/h)(c_2-b_3)=0,\\ &R_{3221}=-\tfrac{1}{2}h_{32}+(h_3/4c)(bh_3/2h-c_2)=0,\\ &R_{3112}\equiv 0, \quad R_{2331}=\tfrac{1}{2}h_{33}-h_3c_3/2c^2=0.\end{aligned} \tag{1.32}$$

These equations immediately give $h_3=0$ and by redefinition of x_2 we can take $h=1$. Multiplying (1.29) by c^2 we have the further condition $bc=F(x_3)$. By redefinition of the variable x_3 we can take $bc=1$.

The separation conditions $a_{33}=u(x_2)v(x_3)$ and $a_{13}=u(x_2)v(x_3)$ imply that b^2 and c^2 may be taken in the form

$$b^2=F(x_3)/H(x_2), \quad c^2=H(x_2)/F(x_3). \tag{1.33}$$

With this choice the only nontrivial curvature equation is $R_{2332}=0$, and it has the form

$$2FF_{33}+F_3^2+2HH_{22}-H_2^2=0. \tag{1.34}$$

The separation equations for (1.34) are then

$$2FF_{33}+F_3^2=\alpha, \quad 2HH_{22}-H_2^2=-\alpha. \tag{1.35}$$

There are two cases to consider,

(a) $\alpha=0$: In this case equations (1.35) have the general solution.

$$H=(\beta x_2+\gamma)^2, \quad F=(\delta x_3+\epsilon)^{2/3}. \tag{1.36}$$

This gives four possibilities for the differential form according as the constants β,γ,δ, and ϵ are or are not zero:

$$[10]\quad ds^2=2dx_1dx_2+2dx_2dx_3+\omega dx_2^2+\frac{1}{\omega}dx_3^2, \tag{1.37}$$

$$[11]\quad ds^2=2dx_1dx_2+2dx_2dx_3+\omega x_3^{2/3}dx_2^2+\frac{dx_3^2}{\omega x_3^{2/3}}, \tag{1.38}$$

$$[12]\quad ds^2=2dx_1dx_2+2dx_2dx_3+\frac{\omega}{x_2^2}dx_2^2+\frac{x_2^2}{\omega}dx_3^2, \tag{1.39}$$

$$[13]\quad ds^2=2dx_1dx_2+2dx_2dx_3+\frac{\omega x_3^{2/3}}{x_2^2}dx_2^2+\frac{x_2^2}{\omega x_3^{2/3}}dx_3^2. \tag{1.40}$$

(b) $\alpha=1$: In this case we can integrate Eqs. (1.35) at once to get the relations

$$dx_3=F^{1/2}\,dF/(F+\beta)^{1/2}, \quad dx_2=dH/(1+\gamma H)^{1/2}. \tag{1.41}$$

Rather than integrate these equations further, we redefine the variables x_2 and x_3 by taking the new variables as H and F, respectively. We then distinguish four cases according as the constants β and γ are or are not zero. The resulting differential forms are

$$[14]\quad ds^2=2dx_1dx_2+2dx_2dx_3+\frac{x_3}{x_2}dx_2^2+\frac{x_2}{x_3}dx_3^2, \tag{1.42}$$

$$[15]\quad ds^2=2dx_1dx_2+\frac{2x_3^{1/2}}{(x_3+\beta)^{1/2}}dx_2dx_3+\frac{x_3}{x_2}dx_2^2+\frac{x_2}{(x_3+\beta)}dx_3^2, \tag{1.43}$$

$$[16]\quad ds^2=\frac{2}{(1+\gamma x_2)^{1/2}}dx_1dx_2+\frac{2}{(1+\gamma x_2)^{1/2}}dx_2dx_3+\frac{x_3}{x_2(1+\gamma x_2)}dx_2^2+\frac{x_2}{x_3}dx_3^2, \tag{1.44}$$

$$[17]\quad ds^2=\frac{2}{(1+\gamma x_2)^{1/2}}dx_1dx_2+\frac{2x_3^{1/2}dx_2dx_3}{[(1+\gamma x_2)(x_3+\beta)]^{1/2}}+\frac{x_3dx_2^2}{x_2(1+\gamma x_2)}+x_2\frac{dx_3^2}{(x_3+\beta)}. \tag{1.45}$$

(2) For the case of x_1 dependence the curvature equation $R_{1331}=0$ reduces to $c=0$, which is inadmissable. There are therefore no solutions of interest in this class.

(B) R-separation

If we assume that ψ in (*) has an R-separable solution of the form $\psi=e^R\phi$, then the equation satisfied by ϕ has the form

$$b_{33}\partial_{33}\phi + b_{12}\partial_{12}\phi + b_{13}\partial_{13}\phi + b_1\partial_1\phi + b_2\partial_2\phi + b_3\partial_3\phi + b_0\phi = 0, \tag{1.46}$$

where the b_{ij}, b_i are related to the a_{ij}, a_i in (1.6) by the equations

$$\begin{aligned} &b_{33} = a_{33}, \quad b_{12} = a_{12}, \quad b_{13} = a_{13}, \\ &b_1 = a_{12}R_2 + a_{13}R_3 + a_1, \quad b_2 = a_{12}R_1 + a_2, \\ &b_3 = a_{13}R_1 + 2a_{33}R_3 + a_3, \\ &b_0 = a_{12}(R_{12} + R_1R_2) + a_{13}(R_{13} + R_1R_3) + a_{33}(R_{33} + R_3^2) \\ &\quad + a_3R_3 + a_2R_2 + a_1R_1. \end{aligned} \tag{1.47}$$

As usual, we look at the possibilities for x_1 dependence. The conditions on the coefficients of (1.46) are

$$\begin{aligned} &b_{33} = G(x_1)\hat{b}_{33}, \quad b_2 = G(x_1)\hat{b}_2, \quad b_3 = G(x_1)\hat{b}_3, \\ &b_0 = G(x_1)\hat{b}_0, \quad b_{12} = F(x_1)\hat{b}_{12}, \quad b_{13} = F(x_1)\hat{b}_{13}, \\ &b_1 = F(x_1)\hat{b}_1. \end{aligned} \tag{1.48}$$

As in the case of pure separation, $F = 1$ by redefinition of the variable x_1 and, consequently, $h = h(x_2, x_3)$. The remaining conditions require, as in the case of pure separation, that $G = \text{const}$ or $G \propto x_1^{1/2}$. This latter case is inadmissable by the curvature conditions.

We may then take $h = 1$ and $c = c(x_2)$. The condition $b_2 = b_2(x_2, x_3)$ requires that R have the form $x_1u(x_2, x_3) + v(x_2, x_3)$. If the x_1 dependence in (1.46) is now extracted via the separation equation $dA(x_1)/dx_1 = lA(x_1)$, the resulting equation has the form

$$b_{33}\partial_{33}\phi + (lb_{12} + b_2)\partial_2\phi + (lb_{13} + b_3)\partial_3\phi + (b_1l + b_0)\phi = 0, \tag{1.49}$$

where $\phi = B(x_2)C(x_3)$. The separation condition $lb_{13} + b_3 = s(x_2, x_3)$ implies $u(x_2, x_3) = 0$. The further condition that $c^2(lb_{13} + b_3) = t(x_3)$ requires that $R_3 = lbc$ to within a sum of functions of single variables. The only nontrivial curvature equation is

$$R_{2332} = bb_{33} + b_3^2 + cc_{22} - b_{23}c - c_2b_3 = 0, \tag{1.50}$$

which has the solution $b_3 = c_2$ so that $b = c_2x_3 + g(x_2)$ and the modulation function R has the form

$$R = \tfrac{1}{2}lcc_2x_3^2 + lcgx_3. \tag{1.51}$$

Finally from the requirement $b_1l + b_0 = v(x_2) + w(x_3)$ we obtain the constraints

$$c^3c_{22} = \beta, \quad c^3g_2 = \gamma, \tag{1.52}$$

with $\beta, \gamma \in \mathbf{R}$. The general solution of the first equation is $c = (\delta x_2^2 + \epsilon)^{1/2}$. We now evaluate the possibilities depending on the values of the constants δ, ϵ:

(i) $\delta = 0$ and $\epsilon = 1$; then $g = \omega x_2$: The resulting metric is

$$[18]\quad ds^2 = 2dx_1\,dx_2 + 2\omega x_2\,dx_2\,dx_3 + dx_3^2 + \omega^2x_2^2\,dx_2^2, \tag{1.53}$$

and the modulation function is $R = \omega lx_2x_3$.

(ii) $\epsilon = 0$ and $\delta = 1$; then $g = \omega/x_2^2$ and the differential form is

$$[19]\quad ds^2 = 2dx_1\,dx_2 + 2(x_2x_3 + \omega/x_2)\,dx_2\,dx_3 + (x_3 + \omega/x_2^2)^2\,dx_2^2 + x_2^2\,dx_3^2. \tag{1.54}$$

The modulation function is then $R = \frac{1}{2}lx_2x_3^2 + \omega lx_3/x_2$.

(iii) $\delta = \epsilon = 1$; then $g = \omega$ and the differential form is

$$[20]\quad ds^2 = 2dx_1\,dx_2 + 2[x_2x_3 + \omega(1 + x_2^2)^{1/2}]\,dx_2\,dx_3 + [x_2x_3/(1 + x_2^2)^{1/2} + \omega]^2\,dx_2^2 + (1 + x_2^2)\,dx_3^2, \tag{1.55}$$

with the modulation function given by

$$R = \tfrac{1}{2}lx_2x_3^2 + l\omega x_3(1 + x_2^2)^{1/2}$$

(iv) $\delta = -\epsilon = 1$: In this case $g = \omega$ and the differential form is

$$[21]\quad ds^2 = 2dx_1\,dx_2 + 2[\nu x_2x_3 + \omega(|1 - x_2^2|)^{1/2}]\,dx_2\,dx_3 + [\nu x_3x_2/(|1 - x_2^2|)^{1/2} + \omega]^2\,dx_2^2 + |1 - x_2^2|\,dx_3^2, \tag{1.56}$$

where $\nu = \text{sgn}(-1 + x_2^2)$. The modulation function is $R = \frac{1}{2}\nu lx_2x_3^2 + l\omega x_3|1 - x_2^2|^{1/2}$. This completes our list of coordinate systems of this type.

(III) R-separable differential forms in which two nondiagonal elements of the covariant metric tensor are nonzero and two separation equations are of second order

Pure separation

The determination of the contravariant metric is rather involved. The wave equation for coordinate systems of this type will be taken as

$$a_{22}\partial_{22}\psi + a_{33}\partial_{33}\psi + a_{12}\partial_{12}\psi + a_{13}\partial_{13}\psi + a_1\partial_1\psi + a_2\partial_2\psi + a_3\partial_3\psi = 0. \tag{1.57}$$

The contravariant metric can then be taken to be

$$g^{ij} = \begin{bmatrix} a & f & abc/f \\ f & b^2 & bc \\ abc/f & bc & c^2 \end{bmatrix} \tag{1.58}$$

so that the components of the covariant metric tensor are

$$g_{ij} = \frac{1}{(f^2 - ab^2)} \begin{bmatrix} 0 & f & -bf/c \\ f & -a & 0 \\ -bf/c & 0 & f^2/c^2 \end{bmatrix}. \tag{1.59}$$

From the conditions for separation of the x_1 variable, which we do not repeat here [these are the analogs of Eqs. (1.30)], we find

$$a = G(x_1)\hat{a}, \quad b = \hat{b}/\sqrt{G(x_1)}, \quad c = \hat{c}/\sqrt{G(x_1)}, \tag{1.60}$$

where $G = 1$ or $1/x_1$. There are then two distinct cases to consider:

(1) $f = f(x_2, x_3)$, $a = a(x_2, x_3)$, $b = b(x_2, x_3)$, and $c = c(x_2, x_3)$.

(2) $f = f(x_2, x_3)$, $a = \hat{a}(x_2, x_3)/x_1$, $b = \hat{b}(x_2, x_3)x_1^{1/2}$, and $c = \hat{c}(x_2, x_3)x_1^{1/2}$.

(1) From the separation conditions $a_{12}/a_{22} = r(x_2)$ and $a_{13}/a_{33} = s(x_3)$ we have the relations $a = t(x_2)f$, $bc = u(x_3)f$, and $h = abc/f = t(x_2)u(x_3)f$. By suitable redefinition of the variables x_2 and x_3 these relations can be reduced to $a = f$, $bc = f$, $h = f$. [Note these results follow also for (2)

with a, b, and c replaced by $\hat{a}, \hat{b}$, and $\hat{c}$.] With $d=f^2/c^2$ the contravariant metric then assumes the relatively simple form

$$g^{ij}=\begin{bmatrix} f & f & f \\ f & d & f \\ f & f & f^2/d \end{bmatrix} \tag{1.61}$$

with corresponding covariant metric tensor

$$g_{ij}=\frac{1}{f(f-d)}\begin{bmatrix} 0 & f & -b \\ f & -f & 0 \\ -b & 0 & b \end{bmatrix}. \tag{1.62}$$

For this case the curvature equations are

$$2R_{1221}=f_{22}+[1/2(f-d)] \times[(2-d/f)f_2^2+f_2d_2-(d/f)f_3^2+(d/f)f_3d_3]=0, \tag{1.63a}$$

$$2R_{1331}=f_{33}+[1/2(f-d)] \times[(f/d)f_3d_3+(f^2/d^2)f_2d_2+(1-2f/d)f_2^2 +(2d-3f)f_3^2]=0, \tag{1.63b}$$

$$2R_{2332}=d_{33}-2f_{33}+(2f/d)f_{32}-(f^2/d^2)d_{22}+[1/(f-d)] \times[-(2f/d)f_2^2+(3f/d^2)(2d-f)f_2d_2+\tfrac{1}{2}(1+f/d)d_3^2 -4f_2d_3+2(2-d/f)f_2f_3+(d/f-2)f_3d_3-(f/d)d_2f_3] =0, \tag{1.63c}$$

$$2R_{3221}=f_{22}-f_{23}+[1/(f-d)][\tfrac{1}{2}f_2d_2+(d/2f-1)f_2^2 +(d/2f)f_3d_3+(2-d/f)f_2f_3-(f/2d)f_3d_2-\tfrac{1}{2}f_2d_3 +(b/2f)f_3^2]=0, \tag{1.63d}$$

$$2R_{3112}=f_{32}+[1/(f-d)] \times(f/2d)f_3d_2+(d/f-2)f_2f_3+\tfrac{1}{2}f_2d_3]=0, \tag{1.63e}$$

$$2R_{2331}=f_{33}-f_{32}+[1/(f-d)][(f/2d)f_3d_3+(d/2f-1)f_2^2 +(d/2f)f_3d_3+(2-d/f)f_2f_3-(f/2d)f_3d_2-\tfrac{1}{2}f_2d_3 +(d/2f)f_3^2]=0.$$

From these equations we deduce

$$2R_{1332}+2R_{3112}-2R_{1331}=[ff_2/d(f-d)][-(f/2d)d_2+f_2]=0. \tag{1.64}$$

There are then two possibilities: (i) $f_2=0$ or (ii) $d=r(x_3)f^2$. We consider each of these cases separately.

(i) From (1.63c) we have that $f_3d_2=0$ so that either $f_3=0$ or $d_2=0$. In the first case we can take $f=1$. Equation $R_{2332}=0$ requires

$$d_{33}+\frac{(1+d)}{2d(1-d)}d_3^2-\frac{d_{22}}{d^2}+\frac{(5d-3)}{2d(d-1)}\frac{d_2^2}{d^2}=0. \tag{1.65}$$

The separation condition $a_{33}/a_{22}=r(x_2)s(x_3)$ must also be satisfied. There are then three possibilities of this type.

a. $d=d(x_3)$: The variable x_3 can be redefined to be d via the relation

$$d_3=d^{1/2}-d^{1/2}. \tag{1.66}$$

The corresponding differential form is

$$[22]\quad ds^2=2dx_1\,dx_2+[2x_3^{1/2}/(1-x_3)]\,(dx_1\,dx_3+dx_2\,dx_3) +dx_1^2+x_3\,dx_2^2+dx_3^2/(1-x_3)^2. \tag{1.67}$$

b. $d=d(x_2)$: The variable x_2 can be redefined to be d via the relation

$$d_2=d^{3/2}(d-1). \tag{1.68}$$

$$[23]\quad ds^2=[2/(x_2-1)x_2^{3/2}]\,(dx_1\,dx_2+dx_2\,dx_3)+2dx_1\,dx_3 +dx_1^2+dx_2^2/(x_2-1)^2x_2^2+dx_3^2/x_2. \tag{1.69}$$

c. $d=\omega$, const: The differential form is

$$[24]\quad ds^2=2dx_1\,dx_2+2dx_1\,dx_3+2dx_2\,dx_3+dx_1^2 +\omega dx_2^2+(1/\omega)\,dx_3^2. \tag{1.70}$$

In addition we must consider the case when $f_3\neq 0$ and $d_2=0$. From (1.63a) this implies $f_3=d_3$ so that $f=d+\delta$ with $\delta\neq 0$. Integrating (1.63c) once, we get $d_3=\delta d^{-1/2}+d^{1/2}$. The variable x_3 can then be redefined to be d. The resulting differential form is

$$[25]\quad ds^2=2(x_3+\delta)\,dx_1\,dx_2+2x_3^{1/2}(dx_2\,dx_3+dx_1\,dx_3) +(x_3+\delta)\,dx_1^2+x_3\,dx_2^2+dx_3^2. \tag{1.71}$$

(ii) In this case the separation condition $a_{33}/a_{22}=u(x_2)v(x_3)$ ensures that d and f can each be expressed as products of functions in each of the variables x_2 and x_3. We may therefore take $f=h(x_2)r(x_3)$, $d=h^2(x_2)s(x_3)$. If r and s are both constants, then (1.63f) implies $h_2=0$. This case has already been found and corresponds to (1.69), (1.70). For nonconstant r and $s=\text{const}=1$, (1.63a) can be put into the form

$$\frac{h_2^2}{2h}-h_{22}=\frac{1}{2(r-h)}\left(-\frac{h_2^2}{h}r+2h_2^2-\frac{h^2r_3^2}{r^2}\right). \tag{1.72}$$

For the right-hand side of (1.72) to be a function of x_2 only, we require that $h=\exp(x_2)$ and $r=\exp(x_3)$. By choosing now variables h and r the differential form becomes

$$[26]\quad ds^2=2x_3\,dx_1\,dx_2+2x_2\,dx_1\,dx_3+2dx_2\,dx_3 +x_2x_3\,dx_1^2+dx_2^2+dx_3^2. \tag{1.73}$$

It is not hard to show that this is the only form of the functions r and s which are compatible with the curvature equations.

(2) For the case of explicit x_1 dependence it can be shown by straightforward but lengthy calculations that there are no differential forms of this type. Similar remarks apply to the case of R-separation.

This concludes our derivation of the differential forms.

II. EXPLICIT COORDINATES AND R-SEPARABLE SOLUTIONS

Here we present the list of coordinates corresponding to the differential forms given in Sec. I. We also present

with each coordinate system the separation equations and a representative solution where possible. We connect the listed coordinate systems with the symmetry group of (*) by giving the operators which specify the separation constants in terms of the generators of the symmetry algebra. We need only recall here the form of the generators in the coordinate representation. [For more information on the group structure associated with (*) we refer the reader to Paper 8 in this series.] The generators are

1. Translations: $P_0=\partial_t,\quad P_1=\partial_x,\quad P_2=\partial_y.$ (2.1)

2. Two-dimensional Lorentz subgroup SO(2,1):

$$M_{12}=x\partial_y-y\partial_x,\quad M_{01}=t\partial_x+x\partial_t,\quad M_{02}=t\partial_y+y\partial_t. \tag{2.2}$$

3. Dilatation: $D=-t\partial_t-x\partial_x-y\partial_y-\tfrac{1}{2}.$ (2.3)

4. Special conformal transformations:

$$\begin{aligned}K_0&=-t-(t^2+x^2+y^2)\partial_t-2tx\partial_x-2t_y\partial_y,\\ K_1&=x+(t^2+x^2-y^2)\partial_x+2xt\partial_t+2xy\partial_y,\\ K_2&=y+(t^2+y^2-x^2)\partial_y+2yt\partial_t+2yx\partial_x.\end{aligned} \tag{2.4}$$

In a number of cases we give simpler forms of the differential forms than given in Sec. I. This is achieved by making use of earlier results in this series of papers and is mentioned when it occurs. We now list the coordinate systems:

[1] $ds^2=2dx_1dx_2+x_2^2dx_3^2.$ (2.5)

The coordinates are given by

$$\begin{aligned}&t=x_1+\tfrac{1}{2}x_2x_3^2+x_2(x_3+1),\quad x=x_1+\tfrac{1}{2}x_2x_3^2+x_2x_3,\\ &y=x_2(x_3+1).\end{aligned} \tag{2.6}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad 2x_2^2\frac{dB}{dx_2}-x_2B=\frac{l_2}{l_1}B,\quad \frac{d^2C}{dx_3^2}+l_2C=0, \tag{2.7}$$

where $\psi=A(x_1)B(x_2)C(x_3)$ is a separable solution of (*). A typical solution is

$$\psi=\exp(l_1x_1)x_2^{1/2}\exp(-l_2/2l_1x_2)\begin{Bmatrix}\cos\sqrt{l_2}\,x_3\\ \sin\sqrt{l_2}\,x_3\end{Bmatrix}. \tag{2.8}$$

The operators L_1 and L_2 which specify this coordinate system are given in terms of the generators by

$$L_1=P_0+P_1,\quad L_2=(M_{12}-M_{02})^2. \tag{2.9}$$

[2] $ds^2=dx_1^2+2dx_1dx_2+x_2^2dx_3^2.$ (2.10)

The three space coordinates are given by

$$t=x_2\cosh x_3,\quad x=x_2\sinh x_3,\quad y=x_1+x_2. \tag{2.11}$$

The separation equations are

$$\begin{aligned}&\frac{dA}{dx_1}=l_1A,\quad \frac{d^2B}{dx_2^2}+\left(\frac{1}{x_2}-2l_1\right)\frac{dB}{dx_2}+\left(\frac{l_2}{x_2^2}-\frac{l_1}{x_2}\right)B=0,\\ &\frac{d^2C}{dx_3^2}=l_2C.\end{aligned} \tag{2.12}$$

A typical solution is

$$\psi=\exp(l_1x_1)\exp(l_2x_2)C_{\sqrt{l_2}}(il_1x_2)\exp(i\sqrt{l_2}x_3), \tag{2.13}$$

where $C_\nu(z)$ is a solution of Bessel's equation. The operators which specify the coordinate system are

$$L_1=P_2,\quad L_2=M_{01}^2. \tag{2.14}$$

[3] $ds^2=2dx_1dx_2+dx_3^2.$ (2.15)

The three space coordinates are

$$\sqrt{2}\,t=x_1+x_2,\quad \sqrt{2}\,x=x_1-x_2,\quad y=x_3. \tag{2.16}$$

The separation equations have the form

$$\frac{dA}{dx_1}=l_1A,\quad \frac{dB}{dx_2}=l_2B,\quad \frac{d^2C}{dx_3^2}=-2l_1l_2C. \tag{2.17}$$

A typical solution is

$$\psi=\exp(l_1x_1+l_2x_2)\begin{Bmatrix}\cos\sqrt{2l_1l_2}x_3\\ \sin\sqrt{2l_1l_2}x_3\end{Bmatrix}. \tag{2.18}$$

The operators which specify the coordinate system are

$$L_1=\sqrt{2}\,(P_0+P_1),\quad L_2=\sqrt{2}\,(P_0-P_1). \tag{2.19}$$

[4] $ds^2=dx_1^2+2dx_1dx_2+dx_3^2.$ (2.20)

The three space coordinates are

$$\begin{aligned}&\text{(i)}\ t=x_1,\quad x=x_1+x_2,\quad y=x_3,\\ &\text{(ii)}\ t=x_1+x_2,\quad x=x_2,\quad y=x_3.\end{aligned} \tag{2.21}$$

The separation equations have the form

$$\frac{dA}{dx_1}=l_1A,\quad \frac{d^2B}{dx_2^2}-2l_1\frac{dB}{dx_2}-l_2B=0,\quad \frac{d^2C}{dx_3^2}=\pm l_2C. \tag{2.22}$$

A typical solution is

$$\psi=\exp(l_1x_1)\exp(l_1x_2)\begin{Bmatrix}\cos\sqrt{l_2+l_1^2}x_2\\ \sin\sqrt{l_2+l_1^2}\,x_2\end{Bmatrix}\begin{Bmatrix}\cos\sqrt{\pm l_2}x_3\\ \sin\sqrt{\pm l_2}x_3\end{Bmatrix}. \tag{2.23}$$

The operators which specify the coordinate system are

$$\begin{aligned}&\text{(i)}\ L_1=P_0+P_1,\quad L_2=P_2^2,\\ &\text{(ii)}\ L_1=P_0,\quad L_2=P_2^2.\end{aligned} \tag{2.24}$$

[5] $ds^2=4\frac{x_2dx_1^2}{x_1^2}+4\frac{dx_1dx_2}{x_1}+dx_3^2.$ (2.25)

The three space coordinates are

$$t=x_2x_1-1/x_1,\quad x=x_2x_1+1/x_1,\quad y=x_3. \tag{2.26}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad x_2\frac{d^2B}{dx_2^2}+(1-2l_1)\frac{dB}{dx_2}-l_2B=0,\quad \frac{d^2C}{dx_3^2}=l_2C. \tag{2.27}$$

A typical solution is

$$\psi = x_1^{l_1} x_2^{l_1/2} C_{l_1}(2i\sqrt{l_2}x_2) \begin{cases} \cos\sqrt{-l_2}x_3 \\ \sin\sqrt{-l_2}x_3 \end{cases}. \tag{2.28}$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = M_{01}, \quad \mathcal{L}_2 = P_2^2. \tag{2.29}$$

[6] $$ds^2 = x_3^2 x_2^2 dx_1^2 + 2x_3^2 dx_1 dx_2 + dx_3^2. \tag{2.30}$$

The three space coordinates are

$$\begin{aligned} t &= x_3[\tfrac{1}{2}x_2(1 - Ex_1) + (E - 1/E)(1 - Ex_1)(1 - x_1x_2/2) \\ &\quad + 1/E\,(1 - x_1x_2/2)], \\ x &= x_3[1 - 2(1 - Ex_1)(1 - x_1x_2/2)], \\ y &= x_3[\tfrac{1}{2}x_2(1 - Ex_1) + (E + 1/E)(1 - Ex_1)(1 - x_1x_2/2) \\ &\quad - (1/E)(1 - x_1x_2/2)], \\ &E \in \mathbb{R}. \end{aligned} \tag{2.31}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad x_2^2 \frac{d^2B}{dx_2^2} + 2(x_2 - l_1)\frac{dB}{dx_2} - l_2 B = 0, \tag{2.32}$$

$$x_2^2 \frac{d^2C}{dx_3^2} + 2x_3 \frac{dC}{dx_3} - l_2 C = 0.$$

A typical solution is

$$\psi = \exp[l_1(x_1 - 1/x_2)]\, C_{j+1/2}(il_1/x_2) x_2^{-1/2} \begin{cases} x_3^j \\ x_3^{-j-1} \end{cases}, \tag{2.33}$$

where $l_2 = j(j+1)$. The operators which specify the coordinate system are

$$\mathcal{L}_1 = (E^2 + 1)M_{12} + (E^2 - 1)M_{01} - 2EM_{02}, \quad \mathcal{L}_2 = -\tfrac{1}{4} + D^2. \tag{2.34}$$

[7] $$ds^2 = x_3^2(x_2^2 - 1)dx_1^2 + 2x_3^2 dx_1 dx_2 + dx_3^2. \tag{2.35}$$

There are two alternative parametrizations in three space which correspond to the above differential form. They are

(i) $$\begin{aligned} t &= x_3[(e^{x_1} + E) + \tfrac{1}{2}e^{-x_1}(x_2 + 1)(e^{x_1} + E)^2 - 2e^{-x_1}(x_2 + 1)] \\ x &= x_3[1 - 2e^{-x_1}(x_2 + 1)(E + e^{x_1})], \\ y &= x_3[(e^{x_1} + E) + \tfrac{1}{2}e^{-x_1}(x_2 + 1)(e^{x_1} + E)^2 + 2e^{-x_1}(x_2 + 1)], \end{aligned} \tag{2.36}$$

where $E \in \mathbb{R}$

(ii) $$\begin{aligned} t &= x_3\{(4/\alpha)(E - \coth\tfrac{1}{2}x_1)[1 + (E\sinh\tfrac{1}{2}x_1 - \cosh\tfrac{1}{2}x_1) \\ &\quad \times(\cosh\tfrac{1}{2}x_1 - x_2\sinh\tfrac{1}{2}x_1)] \\ &\quad + (\alpha/4)\sinh\tfrac{1}{2}x_1(\cosh\tfrac{1}{2}x_1 - x_2\sinh\tfrac{1}{2}x_1)\}, \\ x &= x_3[1 - 2(E\sinh\tfrac{1}{2}x_1 - \cosh\tfrac{1}{2}x_1)(\cosh\tfrac{1}{2}x_1 - x_2\sinh\tfrac{1}{2}x_1)], \\ y &= x_3[(4/\alpha)(E - \coth\tfrac{1}{2}x_1)[1 + (E\sinh\tfrac{1}{2}x_1 - \cosh\tfrac{1}{2}x_1) \\ &\quad \times(\cosh\tfrac{1}{2}x_1 - x_2\sinh\tfrac{1}{2}x_1)] \\ &\quad - (\alpha/4)\sinh\tfrac{1}{2}x_1(\cosh\tfrac{1}{2}x_1 - x_2\sinh\tfrac{1}{2}x_1)], \end{aligned} \tag{2.37}$$

where α, $E \in \mathbb{R}$ and $\alpha \neq 0$.

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad (1 - x_2^2)\frac{d^2B}{dx_2^2} + (2l_1 - x_2)\frac{dB}{dx_2} + l_2 B = 0, \tag{2.38}$$

$$x_3^2 \frac{d^2C}{dx_3^2} + 2x_3 \frac{dC}{dx_3} - l_2 C = 0.$$

A typical solution is

$$\psi = e^{l_1 x_1} \left(\frac{x_2 - 1}{x_2 + 1}\right)^{l_1/2} \begin{cases} P_j^{l_1}(x_2) \\ Q_j^{l_1}(x_2) \end{cases} \begin{cases} x_3^j \\ x_3^{-j-1} \end{cases}, \tag{2.39}$$

where $l_2 = j(j+1)$. The functions $P_\nu^\mu(z)$ and $Q_\nu^\mu(z)$ are Legendre functions of the first and second kind respectively. The operators which specify the coordinate system are

$$\begin{aligned} &\text{(i)} \quad \mathcal{L}_1 = 2M_{02} - E(M_{12} + M_{01}), \quad \mathcal{L}_2 = -\tfrac{1}{4} + D^2, \\ &\text{(ii)} \quad \mathcal{L}_1 = (4/\alpha)(E^2 - 1)(M_{01} + M_{12}) + (\alpha/4)(M_{12} - M_{01}) \\ &\qquad\qquad - 2EM_{02}, \quad \mathcal{L}_2 = -\tfrac{1}{4} + D^2. \end{aligned} \tag{2.40}$$

[8] $$ds^2 = x_3^2(x_2^2 + 1)dx_1^2 + 2x_3^2 dx_1 dx_2 + dx_3^2. \tag{2.41}$$

The three space coordinates are given by

$$\begin{aligned} t &= x_3[-(4/\alpha)(E + \tan\tfrac{1}{2}x_1) \\ &\quad \times[1 + (\sin\tfrac{1}{2}x_1 + E\cos\tfrac{1}{2}x_1)(\sin\tfrac{1}{2}x_1 + x_2\cos\tfrac{1}{2}x_1)] \\ &\quad - (\alpha/4)\cos\tfrac{1}{2}x_1(\sin\tfrac{1}{2}x_1 + x_2\cos\tfrac{1}{2}x_1)], \\ x &= x_3[1 - 2(\sin\tfrac{1}{2}x_1 + x_2\cos\tfrac{1}{2}x_1)(\sin\tfrac{1}{2}x_1 + E\cos\tfrac{1}{2}x_1)], \\ y &= x_3[-(4/\alpha)(E + \tan\tfrac{1}{2}x_1)[1 + (\sin\tfrac{1}{2}x_1 + E\cos\tfrac{1}{2}x_1) \\ &\quad \times(\sin\tfrac{1}{2}x_1 + x_2\cos\tfrac{1}{2}x_1)] \\ &\quad + (\alpha/4)\cos\tfrac{1}{2}x_1(\sin\tfrac{1}{2}x_1 + x_2\cos\tfrac{1}{2}x_1)], \end{aligned} \tag{2.42}$$

where α, $E \in \mathbb{R}$ and $\alpha > 0$.

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad (1 + x_2^2)\frac{d^2B}{dx_2^2} + 2(x_2 - l_1)\frac{dB}{dx_2} - l_2 B = 0, \tag{2.43}$$

$$x_3^2 \frac{d^2C}{dx_3^2} + 2x_3 \frac{dC}{dx_3} - l_2 C = 0.$$

A typical solution is

$$\psi = \exp(l_1 x_1)\left(\frac{x_2 - 1}{x_2 + 1}\right)^{il_1/2} \begin{cases} P_j^{il_1}(ix_2) \\ Q_j^{il_1}(ix_2) \end{cases} \begin{cases} x_3^j \\ x_3^{-j-1} \end{cases}, \tag{2.44}$$

where as usual $l_2 = j(j+1)$. The operators which specify the coordinate system are

$$\mathcal{L}_1 = (4/\alpha)(1 + E^2)(M_{01} + M_{12}) + (\alpha/4)(M_{12} - M_{01}) + 2EM_{02}, \quad \mathcal{L}_2 = -\tfrac{1}{4} + D^2. \tag{2.45}$$

[9] $$ds^2 = (x_2/x_1)dx_1^2 + 2dx_1 dx_2 + x_1 dx_3^2. \tag{2.46}$$

The three space coordinates are given by

$$t+x=2x_2\sqrt{x_1}-\tfrac{1}{2}x_3^2\sqrt{x_1}\,,\quad t-x=-2\sqrt{x_1},\quad y=x_3\sqrt{x_1}\,. \tag{2.47}$$

The separation equations are

$$x_1\frac{dA}{dx_1}=l_1A,\quad x_2\frac{d^2B}{dx_2^2}+(2l_1-\tfrac{1}{2})\frac{dB}{dx_2}+l_2B=0,\quad \frac{d^2C}{dx_3^2}=l_2C. \tag{2.48}$$

A typical solution is

$$\psi=(x_1/x_2)^{l_1}x_2^{3/4}C_{3-4l_1}(2\sqrt{l_2x_2})\begin{cases}\cos\sqrt{l_2}\,x_3\\ \sin\sqrt{l_2}\,x_3\end{cases}. \tag{2.49}$$

The operators which specify this coordinate system are

$$L_1=-\tfrac{1}{2}D-\tfrac{1}{4},\quad L_2=\tfrac{1}{4}(M_{12}-M_{02})^2. \tag{2.50}$$

[10] $$ds^2=2dx_1\,dx_2+2dx_2\,dx_3+\omega dx_2^2+(1/\omega)\,dx_3^2. \tag{2.51}$$

The three space coordinates are

$$\begin{aligned}&t+x=2x_1+2(1-E/\sqrt{\omega})x_3+(\omega-E^2)x_2,\\ &t-x=-x_2,\quad y=x_3/\sqrt{\omega}+Ex_2.\end{aligned} \tag{2.52}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad \frac{dB}{dx_2}=l_2B,\quad \omega^2\frac{d^2C}{dx_3^2}-2\omega^2l_1\frac{dC}{dx_3}+2l_1l_2C=0. \tag{2.53}$$

A typical solution is

$$\psi=\exp(l_1x_1)\exp(l_2x_2)\exp(l_1x_3)\begin{cases}\cos[l_1(2l_2-\omega^2l_1)]^{1/2}x_3\\ \sin[l_1(2l_2-\omega^2l_1)]^{1/2}x_3\end{cases}. \tag{2.54}$$

The operators which specify this coordinate system are

$$L_1=P_0+P_1,\quad L_2=\tfrac{1}{2}(\omega-E^2)(P_0+P_1)+\tfrac{1}{2}(P_1-P_0)+EP_2. \tag{2.55}$$

[11] $$ds^2=2dx_1\,dx_2+2dx_2\,dx_3+\omega x_3^{2/3}dx_2^2+dx_3^2/\omega x_3^{2/3}. \tag{2.56}$$

The three space coordinates are

$$\begin{aligned}&t+x=2x_1+2x_3+(\omega x_2-3E/2\sqrt{\omega})x_3^{2/3}-\omega^{-3/2}(\tfrac{1}{3}\omega x_2^{2/3}-E)^3,\\ &t-x=-x_2,\quad y=(3/2\sqrt{\omega})x_3^{2/3}-\tfrac{1}{6}\omega^{3/2}x_2^2+Ex_2.\end{aligned} \tag{2.57}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad \frac{dB}{dx_2}=l_2B,\quad x_3\frac{d^2C}{dx_3^2}+(\tfrac{1}{3}-2l_1x_3)\frac{dC}{dx_3}-(\tfrac{1}{3}l_1+l_2/\omega\,x_3^{1/3})C=0. \tag{2.58}$$

The operators which specify this coordinate system are

$$\begin{aligned}&L_1=P_0+P_1,\\ &L_2=\tfrac{1}{3}\omega^{3/2}(M_{02}-M_{12})+\tfrac{1}{2}(P_1-P_0)\\ &\quad-\tfrac{1}{2}E^2(P_0+P_1)+EP_2.\end{aligned} \tag{2.59}$$

[12] $$ds^2=2dx_1\,dx_2+2dx_2\,dx_3+(\omega/x_2^2)\,dx_2^2+(x_2^2/\omega)\,dx_3^2. \tag{2.60}$$

The three space coordinates are given by

$$\begin{aligned}&t+x=2x_1+2x_3-x_2x_3^2/\omega-(2E/\sqrt{\omega})x_2x_3-E^2x_2-\omega/x_2,\\ &t-x=-x_2,\quad y=x_2x_3/\sqrt{\omega}+Ex_2.\end{aligned} \tag{2.61}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad 2x_2^2\frac{dB}{dx_2}+(l_2+x_2)B=0,\quad \frac{d^2C}{dx_3^2}-2l_1\frac{dC}{dx_3}-\frac{l_1l_2}{\omega}C=0. \tag{2.62}$$

A typical solution is

$$\psi=\exp(l_1x_1)x_2^{-1/2}\exp(l_2/2x_2)\exp(l_1x_3)\begin{cases}\cosh\sqrt{l_1^2+l_1l_2/\omega}\,x_3\\ \sinh\sqrt{l_1^2+l_1l_2/\omega}\,x_3.\end{cases} \tag{2.63}$$

The operators which specify this coordinate system are

$$L_1=P_0+P_1,\quad L_2=2\omega(P_0+P_1)+K_0+K_1. \tag{2.64}$$

[13] $$ds^2=2dx_1\,dx_2+2dx_2\,dx_3+(\omega x_3^{2/3}/x_2^2)\,dx_2^2+(x_2^2/\omega x_3^{2/3})\,dx_3^2. \tag{2.65}$$

The three space coordinates are given by

$$\begin{aligned}&t+x=2x_1+2x_3-(9/4\omega)x_3^{4/3}x_2+[\omega/2x_2-(3E/\sqrt{\omega})x_2]x_3^{2/3}\\ &\quad-E^2x_2+E\omega^{3/2}/3x_2-\omega^3/108x_2^3,\\ &t-x=-x_2,\quad y=(3/2\sqrt{\omega})x_2x_3^{2/3}+\omega^{3/2}/6x_2+Ex_2.\end{aligned} \tag{2.66}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad 2x_2^2\frac{dB}{dx_2}+\left(\frac{l_2}{l_1}+x_2\right)B=0,\quad x_3\frac{d^2C}{dx_3^2}+(\tfrac{1}{3}-2l_1x_3)\frac{dC}{dx_3}-\left(\frac{l_1}{3}+\frac{l_2}{\omega}x_3^{1/3}\right)C=0. \tag{2.67}$$

The operators which specify this coordinate system are

$$L_1=P_0+P_1,\quad L_2=K_0+K_1+\tfrac{2}{3}E\omega^{3/2}(P_0+P_1)-\tfrac{1}{3}\omega^{3/2}P_2, \tag{2.68}$$

[14] $$ds^2=2dx_1\,dx_2+2dx_2\,dx_3+(x_3/x_2)dx_2^2+(x_2/x_3)\,dx_3^2. \tag{2.69}$$

The three space coordinates are given by

$$t+x=2x_1-E^2x_2-4E\sqrt{x_2x_3},\quad t-x=-x_2,\quad y=2\sqrt{x_2x_3}+Ex_2. \tag{2.70}$$

The separation equations are

$$\frac{dA}{dx_1}=l_1A,\quad 2x_2\frac{dB}{dx_2}=l_2B,\quad x_3\frac{d^2C}{dx_3^2}+(\tfrac{1}{2}-2l_1x_3)\frac{dC}{dx_3}-l_1l_2C=0. \tag{2.71}$$

A typical solution is

$$\psi = \exp[l_1(x_1+x_3)]x_2^{l_2/2} D_{-1/2-il_2}[\pm(1+i)\sqrt{2l_1x_3}], \quad (2.72)$$

where $D_\nu(z)$ is a parabolic cylinder function. The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1, \quad \mathcal{L}_2 = E(M_{12} - M_{02}) - M_{01} - D - \tfrac{1}{2}. \quad (2.73)$$

$$[15]\quad ds^2 = 2dx_1\,dx_2 + [2x_3^{1/2}/(x_3+\beta)^{1/2}]\,dx_2\,dx_3 + (x_3/x_2)\,dx_2^2 + [x_2/(x_3+\beta)]\,dx_3^2. \quad (2.74)$$

The three space coordinates are

$$t+x = 2x_1 - 2x_3 + \sqrt{x_3(x_3+\beta)} - 2\beta\ln(\sqrt{x_3+\beta}+\sqrt{x_3}) - \beta\ln x_2 - E^2x_2 - 4Ex_2^{1/2}(x_3+\beta)^{1/2}, \quad (2.75)$$

$$t-x = -x_2, \quad y = 2\sqrt{x_2(x_3+\beta)} + Ex_2.$$

The separation equations are

$$\frac{dA}{dx_1} = l_1E_1, \quad x_2\frac{dB}{dx_2} = l_2B,$$

$$(x_3+\beta)\frac{d^2C}{dx_3^2} + [\tfrac{1}{2} - 2l_1\sqrt{x_3(x_3+\beta)}]\frac{dC}{dx_3} + \left\{2l_1l_2 + \tfrac{1}{2}l_1\left[1 - \left(\frac{x_3+\beta}{x_3}\right)^{1/2}\right]\right\}C = 0. \quad (2.76)$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1,$$
$$\mathcal{L}_2 = -\tfrac{1}{2}[\beta(P_0+P_1) - D - M_{01} - \tfrac{1}{2} + E(M_{12} - M_{02})]. \quad (2.77)$$

$$[16]\quad ds^2 = \frac{2}{(1+\gamma x_2)^{1/2}}\,dx_1\,dx_2 + \frac{2}{(1+\gamma x_2)^{1/2}}\,dx_2\,dx_3 + \frac{x_3}{x_2(1+\gamma x_2)}\,dx_2^2 + \frac{x_2}{x_3}\,dx_3^2. \quad (2.78)$$

The three space coordinates are given by

$$t+x = 2x_1 - 2(1+\gamma x_2)^{1/2}(E^2/\gamma + x_3) - 4E\sqrt{x_2x_3} + 2x_3,$$
$$t-x = -(2/\gamma)(1+\gamma x_2)^{1/2},$$
$$y = 2\sqrt{x_2x_3} + (2E/\gamma)(1+\gamma x_2)^{1/2}.$$

The separation equations are

$$\frac{dA}{dx_1} = l_1A, \quad 2l_1x_2(1+\gamma x_2)^{1/2}\frac{dB}{dx_2} + [\tfrac{1}{2}(1+\gamma x_2)^{1/2} - l_2]B = 0,$$
$$x_3\frac{d^2C}{dx_3^2} + (\tfrac{1}{2} - 2l_1x_3)\frac{dC}{dx_3} + (l_2 - l_1)C = 0. \quad (2.80)$$

The operators which specify this coordinate system are

$$\mathcal{L}_1 = P_0 + P_1, \quad \mathcal{L}_2 = (\gamma/4)(K_0+K_1) + (1/\gamma)[(1-E^2)P_0 - (1+E^2)P_1 - EP_2]. \quad (2.81)$$

$$[17]\quad ds^2 = \frac{2}{(1+\gamma x_2)^{1/2}}\,dx_1\,dx_2 + \frac{2x_3^{1/2}}{\sqrt{(1+\gamma x_2)(x_3+\beta)}}\,dx_2\,dx_3 + \frac{x_3}{x_2(1+\gamma x_2)}\,dx_2^2 + \frac{x_2}{(x_3+\beta)}\,dx_3^2. \quad (2.82)$$

The three space coordinates are given by

$$t+x = 2x_1 - (2/\gamma)(1+\gamma x_2)^{1/2}(x_3+E^2) - 4E\sqrt{x_2(x_3+\beta)} - \beta\ln[(\sqrt{1+\gamma x_2}-1)/(\sqrt{1+\gamma x_2}+1)] + \sqrt{x_3(x_3+\beta)} - 2\beta\ln(\sqrt{x_3+\beta}+\sqrt{x_3})$$
$$t-x = -(2/\gamma)(1+\gamma x_2)^{1/2}, \quad y = 2\sqrt{x_2(x_3+\beta)} + (2E/\gamma)\sqrt{1+\gamma x_2}. \quad (2.83)$$

The separation equations are

$$\frac{dA}{dx_1} = l_1A, \quad 2l_1x_2\sqrt{1+\gamma x_2}\,\frac{dB}{dx_2} + (\tfrac{1}{2}\sqrt{1+\gamma x_2} - l_2)B = 0, \quad (2.84)$$

$$(x_3+\beta)\frac{d^2C}{dx_3^2} + [\tfrac{1}{2} - 2l_1\sqrt{x_3(x_3+\beta)}]\frac{dC}{dx_3} + \left[l_2 - \frac{1}{2}\left(\frac{x_3+\beta}{x_3}\right)^{1/2}\right]C = 0.$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1,$$
$$\mathcal{L}_2 = (\gamma/4)(K_0+K_1) + (1/\gamma)(P_0-P_1) - (2E/\gamma)P_2 + (E^2/\gamma - \beta)(P_0+P_1). \quad (2.85)$$

$$[18]\quad ds^2 = 2dx_1\,dx_2 + 2\omega x_2\,dx_2\,dx_3 + dx_3^2 + \omega^2x_2^2\,dx_2^2 \quad (2.86)$$

with modulation function $R = \omega l_1x_2x_3$.

The three space variables are given by

$$t+x = 2x_1, \quad t-x = -x_2, \quad y = x_3 + \tfrac{1}{2}\omega x_2^2. \quad (2.87)$$

The separation equations are

$$\frac{dA}{dx_1} = l_1A, \quad 2\frac{dB}{dx_2} - (l_1^2\omega^2x_2^2 + l_2)B = 0,$$
$$\frac{d^2C}{dx_3^2} + (2l_1^2\omega x_3 + l_1l_2)C = 0. \quad (2.88)$$

A typical solution is then

$$\psi = \exp(l_1\omega x_2x_3 + l_1x_1 + \tfrac{1}{6}l_1\omega^2x_2^3 + \tfrac{1}{2}l_2x_2)\begin{Bmatrix}\mathrm{Ai}(z)\\ \mathrm{Bi}(z)\end{Bmatrix} \quad (2.89)$$

with $z = (2l_1^2\omega)^{1/3}x_3 + l_2(2l_1^2\omega)^{-1/3}$. The functions $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are Airy functions.

The operators which define the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1, \quad \mathcal{L}_2 = P_1 - P_0 + 2\omega(M_{12} - M_{02}). \quad (2.90)$$

$$[19]\quad ds^2 = 2dx_1\,dx_2 + 2(x_2x_3 + \omega/x_2)\,dx_2\,dx_3 + (x_3 + \omega/x_2^2)^2\,dx_2^2 + x_2^2\,dx_3^2. \quad (2.91)$$

with modulation factor $R = +\tfrac{1}{2}l_1x_2x_3^2 + \omega l_1x_3/x_2$. The three space coordinates are

$$t+x = 2x_1 - E^2x_2 - 2Ex_2x_3 + 2\omega E/x_2,$$
$$t-x = -x_2, \quad y = x_2x_3 + Ex_2 - \omega/x_2. \quad (2.92)$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad 2x_2^2 \frac{dB}{dx_2} + \left(x_2 - \frac{\omega l_1}{x_2^2} - l_2\right) B = 0,$$
$$\frac{d^2C}{dx_3^2} + (-4\omega l_1^2 l_2^2 x_3 - l_2) C = 0. \tag{2.93}$$

A typical solution is

$$\psi = x_2^{-1/2} \exp\left(\tfrac{1}{2} l_1 x_2 x_3^2 + \frac{\omega l_1 x_3}{x_2} + l_1 x_1 - \frac{l_2}{2 l_1 x_2} - \frac{\omega^2 l_1}{6 x_2^3}\right) \times \begin{Bmatrix} \mathrm{Ai}(z) \\ \mathrm{Bi}(z) \end{Bmatrix}, \tag{2.94}$$

where $z = (4l_1^2\omega)^{1/3} x_3 + l_2 (4 l_1^2 \omega)^{-1/3}$.

The operators which define the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1, \quad \mathcal{L}_2 = K_1 + K_0 + 4\omega P_2. \tag{2.95}$$

[20] $$ds^2 = 2dx_1 dx_2 + 2x_2 x_3 dx_2 dx_3 + [x_2^2 x_3^3/(1 + x_2^2)] dx_2^2 + (1 + x_2^2) dx_3^2 \tag{2.96}$$

with modulation function $R = \frac{1}{2} l_1 x_2 x_3^2$.

The three space variables are

$$t + x = 2x_1 - E^2 x_2 - 2E x_3 (1 + x_2^2)^{1/2},$$
$$t - x = -x_2, \quad y = x_3 (1 + x_2^2)^{1/2} + E x_2. \tag{2.97}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad 2l_1 (1 + x_2^2) \frac{dB}{dx_2} + (l_1 x_2 - l_2) B = 0,$$
$$\frac{d^2C}{dx_3^2} + (l_1^2 x_3^2 + l_2) C = 0. \tag{2.98}$$

A typical solution is then

$$\psi = (1 + x_2^2)^{-1/4} \exp[\tfrac{1}{2} l_1 x_2 x_3^2 + l_1 x_1 + (l_2/2l_1) \tan^{-1} x_2] \times D_{-(1+il_2/l_1)/2}[\pm (1 + i) x_3 \sqrt{l_1}], \tag{2.99}$$

where $D_\nu(z)$ is a parabolic cylinder function. The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1,$$
$$\mathcal{L}_2 = -E^2 (P_0 + P_1) + 2EP_2 + P_1 - P_0 + K_1 + K_0. \tag{2.100}$$

[21] $$ds^2 = 2dx_1 dx_2 + 2\epsilon x_2 x_3 dx_2 dx_3 + (x_2^2 x_3^2/|1 - x_2^2|) dx_2^2 + |1 - x_2^2| dx_3^2, \tag{2.101}$$

where $\epsilon = \mathrm{sgn}(x_2^2 - 1)$ and the modulation function is $R = \frac{1}{2} \epsilon l_1 x_2 x_3^2$.

The three space coordinates are

$$t + x = 2x_1 - E^2 x_2 - 2E x_3 |1 - x_2^2|^{1/2},$$
$$t - x = -x_2, \quad y = x_3 |1 - x_2^2|^{1/2} + E x_2. \tag{2.102}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad 2\epsilon l_1 (x_2^2 - 1) \frac{dB}{dx_2} - (l_1 x_2 + l_2) B = 0,$$
$$\frac{d^2C}{dx_3^2} + (-l_1^2 x_3^2 + l_2) C = 0. \tag{2.103}$$

A typical solution is

$$\psi = \exp(\tfrac{1}{2} \epsilon l_1 x_2 x_2^2 + l_1 x_1)(x_2 - 1)^{\epsilon(l_2 + l_1)/4l_1} \times (x_2 + 1)^{\epsilon(l_2 - l_1)/4l_1} D_{-(l_2 + l_1)/2l_1}(\pm \sqrt{2l_1} x_3). \tag{2.104}$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_0 + P_1,$$
$$\mathcal{L}_2 = -E^2 (P_0 + P_1) - 2EP_2 + P_0 - P_1 + K_0 + K_1. \tag{2.105}$$

[22] $$ds^2 = 2dx_1 dx_2 + [2\sqrt{x_3}/(1 - x_3)] (dx_1 dx_3 + dx_2 dx_3) + dx_1^2 + x_3 dx_2^2 + dx_3^2/(1 - x_3)^2. \tag{2.106}$$

The three space coordinates are given by

$$t = 2\sqrt{1 - x_3} \sinh \tfrac{1}{2} x_2, \quad x = 2\sqrt{1 - x_3} \cosh \tfrac{1}{2} x_2,$$
$$y = x_1 + x_2 + 2\sqrt{x_3} + \ln[(\sqrt{x_3} - 1)/(\sqrt{x_3} + 1)]. \tag{2.107}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad \frac{d^2B}{dx_2^2} - 2l_1 \frac{dB}{dx_2} - l_2 B = 0,$$
$$(x_3 - 1) \frac{d^2C}{dx_3^2} + (x_3 - 1 + 2l_1 \sqrt{x_3}) \frac{dC}{dx_3} + \left(\frac{l_1}{2x_3} + l_2 - \tfrac{1}{2} l_1\right) C = 0. \tag{2.108}$$

The operators which specify this coordinate system are

$$\mathcal{L}_1 = P_2, \quad \mathcal{L}_2 = -\tfrac{3}{4} M_{01}^2 - M_{01} P_2 + P_2^2. \tag{2.109}$$

[23] $$ds^2 = \frac{2}{(x_2 - 1) x_2^{3/2}} (dx_1 dx_2 + dx_2 dx_3) + 2dx_1 dx_3 + dx_1^2 + \frac{dx_2^2}{(x_2 - 1)^2 x_2^2} + \frac{dx_3^2}{x_2}. \tag{2.110}$$

The three space coordinates are given by

$$t = 2\sqrt{(1/x_2) - 1} \cosh \tfrac{1}{2} x_3, \quad x = 2\sqrt{(1/x_2) - 1} \sinh \tfrac{1}{2} x_3,$$
$$y = x_1 + x_3 + 2x_2^{-1/2} + \ln[(\sqrt{x_2} + 1)/(\sqrt{x_2} - 1)]. \tag{2.111}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A,$$
$$x_2^2 (x_2 - 1)^2 \frac{d^2B}{dx_2^2} + \left(3x_2^2 - 2x_2^3 - x_2 + \frac{2l_1}{x_2}\right) \frac{dB}{dx_2} + \left(\frac{l_1}{2} + l_2 - \frac{l_1}{2x_2}\right) B = 0,$$
$$\frac{d^2C}{dx_3^2} - 2l_1 \frac{dC}{dx_3} - l_2 C = 0. \tag{2.112}$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = P_2, \quad \mathcal{L}_2 = -\tfrac{3}{4} M_{01}^2 - M_{01} P_2 + P_2^2. \tag{2.113}$$

[24] $$ds^2 = 2dx_1 dx_2 + 2dx_1 dx_3 + 2dx_2 dx_3 + dx_1^2 + \omega dx_2^2 + (1/\omega) dx_3^2. \tag{2.114}$$

The three space coordinates are given by

$$t = \sqrt{(\omega - 1)/\omega}\, x_3, \quad x = \sqrt{\omega - 1}\, x_2, \quad y = x_1 + x_2 + x_3. \tag{2.115}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad \frac{d^2B}{dx_2^2} - 2l_1 \frac{dB}{dx_2} - l_2 B = 0,$$
$$\frac{d^2C}{dx_3^2} - 2l_1 \frac{dC}{dx_3} - \frac{l_2}{\omega} C = 0. \tag{2.116}$$

A typical solution is

$$\psi = \exp[l_1(x_1+x_2+x_3)] \begin{Bmatrix} \cos\sqrt{l_1^2+l_2}\,x_2 \\ \sin\sqrt{l_1^2+l_2}\,x_2 \end{Bmatrix} \begin{Bmatrix} \cos\sqrt{l_1^2+l_2/\omega}\,x_3 \\ \sin\sqrt{l_1^2+l_2/\omega}\,x_3 \end{Bmatrix}. \tag{2.117}$$

The operators which specify this coordinate system are

$$\mathcal{L}_1 = P_2, \quad \mathcal{L}_2 = (\omega-1)(P_0^2+P_1^2) - (\omega+1)P_2^2. \tag{2.118}$$

[25] $$ds^2 = 2(x_3+\delta)\,dx_1\,dx_2 + 2x_3^{1/2}(dx_2\,dx_3 + dx_1\,dx_3) + (x_3+\delta)\,dx_1^2 + x_3\,dx_2^2 + dx_3^2/(x_3+\delta)^2. \tag{2.119}$$

The three space coordinates are given by

$$t = 2\sqrt{\delta(x_3+\delta)}\,\sinh[(x_1+x_2+2\sqrt{x_3}-2\sqrt{\delta}\ \tan^{-1}\sqrt{x_3/\delta})2\sqrt{\delta}\,],$$
$$x = 2\sqrt{\delta(x_3+\delta)}\,\cosh[(x_1+x_2+2\sqrt{x_3}-2\sqrt{\delta}\ \tan^{-1}\sqrt{x_3/\delta})2\sqrt{\delta}\,],$$
$$y = \sqrt{\delta}\,x_2. \tag{2.120}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad \frac{d^2B}{dx_2^2} - 2l_1\frac{dB}{dx_2} + l_2 B = 0,$$
$$(x_3+\delta)\frac{d^2C}{dx_3^2} + (1-2l_1)\frac{dC}{dx_3} + \left[l_2 - \frac{l_1}{2}\left(1+\frac{\delta}{x_3}\right)\right] C = 0. \tag{2.121}$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = M_{01}, \quad \mathcal{L}_2 = \delta P_2^2 - M_{01}^2. \tag{2.122}$$

[26] $$ds^2 = 2x_3\,dx_1\,dx_2 + 2x_2\,dx_1\,dx_3 + 2dx_2\,dx_3 + x_2x_3\,dx_1^2 + dx_2^2 + dx_3^2. \tag{2.123}$$

The three space coordinates are given by

$$t = x_2x_3\exp(x_1/2) + \exp(-x_1/2),$$
$$x = x_2x_3\exp(x_1/2) - \exp(-x_1/2), \quad y = x_2 + x_3. \tag{2.124}$$

The separation equations are

$$\frac{dA}{dx_1} = l_1 A, \quad x_2\frac{d^2B}{dx_2^2} + (1-2l_1)\frac{dB}{dx_2} + l_2 B = 0,$$
$$x_3\frac{d^2C}{dx_3^2} + (1-2l_1)\frac{dC}{dx_3} + l_2 C = 0. \tag{2.125}$$

A typical solution is

$$\psi = \exp(l_1x_1)\,(x_2x_3)^{l_1} C_{2l_1}(2\sqrt{l_2x_2})C_{2l_1}(2\sqrt{l_2x_3}). \tag{2.126}$$

The operators which specify the coordinate system are

$$\mathcal{L}_1 = \tfrac{1}{2}M_{01}, \quad 2\mathcal{L}_2 = P_0M_{02} + M_{02}P_0 + P_1M_{12} + M_{12}P_1. \tag{2.127}$$

III. OTHER TYPES OF SEPARATION

In this section we examine the coordinate systems associated with the diagonalization of the operator $L = \frac{1}{2}M_{12} - \frac{1}{4}(P_0 - K_0)$. The algebra of (*) when L has been diagonalized is SL(2,R) with basis

$$A = \tfrac{1}{2}M_{12} + \tfrac{1}{4}(P_0 - K_0), \quad B = \tfrac{1}{4}M_{01} + \tfrac{1}{4}(P_2 - K_2),$$
$$C = -\tfrac{1}{2}M_{02} + \tfrac{1}{4}(P_1 - K_1) \tag{3.1}$$

and commutation relations

$$[A,B] = C, \quad [C,A] = B, \quad [C,B] = A. \tag{3.2}$$

The coordinate systems associated with the diagonalization of L and an additional operator from the above SL(2,R) algebra are the semisubgroup coordinates of type 7 of Paper 8 of this series. In this section we give the three subgroup coordinates discussed in Paper 8 and leave open the question of whether there are any more. This will be the subject of subsequent study. The three coordinate systems we present are different from those presented in the earlier two sections in that they do not enable a separation of variables to occur explicitly in the equation. This becomes clear for the individual coordinate systems.

For the choice of variables

$$t = \frac{\sin\frac{1}{2}(\beta-\rho)}{\cos\sigma - \cos\frac{1}{2}(\beta-\rho)}, \quad x = \frac{\sin\sigma\cos\frac{1}{2}(\beta+\rho)}{\cos\sigma - \cos\frac{1}{2}(\beta-\rho)},$$
$$y = \frac{\sin\sigma\sin\frac{1}{2}(\beta+\rho)}{\cos\sigma - \cos\frac{1}{2}(\beta-\rho)}, \tag{3.3}$$

and $\psi = [\cos\sigma - \cos\frac{1}{2}(\beta-\rho)]^{1/2}\exp(i\chi\beta)\,\Theta(\sigma,\rho)$, we have $L\psi = i\chi\beta\psi$, where the function $\Theta(\sigma,\rho)$ satisfies the equation

$$(A^2 - B^2 - C^2)\Theta = (L^2 + \tfrac{1}{4})\Theta = (\tfrac{1}{4} - \chi^2)\Theta. \tag{3.4}$$

The diagonalization of A is easily performed in this coordinate system as $A = \partial_\rho$ when acting on the function Θ, and so for $\Theta(\sigma,\rho) = \Phi(\sigma)\exp(i\tau\rho)$ the corresponding solutions of (*) have the form

$$\psi_{\chi\tau}(\sigma,\beta,\rho) = [\cos\sigma - \cos\tfrac{1}{2}(\beta-\rho)]^{1/2} \times \exp(i\chi\beta)\exp(i\tau\rho)P^{\tau+\chi}_{\chi-\tau-1/2}(\cos\sigma). \tag{3.5}$$

In particular we note that the SL(2,R) generators acting on the functions Θ have the form

$$C + iB = \exp(i\rho)(-\partial_z - i\coth z\,\partial_\rho + \chi/\sinh z + \tfrac{1}{4}\tanh\tfrac{1}{2}z),$$
$$C - iB = \exp(-i\rho)(-\partial_z + i\coth z\,\partial_\rho - \chi/\sinh z + \tfrac{1}{4}\tanh\tfrac{1}{2}z),$$
$$A = \partial_\rho, \tag{3.6}$$

where $\sin\sigma = \tanh\frac{1}{2}z$. The pure derivative parts of these operators are the same as the corresponding operators that would be obtained on the two dimensional hyperboloid parametrized by $t = (\cosh z, \sinh z, \cos\rho, \sinh z\sin\rho)$. This suggests the procedure necessary for the remaining two subgroup coordinate systems which diagonalize C and $A - C$. The appropriate change of variables is given

by a knowledge of the subgroup type coordinates on the hyperboloid.[11] After extraction of the appropriate modulation function, the separation of variables is achieved. The results are:

1. *The diagonalization of* C: The appropriate change of variables is $\cosh z = \cosh a \cosh b$, $\tan\rho = \tanh a \sinh b$, and the R-separation modulation function is

$$f = (\cosh a \cosh b + 1)^{1/4} \exp[i\chi \tan^{-1}(\sinh a \coth b)]. \tag{3.7}$$

The generators acting on the functions Φ, where $\Theta = f\Phi$ have the form

$$A = \sinh b\partial_a - \tanh a \cosh b\partial_b - i\chi \cosh b/\cosh a,$$
$$B = -\cosh b\partial_a + \tanh a \sinh b\partial_b + i\chi \sinh b/\cosh a, \tag{3.8}$$

Then for $\Phi = \exp(i\tau b)H(a)$ the function H satisfies

$$\left(\frac{\partial^2}{\partial a^2} + \tanh a \frac{\partial}{\partial a} + \frac{\chi\tau}{\cosh^2 a} \sinh a + \chi^2 - \tfrac{1}{4}\right)H(a) = 0 \tag{3.9}$$

with solutions

$$H(a) = P^{-1/2+\chi}_{i\sqrt{\chi\tau}/2,\sqrt{\chi\tau}/2}(\cosh a),\quad Q^{-1/2+\chi}_{i\sqrt{\chi\tau}/2,\sqrt{\chi\tau}/2}(\cosh a),$$

where $P^{\mu}_{\nu}(z)$ and $Q^{\mu}_{\nu}(z)$ are the generalized Legendre functions.[11-13]

2. *The diagonalization of* $A - C$: The appropriate change of variables is

$$\cosh z = \cosh a + \tfrac{1}{2}r^2 e^{-a},\quad \tan\rho = re^{-a}/(\sinh a + \tfrac{1}{2}r^2 e^{-a}),$$

and the modulation function is

$$f = [(\cosh a + \tfrac{1}{2}r^2 e^{-a} - 1)/(\cosh a + \tfrac{1}{2}r^2 e^{-a} + 1)]^{i\chi/2}$$
$$\exp\{-\tfrac{1}{2}\tan^{-1}[r/(e^a + 1)]\}. \tag{3.10}$$

The generators acting on the functions $\Phi = f\psi$ have the form

$$B = \partial_a + r\partial_r,\quad A - C = \partial_r, \tag{3.11}$$
$$A + C = 2r\partial_a + (r^2 - e^{2a})\partial_r + \tfrac{1}{4}(2e^a - 1).$$

Then for $\Phi = \exp(i\tau r)H(a)$ the function H satisfies

$$[\partial^2/\partial a^2 - \partial/\partial a - \tau^2 e^{2a} - \tfrac{1}{4}i\tau(2e^a - 1) - \chi^2 + \tfrac{1}{4}]H(a) = 0, \tag{3.12}$$

which has solutions

$$H(a) = M_{i/4,\pm 2(\chi^2 - i\tau)^{1/2}}(2\tau e^a),$$

where $M_{\mu,\nu}(z)$ is a solution of Kummers differential equation.[14] We see that each of the subgroup types has an R-separable solution and does not fit into the scheme of Sec. I. We do not yet know if there are any more systems of these types.

The principal contribution of this article is to provide examples of R-separable solutions, which to our knowledge have not previously been exhibited. A unified group theoretical approach must be able to account for the explicit solutions and coordinate systems produced here.

[1] W. Miller Jr., SIAM J. Math. Anal. 5, 626 (1974).
[2] W. Miller Jr., SIAM J. Math. Anal. 5, 822 (1974).
[3] E.G. Kalnins and W. Miller Jr., J. Math. Phys. 15, 1025 (1974).
[4] E.G. Kalnins and W. Miller Jr., J. Math. Phys. 15, 1263 (1974).
[5] E.G. Kalnins and W. Miller Jr., J. Math. Phys. 15, 1728 (1974).
[6] C.P. Boyer, E.G. Kalnins and W. Miller Jr., J. Math. Phys. 16, 499 (1975).
[7] C.P. Boyer, E.G. Kalnins, and W. Miller Jr., J. Math. Phys. 16, 512 (1975).
[8] E.G. Kalnins and W. Miller Jr., J. Math. Phys. 17, xxx (1976).
[9] E.G. Kalnins and W. Miller Jr., J. Math. Phys. 17, xxx (1976).
[10] M. Bôcher, *Ueber die Reihenentwickelungen der Potentialtheorie* (Teubner, Leipzig, 1894).
[11] E.G. Kalnins, Ph.D. thesis (University of Western Ontario, 1972).
[12] Ya.I. Azimov, Sov. J. Nucl. Phys. 4, 469 (1967).
[13] E.G. Kalnins, J. Math. Phys. 14, 654 (1973).
[14] A. Bucholz, *The Confluent Hypergeometric Function* (Springer-Verlag, Berlin, 1969).

Lie theory and separation of variables. 11. The EPD equation

E. G. Kalnins

Centre de Recherches Mathématiques, Université de Montréal, Montréal 101, P.Q., Canada

Willard Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 2 May 1975)

We show that the Euler–Poisson–Darboux equation $\{\partial_{tt}-\partial_{rr}-[(2m+1)/r]\partial_r\}\Theta=0$ separates in exactly nine coordinate systems corresponding to nine orbits of symmetric second-order operators in the enveloping algebra of $SL(2,R)$, the symmetry group of this equation. We employ techniques developed in earlier papers from this series and use the representation theory of $SL(2,R)$ to derive special function identities relating the separated solutions. We also show that the complex EPD equation separates in exactly five coordinate systems corresponding to five orbits of symmetric second-order operators in the enveloping algebra of $SL(2,\mathbb{C})$.

INTRODUCTION

This paper is one of a series concerning the relationships between the symmetry group of a linear second order partial differential equation and the coordinate systems in which variables separate for that equation. The previous three papers[1] were devoted to separation of variables for the wave equation $(\partial_{tt}-\Delta_2)\psi(x)=0$. If we pass to polar coordinates,

$$x_1=r\cos\varphi,\quad x_2=r\sin\varphi,$$

and consider solutions of the form $\psi(x)=\exp(im\varphi)\Phi(t,r)$, the wave equation transforms to the Euler–Poisson–Darboux (EPD) equation

$$[\partial_{tt}-\partial_{rr}-(1/r)\,\partial_r+m^2/r^2]\,\Phi=0. \tag{0.1}$$

Many authors write $\Phi(t,r)=r^m\Theta(t,r)$ and take the EPD equation in the form

$$(\partial_{tt}-\partial_{rr}-[(2m+1)/r]\,\partial_r)\,\Theta=0, \tag{0.2}$$

but for our purposes (0.1) is more convenient. This equation also arises from the wave equations $(\partial_{tt}-\Delta_n)\,\psi(x)=0$, $n>2$, if one looks for spherically symmetric solutions. For $n=2$, m is usually taken to be an integer while, for $n>2$, m may be half-integral. In this paper we will treat these cases simultaneously by allowing m to be a nonnegative real number.

It follows from the results of Refs. 1 that (0.1) can be solved by separation of variables in exactly nine coordinate systems associated with nine orbits of second order operators in the enveloping algebra of $SL(2,R)$. Here $SL(2,R)$ is the local symmetry group of the EPD equation.

In this paper we undertake a detailed study of these coordinate systems and show how one can use the representation theory of $SL(2,R)$ and its universal covering group to derive special function identities relating separable solutions corresponding to distinct coordinate systems.

In Sec. 1 we compute the symmetry algebra $sl(2,R)$ of the EPD equation and show that we can introduce a Hilbert space structure on the solution space of the EPD equation such that this space transforms according to a unitary irreducible representation of the universal covering group of $SL(2,R)$, taken from the discrete series. We also relate the space to two other models of this representation which are more useful for computational purposes.

In Secs. 2 and 3 we classify the nine possible coordinate systems such that variables separate in (0.1) and relate them to nine orbits of symmetric second order operators in the enveloping algebra of $sl(2,R)$. We also compute the spectral resolutions of these operators. In Sec. 4 we use our earlier results to compute the separable solutions of (0.1), and in Sec. 5 we determine overlap functions relating various distinct bases.

Finally, in Sec. 6 we discuss the separation of variable problem for the complex EPD equation and show that this equation permits separation in exactly five coordinate systems corresponding to five orbits of symmetric second order operators in the enveloping algebra of $sl(2,\mathbb{C})$. We relate these results to a paper by Viswanathan,[2] which employs Weisner's method[3,4] to derive generating functions for Gegenbauer polynomials. For a slightly different approach to the complex EPD equation, see Ref. 5.

All special functions appearing in this paper are defined as in the Bateman Project.[6]

1. SYMMETRIES OF THE EPD EQUATION

The symmetry algebra of the EPD equation

$$[\partial_{tt}-\partial_{rr}-(1/r)\,\partial_r+m^2/r^2]\,\Phi(t,r)=0,\qquad r\geq 0,\quad -\infty<t<\infty, \tag{1.1}$$

is the set of all linear differential operators

$$L=a_1(t,r)\,\partial_t+a_2(t,r)\,\partial_r+b(t,r)$$

such that $L\Phi$ is a (local) solution of (1.1) whenever Φ is a (local) solution. By using standard techniques in Lie theory,[1,4] it is straightforward to show that this algebra is isomorphic to $sl(2,R)$. Indeed, the operators A, B, C form a basis where

$$A=\tfrac{1}{2}[(1-t^2-r^2)\,\partial_t-2tr\partial_r-t],$$
$$B=-(\tfrac{1}{2}+t\partial_t+r\partial_r), \tag{1.2}$$
$$C=\tfrac{1}{2}[(1+t^2+r^2)\,\partial_t+2tr\partial_r+t].$$

Note that $A+C=\partial_t$. Here, we are ignoring the trivial symmetry E of multiplication by the scalar one: $E\Phi=\Phi$. The commutation relations are

$$[A,B]=-C,\quad [A,C]=-B,\quad [B,C]=A. \tag{1.3}$$

We can express (1.1) in terms of the Lie algebra generators with the result

$$(C^2-A^2-B^2)\,\Phi=(\tfrac{1}{4}-m^2)\,\Phi, \tag{1.1'}$$

where $C^2-A^2-B^2$ is the Casimir operator for $\mathrm{sl}(2,R)$.

By definition, $\mathrm{sl}(2,R)$ is the Lie algebra of 2×2 real traceless matrices. We choose the isomorphism between our symmetry algebra and $\mathrm{sl}(2,R)$ such that the operators A,B,C correspond to the matrices $\mathcal{A},\mathcal{B},\mathcal{C}$, respectively, where

$$\mathcal{A}=\begin{pmatrix}0 & \frac{1}{2}\\ \frac{1}{2} & 0\end{pmatrix},\quad \mathcal{B}=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{pmatrix},\quad \mathcal{C}=\begin{pmatrix}0 & \frac{1}{2}\\ -\frac{1}{2} & 0\end{pmatrix}. \tag{1.4}$$

Then, using standard results from Lie theory,[4] we find that the operators (1.2) exponentiate to a local Lie representation of the group $\mathrm{SL}(2,R)$ by operators $T(G)$, where

$$T(G)\,\Phi(t,r)=[(\alpha+\gamma t)^2-\gamma^2r^2]^{-1/2}\times\Phi\left[\frac{(\delta t+\beta)(\alpha+\gamma t)-\gamma\delta r^2}{(\alpha+\gamma t)^2-\gamma^2r^2},\ \frac{r}{(\alpha+\gamma t)^2-\gamma^2r^2}\right],$$
$$G=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in \mathrm{SL}(2,R). \tag{1.5}$$

Here, $\mathrm{SL}(2,R)$ is a local symmetry group of (1.1) in the sense that if Φ is a local solution of the EPD equation, then $T(G)\,\Phi$ is also a local solution.

Motivated by the connection between the EPD equation and the wave equation discussed in Refs. 1, we note that for any C^∞ function $f(k)$ with compact support in $(0,\infty)$ the corresponding function

$$\Phi(t,r)=\exp(-im\pi/2)\int_0^\infty \exp(itk)\,J_m(kr)f(k)\,dk=\cup[f] \tag{1.6}$$

is a solution of (1.1). If we introduce the inner product

$$\langle f_1,f_2\rangle=\int_0^\infty f_1(k)\bar{f}_2(k)\,dk \tag{1.7}$$

on the space of C^∞ functions, we find that, in terms of the corresponding solutions $\Phi_i(t,r)$ of (1.1), the inner product reads

$$(\Phi_1,\Phi_2)\equiv\langle f_1,f_2\rangle=i\int_0^\infty \Phi_1(r,t)\,\partial_t\bar{\Phi}_2(r,t)\,r\,dr=-i\int_0^\infty \bar{\Phi}_2(r,t)\,\partial_t\Phi_1(r,t)\,r\,dr. \tag{1.8}$$

Here, the last two integrals are actually independent of t.

It follows from this that if we complete our pre-Hilbert space of C^∞ functions f to form the Hilbert space $L_2(0,\infty)$, the space of corresponding functions $\Phi=\cup[f]$ defined formally by (1.6) will form a Hilbert space $\mathcal{H}$ of weak solutions of the EPD equation with inner product (1.8). The transformation $\cup$ determines a unitary equivalence between these Hilbert spaces.

The action of the symmetry algebra $\mathrm{sl}(2,R)$ on $\mathcal{H}$ is given formally by (1.2). Indeed, these expressions make sense when applied to the dense subspace of $\mathcal{H}$ consisting of those solutions Φ which arise from C^∞ functions f with compact support on $(0,\infty)$. Moreover, as we shall see, they define symmetric operators on this subspace. The operators $\cup^{-1}K\cup$ on $L_2(0,\infty)$ related to operators $K\in\mathrm{sl}(2,R)$ on $\mathcal{H}$ are easily determined using integration by parts:

$$A=\frac{i}{2}k\left(\frac{d^2}{dk^2}+\frac{1}{k}\frac{d}{dk}-\frac{m^2}{k^2}+1\right),$$
$$B=\frac{1}{2}+k\frac{d}{dk}, \tag{1.9}$$
$$C=\frac{i}{2}k\left(-\frac{d^2}{dk^2}-\frac{1}{k}\frac{d}{dk}+\frac{m^2}{k^2}+1\right).$$

[We are using the same letter to designate corresponding operators on $\mathcal{H}$ and $L_2(0,\infty)$.] It is now straightforward to show that iA, iB, and iC are essentially self-adjoint on $L_2(0,\infty)$. Moreover, it is well known that these operators determine a unitary irreducible representation of the universal covering group of $\mathrm{SL}(2,R)$ from the discrete series.[7,8]

Indeed, it is easy to check that C has discrete spectrum $i\lambda=i(m+s+1/2)$, $s=0,1,2,\cdots$ with a corresponding ON basis for $L_2(0,\infty)$ consisting of eigenfunctions

$$f_s^{(1)}(k)=\left(\frac{2\Gamma(s+1)}{\Gamma(2m+s+1)}\right)^{1/2}(2k)^m\exp(-k)\,L_s^{(2m)}(2k),$$
$$Cf_s^{(1)}=i(m+s+\tfrac{1}{2})f_s^{(1)},\quad \langle f_s^{(1)},f_{s'}^{(1)}\rangle=\delta_{ss'}. \tag{1.10}$$

From this fact and the relation

$$C^2-A^2-B^2=\tfrac{1}{4}-m^2 \tag{1.11}$$

we see that the operators (1.9) determine the irreducible representation $D^-_{m-1/2}$ from the negative discrete series.[4,7,8]

For $2m$ an integer this Lie algebra representation exponentiates to a single-valued unitary irreducible representation of $\mathrm{SL}(2,R)$ defined by unitary operators $T(G)$, where

$$T(G)f(k)=-\gamma^{-1}\exp[i(k\alpha/\gamma+\pi/2-\pi m)]\times\int_0^\infty \exp(il\delta/\gamma)\,J_{2m}((2/\gamma)\sqrt{kl})f(l)\,dl,\quad \gamma\neq 0,$$
$$T(G)f(k)=\alpha\exp(ik\alpha\beta)f(\alpha^2k),\quad \gamma=0,$$
$$G=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\in\mathrm{SL}(2,R),\quad f\in L_2(0,\infty). \tag{1.12}$$

For $2m$ not an integer, operators (1.12) define a multiple-valued representation of $\mathrm{SL}(2,R)$. In this case we obtain a single-valued representation of the universal covering group of $\mathrm{SL}(2,R)$, and expressions (1.12) are valid only in a neighborhood of the identity element of the covering group. For a discussion of the parametrization of the covering group see Refs. 7, 9.

There is another model of $D^-_{m-1/2}$, due to Bargmann,[9] which we will also find useful. This model is defined on the Hilbert space $\mathcal{H}_m$ of all functions $\mathcal{J}(z)=\sum_{n=0}^{\infty} a_n z^n$, $a_n \in \mathbb{C}$, analytic in the disk $|z|<1$ and such that

$$\lim_{l\to 2m+1} [(l-1)/\pi] \int_{|z|<1} |\mathcal{J}(z)|^2 (1-z\bar{z})^{l-2}\, dx\, dy < \infty, \qquad l>1,\ z=x+iy.$$

The inner product is

$$\langle \mathcal{J}, K\rangle_m = \lim_{l\to 2m+1} [(l-1)/\pi] \times \int_{|z|<1} \mathcal{J}(z)\overline{K}(z)(1-z\bar{z})^{l-2}\, dx\, dy \tag{1.13}$$

and a convenient ON basis is provided by the monomials

$$\mathcal{J}_s(z) = [\Gamma(2m+s+1)/\Gamma(2m+3)\, s!\,]^{1/2} z^s, \qquad s=0,1,2,\cdots. \tag{1.14}$$

The operators A, B, C corresponding to (1.2) and (1.9) are

$$A=\frac{i}{2}\left\{(1+z^2)\frac{d}{dz}+(2m+1)z\right\},$$
$$B=\frac{1}{2}\left\{(1-z^2)\frac{d}{dz}-(2m+1)z\right\}, \tag{1.15}$$
$$C=i\left\{z\frac{d}{dz}+m+\frac{1}{2}\right\}.$$

For more details about this representation, see Refs. 7, 9. Since $C\mathcal{J}_s = i(m+s+\frac{1}{2})\mathcal{J}_s$, it follows that the basis vectors $f_s^{(1)}(k)$ and $\mathcal{J}_s$ correspond. The unitary mapping V from $\mathcal{H}_m$ onto $L_2(0,\infty)$ which carries $\mathcal{J}_s$ to $f_s^{(1)}$ and the operators (1.15) to (1.9) is

$$V\mathcal{J}(k) = \langle \mathcal{J}, V(k,\cdot)\rangle_m, \quad \mathcal{J}\in\mathcal{H}_m,$$
$$V(k,z) = \sum_{s=0}^{\infty} f_s^{(1)}(k)\mathcal{J}_s(z) \tag{1.16}$$
$$= [2/\Gamma(2m+3)]^{1/2}(2k)^m (1-z)^{-2m-1} \times \exp(-k)\exp[-2kz/(1-z)].$$

Similarly, the unitary mapping $W=UV$ from $\mathcal{H}_m$ onto $\mathcal{H}$ is

$$W\mathcal{J}(t,r) = \langle \mathcal{J}, W(t,r,\cdot)\rangle_m, \quad \mathcal{J}\in\mathcal{H}_m,$$
$$W(t,r,z) = [2^{2m} r^m \Gamma(m+\tfrac{1}{2})\exp(+im\pi/2)/\sqrt{2\pi\Gamma(2m+3)}] \times \{[(1+it)(1-z)+2z]^2 + (1-z)^2 r^2\}^{-m-1/2}.$$

2. THE SEPARABLE COORDINATE SYSTEMS

As shown in Refs. 1 the EPD equation permits separation or R-separation of variables in nine coordinate systems corresponding exactly to the nine SL(2,R)-orbits in the space $\mathcal{S}/\{C^2-A^2-B^2\}$, where $\mathcal{S}$ is the space of symmetric second-order elements in the universal enveloping algebra of sl(2,R). A particular separated solution Φ is characterized by (1.1) and the eigenvalue equation $S\Phi=\lambda\Phi$. The eigenvalue λ is the separation constant. If two operators S, S' are on the same orbit, i.e., if $cS'=T(G)ST(G^{-1})$, where $c\in R$, $c\neq 0$ and $G\in$ SL(2,R), then the coordinate systems associated with S and S' are considered equivalent. A complete list of orbit representatives is

$$\begin{aligned}
&1].\ C^2,\\
&2].\ (A+C)^2,\\
&3].\ B^2,\\
&4].\ 2A^2+AC+CA-aB^2,\quad a>-1,\\
&5].\ 2C^2+AC+CA+aB^2,\quad a>-1,\\
&6].\ rB^2+AC+CA,\quad 0\le r<\infty,\\
&7].\ B^2-s^2C^2,\quad 0<s^2<1,\\
&8].\ C^2+k^2B^2,\quad 0<k^2<\infty,\\
&9].\ (A+C)B+B(A+C)
\end{aligned} \tag{2.1}$$

(In case 4] and 5] we shall always normalize so that $a=0$).

It is clear that the operators 1]–9] are symmetric in the $L_2(0,\infty)$ model of the representation $D^-_{m-1/2}$. [Here we consider these operators as initially defined on the dense subspace of C^∞ functions with compact support in $(0,\infty)$.] In this section we will determine the spectral resolutions of six of these operators in this model. Operators 6], 7], and 8] are most conveniently studied in the $\mathcal{H}_m$ model and will be treated in the following section.

System 1] has been treated above. It is straightforward to show that the operator iC has deficiency indices (0, 0). Thus iC is essentially self-adjoint. The spectrum of the closure of C is $i(m+s+\frac{1}{2})$, $s=0,1,2,\cdots$, and each of these discrete eigenvalues has multiplicity one. The corresponding ON basis of eigenvectors is listed in (1.10).

For system 2] we have $A+C=ik$. Clearly the closure of $-i(A+C)$ has continuous spectrum covering the positive real axis and a basis of generalized eigenfunctions

$$f_\lambda^{(2)}(k)=\delta(k-\lambda),\quad 0<\lambda<\infty,$$
$$\langle f_\lambda^{(2)}, f_{\lambda'}^{(2)}\rangle = \delta(\lambda-\lambda'),\quad (A+C)f_\lambda^{(2)} = i\lambda f_\lambda^{(2)}, \tag{2.2}$$

where $\delta(\lambda)$ is the Dirac delta function.

For system 3] we have $B=\frac{1}{2}+kd/dk$. It is easy to show that the closure of $-iB$ is self-adjoint with continuous spectrum covering the real axis and a basis of generalized eigenfunctions

$$f_\mu^{(3)}(k) = (2\pi)^{-1/2} k^{i\mu-1/2},\quad -\infty<\mu<\infty$$
$$\langle f_\mu^{(3)}, f_{\mu'}^{(3)}\rangle = \delta(\mu-\mu'),\quad Bf_\mu^{(3)} = i\mu f_\mu^{(3)}. \tag{2.3}$$

System 4] with $a=0$ is more interesting. Here the operator $L=2A^2+AC+CA$ is symmetric with deficiency indices (1, 1). The possible self-adjoint extensions L_α can be parametrized by the real number α, $0<\alpha\le 2$. For each α, L_α has discrete spectrum $\lambda=m^2-(\alpha+2s)^2-\frac{1}{4}$, $s=0,1,2,\cdots$, each eigenvalue of multiplicity one, and continuous spectrum. The normalized eigenvectors $f^{(4)}_{\alpha,s}(k)$ form an ON set for $L_2(0,\infty)$:

$$f^{(4)}_{\alpha,s}(k) = [2(\alpha+2s)k^{-1}]^{1/2} J_{\alpha+2s}(k),\quad s=0,1,\cdots,$$
$$\langle f^{(4)}_{\alpha,s}, f^{(4)}_{\alpha,s'}\rangle = \delta_{ss'},\quad L_\alpha f^{(4)}_{\alpha,s} = [m^2-(\alpha+2s)^2-\tfrac{1}{2}]f^{(4)}_{\alpha,s}. \tag{2.4}$$

The overlaps between distinct self-adjoint extensions L_α, $L_{\alpha'}$, $\alpha\neq\alpha'$, are given by

$$\langle f^{(4)}_{\alpha,s}, f^{(4)}_{\alpha',s'}\rangle = \frac{\sqrt{(\alpha+2s)(\alpha'+2s')}\sin\pi[(\alpha-\alpha')/2+s-s']}{\pi[(\alpha-\alpha')/2+s-s'][(\alpha+\alpha')/2+s+s']}. \quad (2.5)$$

Restricting ourselves to the case $\alpha=2$ for simplicity, we find that L_2 has continuous spectrum $\lambda=m^2+\beta-\frac{1}{4}$, $\beta\geq 0$, with corresponding generalized eigenfunctions

$$\tilde{f}^{(4)}_{2,\beta}(k)=[J_{i\sqrt{\beta}}(\gamma k)+J_{-i\sqrt{\beta}}(\gamma k)]/2\sqrt{k\sinh(\pi\sqrt{\beta})},$$
$$\langle \tilde{f}^{(4)}_{2,\beta}, \tilde{f}^{(4)}_{2,\beta'}\rangle=\delta(\beta-\beta').$$

The functions $\{f^{(4)}_{\alpha,s}, \tilde{f}^{(4)}_{\alpha,\beta}\}$ form a complete set for $L_2(0,\infty)$. More details can be found on pp. 93–95 of Ref. 10.

For system 5] with $a=0$ we find that the operator $M=2C^2+AC+CA$ is essentially self-adjoint. The closure of M (which we also call M) has continuous spectrum $\lambda=\frac{1}{4}-m^2-\mu^2$, $\mu\geq 0$, of multiplicity one and a basis of generalized eigenvectors.

$$f^{(5)}_{\mu}(k)=[\pi/\sqrt{2k\mu\sinh k(\mu\pi)}]K_{i\mu}(k), \quad 0<\mu<\infty,$$
$$\langle f^{(5)}_{\mu}, f^{(5)}_{\mu'}\rangle=\delta(\mu-\mu'), \quad Mf^{(5)}_{\mu}=(\tfrac{1}{4}-m^2-\mu^2)f^{(5)}_{\mu}. \quad (2.6)$$

For system 9] we find that the operator $N=(A+C)B+B(A+C)$ has unequal deficiency indices $(1,0)$. However, there exists an obvious extension of N to the space $L_2(R)=L_2(-\infty,0)\oplus L_2(0,\infty)$ with deficiency indices $(1,1)$. Of the self-adjoint extensions of this latter operator we choose the one with continuous spectrum covering the real axis and generalized eigenfunctions:

$$f^{(9)}_{\lambda}(k)=\exp(i\lambda/k)/k\sqrt{2\pi}, \quad -\infty<\lambda<\infty,$$
$$Nf^{(9)}_{\lambda}=2\lambda f^{(9)}_{\lambda}, \quad \int_{-\infty}^{\infty} f^{(9)}_{\lambda}(k)\bar{f}^{(9)}_{\lambda'}(k)\,dk=\delta(\lambda-\lambda'). \quad (2.7)$$

Note that $\{f^{(9)}_{\lambda}\}$ satisfies the usual orthogonality relations on $L_2(R)$ but not on $L_2(0,\infty)$.

3. LAMÉ BASES

The spectral resolution of the operators 6], 7], and 8] is carried out in this section using the model of $D^{-}_{m-1/2}$ due to Bargmann,[9] defined on the Hilbert space $\mathcal{H}_m$ as given in Sec. 1. The reason for treating these operators in this model rather than the $L_2(0,\infty)$ Hilbert space model with SL(2,R) generators as in (1.9) is that they are second order differential operators rather than fourth order.

In fact, if we consider the functions $\mathcal{G}(z)$ defined by $\mathcal{F}(z)=z^{m-1/2}\mathcal{G}(z)$, where $\mathcal{F}(z)\in\mathcal{H}_m$, and put $z=ie^{i\theta}$ with θ complex, the generators A, B, and C acting on the functions $\mathcal{G}(z)$ have the form

$$A=-\sin\theta\frac{d}{d\theta}+(m-\tfrac{1}{2})\cos\theta,$$
$$B=-\cos\theta\frac{d}{d\theta}-(m-\tfrac{1}{2})\sin\theta, \quad C=\frac{d}{d\theta}, \quad (3.1)$$

The form of the generators (3.1) is the same as used in Ref. 11, where the bases described by the operators 1]–9] were studied for the principal series of SL(2, R). The appropriate variables which change eigenvalue equations for the operators 6], 7], and 8] to Lamé equations have been given in that article. The spectral resolution for each of these operators also follows along the lines of our previous article. We now discuss each of the three Lamé bases in the order 8], 7], and 6], treating the simplest cases first.

For the coordinate system 8] it is convenient to choose the functions $\Lambda(v)$ defined by

$$\mathcal{G}(z)=[\mathrm{dn}(v,s)]^{1/2-m}\Lambda(v), \quad (3.2)$$

where

$$\cos\theta=\frac{1}{(1+k^2)^{1/2}}\frac{\mathrm{sn}(v,s)}{\mathrm{dn}(v,s)}$$

and $s=k/(1+k^2)^{1/2}$. The eigenvalue equation for this coordinate system is then

$$\left(\frac{d^2}{dv^2}+s^2(m^2-\tfrac{1}{4})\mathrm{cn}^2(v,s)-\frac{\lambda}{1+k^2}\right)\Lambda_{\lambda}(v)=0. \quad (3.3)$$

Suitable eigenfunctions are the Lamé Wangerin functions with boundary conditions

(i) $[\mathrm{sn}(v,s)]^{1/2}\Lambda_{\lambda}(v)$ bounded at $v=iK'$ and $\Lambda'_{\lambda}(K+iK')=0$.

This gives the solution $\Lambda_{\lambda}(v)=F^{2p}_{m-1/2}(v,s)$ with $2p$ zeros in the interval $(iK', iK'+2K)$.

(ii) $[\mathrm{sn}(v,s)]^{1/2}\Lambda_{\lambda}(v)$ bounded at $v=iK'$ and $\Lambda_{\lambda}(K+iK')=0$ giving the solution $\Lambda_{\lambda}(v)=F^{2p+1}_{m-1/2}(v,s)$ with $2p+1$ zeros in the interval $(iK', iK'+2K)$.

It should also be mentioned that the resulting eigenfunctions

$$\mathcal{F}^{(8)}_{\lambda}(z)=\{[s'\mathrm{sn}(v,s)+i\mathrm{cn}(v,s)]/s'\mathrm{dn}^2(v,s)\}^{m-1/2}\Lambda_{\lambda}(v), \quad (3.4)$$

where $s'=(1-s^2)^{1/2}$, are analytic inside the unit circle of the complex z plane and are elements of $\mathcal{H}_m (m=1,2,3,\cdots)$.

For the coordinate system 7] we choose the functions $\mathcal{H}(v)$ defined by

$$\mathcal{G}(z)=[is'/\mathrm{cn}(v,s)]^{m-1/2}\mathcal{H}(v), \quad (3.5)$$

where

$$\cos\theta=\mathrm{dn}(v,s)/\mathrm{cn}(v,s).$$

The eigenvalue equation for this coordinate system then becomes

$$\left(\frac{d^2}{dv^2}-s^2(m^2-\tfrac{1}{4})\mathrm{sn}^2(v,s)-\lambda\right)\mathcal{H}_{\lambda}(v)=0. \quad (3.6)$$

Natural choices for eigenfunctions are the Lamé Wangerin functions with boundary conditions:

(i) $[\mathrm{sn}(v,s)]^{1/2}\mathcal{H}_{\lambda}(v)$ bounded at $v=iK$ and $\mathcal{H}_{\lambda}(K+iK')=0$ giving the solution $\mathcal{H}_{\lambda}(v)=F^{2p}_{m-1/2}(v,s)$ with $2p$ zeros in the interval $(iK', iK'+2K)$;

(ii) $[\mathrm{sn}(v,s)]^{1/2}\mathcal{H}_{\lambda}(v)$ bounded at $v=iK'$ and $\mathcal{H}_{\lambda}(K+iK')=0$ giving the solution $\mathcal{H}_{\lambda}(v)=F^{2p+1}_{m-1/2}(v,s)$ with $2p+1$ zeros in the interval $(iK', iK'+2K)$.

In each case the spectrum is discrete and the eigenvalues are labeled by the index p, $p=0,1,2,\cdots$. The resulting eigenfunctions

$$\mathcal{F}^{(7)}_{\lambda}(z)=\{s'[s'\mathrm{sn}(v,s)-\mathrm{dn}(v,s)]/\mathrm{cn}^2(v,s)\}^{m-1/2}\mathcal{H}_{\lambda}(v) \quad (3.7)$$

are analytic inside the unit circle of the complex z plane and belong to H_m. The coordinate system 6] is the most complicated of the three under consideration. A convenient choice of function $N(v)$ is

$$\mathcal{G}(z)=\left(\frac{N\mathrm{sn}(v,s)\mathrm{dn}(v,s)}{[-(1+r^2)^{1/2}+r+1]\mathrm{sn}^2(v,s)-2r}\right)^{m-1/2}N(v) \tag{3.8}$$

where

$$\sin\theta=\frac{2[1-(1+r^2)^{1/2}]+[(1+r^2)^{1/2}-1-r]\mathrm{sn}^2(v,s)}{[1+r-(1+r^2)^{1/2}]\mathrm{sn}^2(v,s)-2r}$$

and

$$s^2=\frac{(1+r^2)^{1/2}-r}{2(1+r^2)^{1/2}},\quad N=\frac{8(1+r^2)^{1/2}}{r^2}[(1+r^2)^{1/2}-1].$$

The eigenvalue equation for this coordinate system is

$$\left(\frac{d^2}{d\omega^2}-t^2(m^2-\tfrac{1}{4})\mathrm{sn}^2(\omega,t)+\frac{(m^2-\frac{1}{4})r}{(1+r^2)^{1/2}(s-is')^2}+\frac{(1+r^2)^{1/2}\lambda}{(s-is')^2}\right)N_\lambda(\omega)=0. \tag{3.9}$$

Here we have introduced the variables, $\omega=(s+is')v+iK'(t)$ and $t=(s+is')/(-s+is')$. The resulting equation is of the Lamé type with modulus t on the unit circle. Natural choices for eigenfunctions are the Lamé Wangerin functions with boundary conditions as for the coordinate system 7], where v is replaced by ω and r by t. The eigenfunctions are then $N_\lambda(\omega)=F^{e}_{m-1/2}(\omega,t)$ with $p=0,1,2,\cdots$, and the spectrum is discrete. The corresponding eigenfunctions $\mathcal{F}^{(6)}_\lambda(z)$ are analytic in the unit circle in the complex z plane and are members of H_m.

We see that for the discrete series $D^-_{m-1/2}$ of SL(2, R) the most convenient basis eigenfunctions for the three Lamé bases are the Lamé Wangerin or finite Lamé functions. This is opposed to the situation in Ref. 11, where we dealt with the principal series of SL(2, R) and the corresponding basis functions for system 8] were the periodic Lamé functions. For the discrete series $D^-_{m-1/2}$ in the Bargmann model the operator specifying system 8] is singular inside the unit disc. The operators of the other two systems are singular on the unit disc. The imposition of boundary conditions that gives Lamé Wangerin functions for these systems yields eigenfunctions in H_m, which are analytic inside the unit disc and zero at the singular points.

4. THE TWO-VARIABLE MODEL

If $\{f^{(j)}_\lambda(k)\}$ is a basis for $L_2(0,\infty)$ consisting of eigenfunctions of the operator S, symmetric and second order in the generators (1.9) of sl(2, R), then $\{F^{(j)}_\lambda(t,r)\}$ is a basis for H consisting of eigenfunctions of the corresponding operator S' constructed from the generators (1.2), where

$$F^{(j)}_\lambda(t,r)=\cup[f^{(j)}_\lambda],\quad j=1,\ldots,9, \tag{4.1}$$

and $\cup$ is the unitary transformation (1.6). Indeed we have the relations

$$Sf^{(j)}_\lambda=\lambda f^{(j)}_\lambda,\quad S'F^{(j)}_\lambda=\lambda F^{(j)}_\lambda,$$
$$\langle f^{(j)}_\lambda, f^{(j)}_{\lambda'}\rangle=(\cup f^{(j)}_\lambda,\cup f^{(j)}_{\lambda'})=(F^{(j)}_\lambda,F^{(j)}_{\lambda'})=\delta_{\lambda\lambda'}. \tag{4.2}$$

Furthermore, $F^{(j)}_\lambda(t,r)$ is a solution of the EPD equation (1.1). It follows from results proved in Refs. 1 that for fixed j there exists a coordinate system $\{u(t,r),v(t,r)\}$ such that variables separate (or R-separate) in the EPD equation and such that $F^{(j)}_\lambda(t,r)=\exp[Q(u,v)]J_\lambda(u)K_\lambda(v)$, where J_λ, K_λ are solutions of the separated second order ordinary differential equations and either $Q\equiv 0$ (separation) or $Q\neq 0$ and Q cannot be expressed as a sum $Q(u,v)=q_1(u)+q_2(v)$ (R-separation). In particular the possible coordinate systems and separated equations (as well as the functions Q) are listed in Ref. 1, Paper 9.

In this section it is our aim to compute all the functions $F^{(j)}_\lambda(t,r)$ defined by (4.1) and (1.6). In general, the integrals (4.1) are rather difficult to evaluate. In particular we have not been able to find the integral for $j=5$ nor two of the three integrals needed for $j=4$ in the Bateman Project.

However, our work is enormously simplified because we know in advance the coordinates in which variables separate for (4.1). Thus we can immediately evaluate the integral as a linear combination of four terms (since J_λ and K_λ each satisfy a known second order ordinary linear differential equation). The four constants can then be determined by evaluating the integral for specially chosen values of the variables u, v. In this respect the functions $F^{(j)}_\lambda(t,r)$ listed below can also be regarded as evaluations of a number of interesting integrals related to the EPD equation.

For several cases we find that the integral (4.1) does not converge sufficiently rapidly so that differentiation under the integral sign is permitted. It is not immediately apparent in these cases that $F^{(j)}_\lambda$ is actually a solution of the EPD equation. However, it is always possible to justify our assertions by noting that if we allow t to become complex and take $\mathrm{Im}\,t>0$ in (1.6), then the convergence in each integral (4.1) is sufficiently rapid that multiple differentiation with respect to r and t is permitted under the integral sign. In each case one can verify by inspection that the coordinates $u(t,r)$, $v(t,r)$ can be extended to the domain $\mathrm{Im}\,t\geq 0$ and that variables still separate in the EPD equation. Finally one can evaluate the integrals (4.1) for $\mathrm{Im}\,t>0$ and then use the Lebesgue dominated convergence theorem or a similar device to justify going to the limit as t becomes real through positive imaginary values.

We have the following results:

1]: $F^{(1)}_s(t,r)\equiv F^{(1)}_s(\sigma,\varphi)$

$$=\left(\frac{2(s!)}{\Gamma(2m+s+1)}\right)^{1/2}\frac{\Gamma(2m+1)}{\Gamma(m+1)}2^{-m-1/2}\times\exp[-i(m-1)\pi/2]\sqrt{\cos\sigma-\cos\varphi}\,\sin^m\sigma\times\exp[-i(s+m+1/2)\varphi]C^{(m+1/2)}_s(\cos\sigma),$$

$s=0,1,2,\cdots$.

$$t=\frac{\sin\varphi}{\cos\sigma-\cos\varphi},\quad r=\frac{\sin\sigma}{\cos\sigma-\cos\varphi},$$

$$0 \leq \sigma \leq \pi, \quad 2\pi - \sigma > \varphi > \sigma, \tag{4.3}$$

$$e^{i\varphi} = \frac{\frac{1}{2}(t^2 - r^2 - 1) + it}{[r^2 + \frac{1}{4}(r^2 - t^2 - 1)^2]^{1/2}},$$

$$e^{i\sigma} = \frac{\frac{1}{2}(t^2 - r^2 + 1) + ir}{[r^2 + \frac{1}{4}(r^2 - t^2 - 1)^2]^{1/2}},$$

$$e^{Q} = \sqrt{\cos\sigma - \cos\varphi}, \quad (F_s^{(1)}, F_{s'}^{(1)}) = \delta_{ss'}.$$

Here, $C_s^{(\nu)}(z)$ is a Gegenbauer polynomial.

2]: $F_\lambda^{(2)}(t, r) = \exp(-im\pi/2)\exp(it\lambda)J_m(\lambda r), \quad \lambda > 0,$

$$u = t, \quad v = r, \quad Q \equiv 0, \quad (F_\lambda^{(2)}, F_{\lambda'}^{(2)}) = \delta(\lambda - \lambda'). \tag{4.4}$$

3]:

$$F_\mu^{(3)}(t, r) = \exp[\pm i\pi(m + 1/2 + i\mu)/2](\pm t)^{-m-i\mu-1/2}$$

$$\times \frac{\exp(im\pi/2)r^m}{\Gamma(m+1)\sqrt{2\pi}}\Gamma(m + i\mu + \tfrac{1}{2}) \tag{4.5}$$

$$\times {}_2F_1(i\mu/2 + m/2 + \tfrac{1}{4}, i\mu/2 + m/2 + \tfrac{3}{4}; m + 1; r^2/t^2),$$

$$r < |t|,$$

$$F_\mu^{(3)}(t, r) = [\exp(-im\pi/2)/\sqrt{2\pi}]\{2^{i\mu-1/2}$$

$$[\Gamma(m/2 + i\mu/2 + \tfrac{1}{4})r^{-m}/\Gamma(m/2 - i\mu/2 + \tfrac{3}{4})]$$

$$\times {}_2F_1(i\mu/2 + m/2 + \tfrac{1}{4}, i\mu/2 - m/2 + \tfrac{1}{4}; \tfrac{1}{2}; t^2/r^2)$$

$$+ 2^{i\mu+1/2}it r^{-i\mu-3/4}[\Gamma(m/2 + i\mu/2 + \tfrac{3}{4})/\Gamma(m/2 + i\mu/2 + \tfrac{1}{4})]$$

$$\times {}_2F_1(i\mu/2 + m/2 + \tfrac{3}{4}; i\mu/2 - m/2 + \tfrac{3}{4}; \tfrac{3}{2}; t^2/r^2)\},$$

$$r > |t|,$$

Here the (+) sign holds for $t > 0$ and the (−) sign for $t < 0$. In this case $u = t$, $v = r/t$, $Q \equiv 0$, and $(F_\mu^{(3)}, F_{\mu'}^{(3)}) = \delta(\mu - \mu')$.

For systems of type 4] we consider three cases.

4a]: $|t| \geq r + 1$: For $t \geq 1$ we have $(\nu = \alpha + 2s)$

$$F_{\alpha,s}^{(4)}(t, r) = F_{\alpha,s}^{(4)}(\theta, \varphi)$$

$$= \exp[i\pi(\nu/2 - m + \tfrac{1}{4})]\sqrt{2/\pi} \tag{4.6}$$

$$\times P_{\nu-1/2}^{-m}(\cosh\theta)Q_{\nu-1/2}^{m}(\cosh\varphi),$$

$s = 0, 1, 2, \ldots$, for $\theta \geq \varphi$, and $F_{\alpha,s}^{(4)}(\theta, \varphi) = F_{\alpha,s}^{(4)}(\varphi, \theta)$ for $\theta < \varphi$. Here P_ν^μ and Q_ν^μ are Legendre functions and

$$t = \cosh\theta\cosh\varphi, \quad r = \sinh\theta\sinh\varphi, \quad \theta, \varphi \geq 0. \tag{4.7}$$

For $t \leq -1$ we have $F_{\alpha,s}^{(4)}(t, r) = \exp(-im\pi)\overline{F_{\alpha,s}^{(4)}(-t, r)}$.

4b]: $|t| \leq r - 1$:

$$F_{\alpha,s}^{(4)}(\theta, \varphi) = \frac{\exp[i\pi(-5m/2 + \nu + \frac{1}{2})]2^{\nu-m-1}\Gamma(\frac{3}{4} + \nu/2 - m/2)}{\pi\Gamma(\frac{1}{2} + \nu + m)\Gamma(\frac{3}{4} + m/2 - \nu/2)}$$

$$Q_{\nu-1/2}(i\sinh\theta)Q_{\nu-1/2}(i\sinh\varphi), \tag{4.8}$$

$$t = \sinh\theta\sinh\varphi, \quad r = \cosh\theta\cosh\varphi, \quad -\infty < \theta < \infty$$

4c]: $|t| + r \leq 1$: For $t \geq 0$,

$$F_{\alpha,s}^{(4)}(\theta, \varphi) = \frac{\sqrt{\pi\nu}\,\Gamma(\nu + m + \frac{1}{2})\exp(-im\pi/2)}{\Gamma(\nu - m + \frac{1}{2})\cos[(\pi/2)(\nu - m - \frac{1}{2})]}$$

$$\times P_{\nu-1/2}^{-m}(\cos\theta)P_{\nu-1/2}^{-m}(\cos\varphi), \quad \nu = \alpha + 2s, \tag{4.9}$$

$t = \cos\theta\cos\varphi, \quad r = \sin\theta\sin\varphi, \quad 0 \leq \theta, \varphi \leq \pi/2.$

For $t < 0$, $F_{\alpha,s}^{(4)}(t, r) = \exp(-im\pi)\overline{F_{\alpha,s}^{(4)}(-t, r)}$

These three parametrizations do not cover the full r–t plane but, as shown in Ref. 1, Paper 9, variables do not separate in the remaining domain. We omit the computation of the continuum eigenfunctions $\tilde{F}_{\alpha,\beta}^{(4)}$

5]: $F_\mu^{(5)}(t, r) \equiv F_\mu^{(5)}(\theta, \varphi)$

$$= \frac{\pi^{3/2}\Gamma(\frac{1}{2} - i\mu + m)\Gamma(\frac{1}{2} + i\mu + m)}{2\sqrt{k\mu\sinh(\mu\pi)}}$$

$$\times P_{-1/2+i\mu}^{-m}(\cosh\theta)P_{-1/2+i\mu}^{-m}(-i\sinh\varphi),$$

$$0 < \mu < \infty, \tag{4.10}$$

$t = \cosh\theta\sinh\varphi, \quad r = \sinh\theta\cosh\varphi, \quad 0 \leq \theta, \varphi.$

For $t < 0$ we have $F_\mu^{(5)}(t, r) = \exp(-im\pi)\overline{F_\mu^{(5)}(-t, r)}$.

9]:

$$F_\lambda^{(9)}(t, r) = F_\lambda^{(9)}(x, X)$$

$$= \begin{cases} \sqrt{2/\pi}\,K_m(2x\sqrt{-\lambda})I_m(2X\sqrt{-\lambda}) & \text{if } t > 0, \ \lambda < 0, \\ i\sqrt{\pi/2}\,H_m^{(1)}(2x\sqrt{\lambda})J_m(2X\sqrt{\lambda}) & \text{if } t > 0, \ \lambda > 0. \end{cases} \tag{4.11}$$

Here,

$$x = \tfrac{1}{2}(\sqrt{t+r} + \sqrt{t-r}), \quad X = \tfrac{1}{2}(\sqrt{t+r} - \sqrt{t-r}) \quad \text{if } t \pm r > 0,$$

$$x = \tfrac{1}{2}(\sqrt{r+t} + \sqrt{r-t}), \quad X = \tfrac{1}{2}(\sqrt{r+t} - \sqrt{r-t}) \quad \text{if } r \pm t > 0. \tag{4.12}$$

Also $F_\lambda^{(9)}(-t, r) = \exp(-im\pi)\overline{F_\lambda^{(9)}(t, r)}$.

For cases $j = 6, 7, 8$ it is most convenient to use the model H_m, (1.13)–(1.15). The passage to the two variable model proceeds along similar lines as in our earlier work with Lamé bases.[11] In each case the resulting basis function is determined to within a phase. This quantity can be chosen by adopting a fixed normalization of the Lamé Wangerin functions.

8]: $F_\lambda^{(8)}(t, r) = \lambda_m^P[\mathrm{dn}(\alpha, s)\,\mathrm{dn}(\beta, s) + iss'\mathrm{cn}(\beta, s)]^{1/2}$

$$\times F_{m-1/2}^{P}(\alpha, s)F_{m-1/2}^{P}(\beta, s), \tag{4.13}$$

where

$$t = s\ \mathrm{sn}(\alpha, s)\,\mathrm{sn}(\beta, s)/R, \quad r = 1/R \tag{4.14}$$

and

$$R = -i\,\mathrm{dn}(\alpha, s)\,\mathrm{dn}(\beta, s)/ss' + \mathrm{cn}(\alpha, s)\mathrm{cn}(\beta, s')$$

and the variables α, β are in the ranges $\alpha \in [0, 2K]$, $\beta \in [iK', iK' + 2K]$.

7]: $F_\lambda^{(7)}(t, r) = \lambda_m^P[s'\mathrm{sn}(\alpha, s)\mathrm{sn}(\beta, s) + \mathrm{cn}(\alpha, s)\mathrm{cn}(\beta, s)]^{1/2}$

$$\times F_{m-1/2}^{P}(\alpha, s)F_{m-1/2}^{P}(\beta, s), \tag{4.15}$$

where

$$t = \mathrm{dn}(\alpha, s)\mathrm{dn}(\beta, s)/ss'R, \quad r = 1/R,$$

and

$$R = s[\mathrm{sn}(\alpha, s)\mathrm{sn}(\beta, s) + \mathrm{cn}(\alpha, s)\mathrm{cn}(\beta, s)/s']. \tag{4.16}$$

The variables α, β can vary in the two ranges $\alpha, \beta \in [0, 2K]$, $\alpha, \beta \in [iK', iK' + 2K]$.

6]: $$F_\lambda^{(6)}(t, r) = \lambda_m^P[(s - is')\mathrm{dn}(\omega, t)\mathrm{dn}(\mu, r) + (s + is')\mathrm{cn}(\omega, t)\mathrm{cn}(\mu, t)]^{1/2} \times F_{m-1/2}^P(\omega, t)F_{m-1/2}^P(\mu, t), \quad (4.17)$$

where

$$t = 2\sqrt{ss'}(s + is')\mathrm{sn}(\omega, t)\mathrm{sn}(\mu, t)/R,$$
$$r = 2\sqrt{ss'}/R \quad (4.18)$$

and $R = [(s - is')\mathrm{dn}(\omega, t)\mathrm{dn}(\mu, t) + (s + is')\mathrm{cn}(\omega, t)\mathrm{cn}(\mu, t)]$, $t = (s + is')/(s - is')$.

The variables ω and μ vary in the ranges $\omega, \mu \in [-iK', iK']$, which is the line segment joining the points $-iK'$, iK' in the complex plane. (Remember K' and K are complex.) Here for the Lamé bases we have used the same notation as in Sec. 3, where the spectral analysis was performed.

5. OVERLAP FUNCTIONS

Here we compute several of the overlap functions $\langle f_\mu^{(j)}, f_\lambda^{(i)}\rangle$, which allow us to expand eigenfunctions $f_\lambda^{(j)}$ in terms of eigenfunctions $f_\lambda^{(i)}$. Since $\langle T(G)f_\lambda^{(j)}, T(G)f_\lambda^{(i)}\rangle = \langle f_\mu^{(j)}, f_\lambda^{(i)}\rangle$, the same functions allow us to expand eigenfunctions $T(G)f_\mu^{(j)}$ in terms of eigenfunctions $T(G)f_\lambda^{(i)}$. Moreover, since $\langle f_\mu^{(j)}, f_\lambda^{(i)}\rangle = (F_\mu^{(j)}, F_\lambda^{(i)})$, the overlaps allow us to expand eigenfunctions $F_\mu^{(j)}$ in $\mathcal{H}$ in terms of eigenfunctions $F_\lambda^{(i)}$. These last expansions converge in the Hilbert space sense. Pointwise convergence has to be checked separately.

However, if we choose $r \geq 0$ and $\mathrm{Im}t > 0$ in (1.6), then the function $H_{t,r}(k) = \exp(+im\pi/2)\exp(-i\bar{t}k)J_m(kr)$ belongs to $L_2[0, \infty]$ and the transformation $\cup[f]$, (1.6), can be represented as an inner product on $L_2[0, \infty]$:

$$\cup[f] = \langle f, H_{t,r}\rangle. \quad (5.1)$$

In this case it follows easily that all of the expansion formulas

$$\mathcal{J}_\mu^{(j)}(t, r) = \int \langle f_\mu^{(j)}, f_\lambda^{(i)}\rangle \mathcal{J}_\lambda^{(i)}(t, r)\, d\lambda, \quad \mathrm{Im}t > 0, \ r \geq 0, \quad (5.2)$$

are valid in the sense of pointwise convergence, a.e., (See the analogous arguments in Refs. 12 and 13.) In each case it is easy to verify that separation of variables persists in the domain $\mathrm{Im}t > 0$ if it holds for $\mathrm{Im}t = 0$.

Overlaps involving system 2] are especially easy to compute:

$$\langle f_\mu^{(j)}, f_\lambda^{(2)}\rangle = f_\mu^{(j)}(\lambda), \quad 0 < \lambda < \infty. \quad (5.3)$$

In addition we list the overlaps $\langle f_\mu^{(j)}, f_s^{(1)}\rangle$ between the j] basis and the discrete basis 1]:

$$\langle f_\mu^{(3)}, f_s^{(1)}\rangle = \left(\frac{\Gamma(2m + s + 1)}{(s!)\pi}\right)^{1/2} 2^m \frac{\Gamma(2m + i\mu + \frac{1}{2})}{\Gamma(2m + 1)} \times {}_2F_1\left(\begin{matrix} -s, m + i\mu + \frac{1}{2} \\ 2m + 1 \end{matrix}\middle| 2\right), \quad (5.4)$$

$$\langle f_{\alpha,s}^{(4)}, f_s^{(1)}\rangle = 2^{m-\nu+1}(1 + i)^{-\nu-m-1/2} \frac{\Gamma(\nu + m + \frac{1}{2})}{\Gamma(\nu + 1)} \left(\frac{\nu\Gamma(2m + s + 1)}{(s!)\Gamma(2m + 1)}\right)^{1/2} \times F_A\left(\nu + m + \tfrac{1}{2}; \nu + \tfrac{1}{2}, -s; 2\nu + 1, 2m + 1; \frac{2i}{1 + i}, \frac{2}{1 + i}\right), \quad \nu = \alpha + 2q, \ q, s = 0, 1, 2, \cdots. \quad (5.5)$$

Here F_A is a Lauricella function.[14] The overlaps $\langle f_\mu^{(5)}, f_s^{(1)}\rangle$ and $\langle f_\lambda^{(9)}, f_s^{(1)}\rangle$, while straightforward to compute, are of a complexity similar to (5.5) and will not be listed here. [Note that the latter overlaps are not unitary since $\{f_\lambda^{(9)}\}$ is an ON basis for $L_2(R)$, not $L_2[0, \infty]$.]

The overlaps $\langle f_\mu^{(j)}, f_s^{(1)}\rangle$, $j = 6, 7, 8$, can be obtained immediately from the H_m models. The computation of the overlap functions between the Lamé bases 6], 7], and 8] and the basis 1] is easiest to perform by giving the recurrence formulas for these coefficients (see Ref. 11.) We consider explicitly the case of coordinate system 8], where the basis function $\mathcal{J}_\lambda^{(8)}(z)$ is even under the interchange $z \to -z$. Applying the operator $C^2 + k^2B^2$ to both sides of the identity,

$$\mathcal{J}_\lambda^{(8)}(z) = \sum_{n=0}^{\infty} a_n z^{2n}, \quad (5.6)$$

we obtain the recurrence relation

$$k^2(2n + 2)(2n + 1)a_{n+1} + [4n(k^2 + 2)(1 - m - 2s) - 4\lambda - (2m - 1)(2m - 1 + k^2)]a_n + 2k^2[2(n - 1)^2 + (2m - 1)(n - m)]a_{n-1} = 0, \quad (5.7)$$
$$2k^2a_1 - [4\lambda + (2m - 1)(2m - 1 + k^2)]a_0 = 0.$$

The normalized overlap functions b_n are then given via the relation

$$a_n = [\Gamma(2m + n + 1)/\Gamma(2m + 3)n!\,]^{1/2} b_n.$$

For the case of eigenfunctions $\mathcal{J}_\lambda^8(z)$ which are odd the analysis goes through as before by making the substitution $n \to n + \frac{1}{2}$. We should mention here that even and odd eigenfunctions $\mathcal{J}_\lambda^{(8)}(z)$ correspond to Lamé Wangerin functions with an even or odd number of zeros in the interval $(iK'(s), iK'(s) + 2K(s))$ (see Sec. 3). Similar recurrence relations for the basis eigenfunctions of system 7] can be derived by making the substitutions $k^2 \to -1/s^2$, $\lambda \to -\lambda/s^2$. The recurrence relation for 6] is somewhat more lengthy and will not be presented here.

Finally we list the interesting overlaps

$$\langle f_{\alpha,s}^{(4)}, f_\mu^{(5)}\rangle = \frac{\pi}{4}\left(\frac{\nu}{\mu \sinh \mu\pi}\right)^{1/2} \frac{\Gamma((\nu + i\mu)/2)\Gamma((\nu - i\mu)/2)}{\Gamma(\nu + 1)} \times {}_2F_1\left(\begin{matrix} (\nu + i\mu)/2, (\nu - i\mu)/2 \\ \nu + 1 \end{matrix}\middle| -1\right), \quad \nu = \alpha + 2s,$$

$$\langle f_\mu^{(3)}, f_{\alpha,s}^{(4)}\rangle = \sqrt{\nu/\pi}\, \frac{2^{i\mu-1}\Gamma((\nu + i\mu)/2)}{\Gamma((\nu - i\mu)/2 + 1)}, \quad \nu = \alpha + 2s.$$

As discussed in earlier papers in this series, the most general overlaps between basis functions are the

mixed basis matrix elements $\langle T(G)f_\mu^{(j)}, f_\lambda^{(i)}\rangle$. The determination of these matrix elements is straightforward, though frequently the result is complicated.

6. THE COMPLEX EPD EQUATION

In the case where the variables r, t in (1.1) are complex and m is a complex constant, we can regard the EPD equation from another point of view. For the symmetry algebra we now choose the *complex* Lie algebra sl(2, $\mathbb{C}$), whose action on solutions of (1.1) is given by (1.5), where now the matrix elements $\alpha, \beta, \gamma, \delta$ are allowed to be complex and constrained only by the requirement $\det G = 1$.

We can now pose the problem of determining the possible coordinate systems $\{u, v\}$ in which the complex EPD equation is separable. Here, we require that the coordinate transformation functions $u(r, t)$, $v(r, t)$ be only *complex* analytic in r, t rather than real analytic as in the case of the real EPD equation. Furthermore, we regard two coordinate systems as equivalent if one can be obtained from the other by a transformation (1.5) from the group SL(2, $\mathbb{C}$). Just as in Sec. 2, we expect the equivalence classes of coordinate systems to correspond to the SL(2, $\mathbb{C}$)-orbits in the space $\mathcal{J}^c = \mathcal{S}^c/\{C^2 - A^2 - B^2\}^c$, where $\mathcal{S}^c$ is the space of symmetric second-order elements in the universal enveloping algebra of sl(2, $\mathbb{C}$).

To determine the adjoint action of SL(2, $\mathbb{C}$) on $\mathcal{S}^c$, we choose a more convenient basis for sl(2, $\mathbb{C}$):

$$S_1 = iA, \quad S_2 = iB, \quad S_3 = C,$$
$$[S_1, S_2] = S_3, \quad [S_3, S_1] = S_2, \quad [S_2, S_3] = S_1. \tag{6.1}$$

A general element Q of $\mathcal{S}^c$ can be expressed uniquely in the form

$$Q = \sum_{j,k=1}^{3} q_{jk} S_j S_k, \quad q_{jk} = q_{kj} \in \mathbb{C}. \tag{6.2}$$

Using the well-known local isomorphism of SO(3, $\mathbb{C}$) and SL(2, $\mathbb{C}$), and identifying Q with the 3×3 symmetric matrix $\hat{Q} = (q_{jk})$, we see that under the adjoint representation Q transforms according to $\hat{Q} \to O^{-1}\hat{Q}O$, $O \in$ SO(3, $\mathbb{C}$). The elements of $\mathcal{J}^c$ can be identified with the matrices Q such that $\mathrm{tr}\hat{Q} = 0$, or more conveniently, we can add arbitrary multiples of the identity matrix to $\hat{Q}$. It is now a simple exercise in matrix theory to classify the orbits in $\mathcal{J}^c$ under the adjoint representation of SL(2, $\mathbb{C}$). We present only the results and label the four possible orbit types by the eigenvalues of $\hat{Q}$. The symbol $\lambda(2)$ [or $\lambda(3)$] signifies that the eigenvalue λ corresponds to a generalized eigenvector of degree 2 (or 3). Every $\hat{Q} \in \mathcal{J}^c$ is conjugate under the adjoint representation to an element in the following list.

	eigenvalues	*orbit representative*
a.	λ, μ, ρ $\lambda+\mu+\rho=0$, $(\lambda-\mu)(\lambda-\rho)(\mu-\rho)\neq 0$	$\lambda S_1^2 + \mu S_2^2 + \rho S_3^2$
b.	$2\lambda, -\lambda, -\lambda$ $\lambda \neq 0$	$\lambda(2S_1^2 - S_2^2 - S_3^2)$
c.	$\lambda(2), -2\lambda$	$(\lambda+\frac{1}{2})S_1^2 + (\lambda-\frac{1}{2})S_2^2 - 2\lambda S_3^2$ $+\frac{1}{2}i(S_1S_2 + S_2S_1)$
d.	0(3)	$(S_1 + iS_2)S_3 + S_3(S_1 + iS_2)$

(6.3)

For our purposes we can add a scalar multiple of $S_1^2 + S_2^2 + S_3^2$ to any of the orbit representations without changing the element of $\mathcal{J}^c$. Then we find that any element of $\mathcal{J}^c$ is equivalent to a scalar multiple of one of the following elements.

a.	$S_3^2 - k^2 S_2^2$, $k \neq 0$,	6], 7], 8],
b.	S_1^2,	1], 3]
c1.	$(\lambda \neq 0)$ $2S_1^2 + (1-6\lambda)S_3^2 + i(S_1S_2 + S_2S_1)$,	4], 5], (6.4)
c2.	$(\lambda = 0)$ $S_1^2 - S_2^2 + i(S_1S_2 + S_2S_1)$,	2],
d.	$(S_1 + iS_2)S_3 + S_3(S_1 + iS_2)$,	9].

Each of the nine SL(2, R)-orbit representatives in (2.1) belongs to one of the five orbits (6.4), and we have indicated the orbit inclusions in the last column of (6.4). We see that each of our five orbit-types contains at least one of the SL(2, R)-orbits and that some contain more than one. From these facts we infer that there are no new separable coordinate systems obtained by complexifying the EPD equation: all coordinate systems follow from an obvious analytic continuation of the systems 1]–9]. However, the systems 1] and 3], the systems 4] and 5], and the systems 6], 7], 8] are equivalent for the complex EPD equation.

A particularly interesting basis for the solutions of the complex EPD equation is that of type b:

$$\mathcal{J}_n^b(w, \tau) = \tau^{n+m+1}(1-w^2)^{m/2}C_n^{m+1/2}(w),$$
$$B\mathcal{J}_n^b = (n+m+\tfrac{1}{2})\mathcal{J}_n^b. \tag{6.5}$$

Here, $C_n^\nu(w)$ is a Gegenbauer function (a polynomial for $n = 0, 1, 2, \cdots$) and the complex variables w, τ are given by

$$w = t(t^2 - r^2)^{-1/2}, \quad \tau = (t^2 - r^2)^{-1/2}. \tag{6.6}$$

In terms of the variables w, τ the local group action (1.5) of SL(2, C) becomes

$$T(G)\Phi(w, \tau) = \cup^{1/2} V^{1/2} \times \Phi((w + 2\beta\gamma w + \alpha\beta\tau + \gamma\delta\tau^{-1})/V;\ \tau\cup^{-1}),$$
$$\cup = [(\delta^2 + \beta^2\tau^2 + 2\beta\delta\tau w)/(\alpha^2 + \gamma^2\tau^{-2} + 2\alpha\gamma\tau^{-1}w)]^{1/2},$$
$$V = [(2w + 2\beta\gamma w + \alpha\beta\tau + \gamma\delta\tau^{-1})(2\beta\gamma w + \alpha\beta\tau + \gamma\delta\tau^{-1}) + 1]^{1/2}. \tag{6.7}$$

Here

$$G = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

We are now in a position to apply Weisner's method[2–4] to expand solutions of the complex EPD equation in terms of the basis (6.5). Suppose $\Phi(r, t) \equiv \Phi(w, \tau)$ is a solution of the EPD equation such that $\tau^{-m-1}(1-w^2)^{-m/2}\Phi(w, \tau)$ is analytic in τ and w in a neighborhood of $(w, \tau) = (0, 0)$. Then there exist complex constants a_n such that

$$\tau^{-m-1}(1-w^2)^{-m/2}\Phi(w,\tau)=\sum_{n=0}^{\infty} a_n C_n^{m+1/2}(w)\tau^n. \tag{6.8}$$

This method was employed by Viswanathan[2] to derive generating functions for the Gegenbauer polynomials. [The awkward factor $\tau^{-m-1}(1-w^2)^{-m/2}$ which appears in (6.8) is due to our insistence in retaining the EPD equation. One can easily remove this factor by transforming the EPD equation to the equivalent equation for ultraspherical functions which appears in Ref. 2.] In order to derive useful results from (6.8), one characterizes a solution Φ of the EPD equation by requiring that it be an eigenfunction of a first or second order operator in the enveloping algebra of SL(2, C). As Viswanathan remarked, in practice one can compute Φ precisely in the cases where it is possible to find coordinate systems in which variables separate in the equations for Φ. The results of our paper show why this is so and exactly when separable variables exist. Once a suitable Φ is computed one can evaluate the constants a_n by choosing special values of the variables, e.g., $w=0$. Similarly one can derive expansions for $T(G)\Phi$, i.e., functions which lie on the same SL(2, C)-orbit as Φ.

According to (6.4) there are five types of orbits to consider to obtain all possible generating functions for the Gegenbauer polynomials via Weisner's method. An examination of Viswanathan's paper shows that he has found four of these orbits, omitting only the Lamé case (type a). This case can be treated by using the coordinates (4.14) for α, β complex and substituting into (6.6). The remainder of the computations follow just as those given in Ref. 2. However, the resulting identities are somewhat complex due to the fact that $\mathrm{sn}(\alpha, s)$ and $\mathrm{sn}(\beta, s)$ are rather complicated algebraic functions of w and τ.

[1] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. **17**, xxx, xxx, xxx (1976), Papers 8—10.
[2] B. Viswanathan, Can. J. Math. 20, 120 (1968).
[3] L. Weisner, Pacific J. Math. 5, 1033 (1955).
[4] W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[5] W. Miller, Jr., SIAM J. Math. Anal. **4**, 314 (1973).
[6] A. Erdélyi *et al.*, *High Transcendental Functions* (McGraw-Hill, New York, 1953), Vols. I and II.
[7] P. Sally, *Analytic Continuation of the Irreducible Unitary Representations of the Universal Covering Group of SL(2,R)*, AMS Mem., No. 69 (Am. Math. Soc., Providence, R.I., 1967).
[8] W. Montgomery and L. O'Raifeartaigh, J. Math. Phys. 15, 380 (1974).
[9] V. Bargmann, Ann. Math. 48, 568 (1947).
[10] E.C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., Oxford, 1962), Part One, 2nd ed.
[11] E.G. Kalnins and W. Miller, Jr., J. Math. Phys. 15, 1263 (1974).
[12] W. Miller Jr., SIAM J. Math. Anal. 5, 822 (1974).
[13] W. Miller Jr., SIAM J. Math. Anal. 5, 822 (1974).
[14] L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., London, 1966).

On a phase interchange relationship for composite materials

Kalman Schulgasser

Department of Mechanical Engineering, Ben Gurion University of the Negev, Beer Sheva, Israel
(Received 18 July 1975)

A theorem exists relating the transverse conductivity of a fiber reinforced material in a determinate manner to the conductivity of the composite with the phase properties interchanged. It is shown that no such theorem can exist in the three-dimensional case, e.g., for a statistically isotropic composite material. However, an inequality is established relating the two effective conductivities.

1. INTRODUCTION

Keller[1] presented and proved a very interesting "phase interchange" theorem for two-phase fiber reinforced materials. He considered two-phase materials whose boundaries are cylindrical surfaces, say parallel to the z axis, and for which effective conductivity[2] perpendicular to the fiber direction has as principal directions the axes x and y. The conductivity properties of each phase are assumed to be homogeneous and isotropic. Let us denote one phase by the subscript 1, the other by the subscript 2. The conductivities of the individual phases are k_1 and k_2 and the effective conductivities of the composite in the x and y directions are $k_x^*(k_1, k_2)$ and $k_y^*(k_1, k_2)$ respectively. Keller showed that

$$k_x^*(k_1, k_2)k_y^*(k_2, k_1) = k_1k_2, \tag{1.1}$$

where $k_y^*(k_2, k_1)$ is the effective conductivity in the y direction when the phase denoted by 1 is now given conductivity k_2 and the phase denoted by 2 is now given conductivity k_1. We have not changed interface geometry, but only interchanged phase properties. If the sample is statistically isotropic in the x—y plane, then

$$k_x^*(k_1, k_2) = k_y^*(k_1, k_2) = k^*(k_1, k_2)$$

and Eq. (1.1) becomes

$$k^*(k_1, k_2)k^*(k_2, k_1) = k_1k_2. \tag{1.2}$$

Actually, Keller proved this theorem only for rectangular arrays in the x—y plane, when one phase, identified as an inclusion (fiber), has specific symmetries. Mendelson[3] has shown that Eqs. (1.1) and (1.2) are valid for any two-phase material with cylindrical phase boundaries, no matter what the phase geometry, even if the phases are not distinguishable as matrix and inclusion (fiber).

That Keller's theorem can be useful in dealing with the problem of the prediction of effective conductivity of two-phase materials has been shown by Schulgasser.[4] A major limitation on its potential application is its applicability only to the two-dimensional (cylindrical phase boundaries) case. The proofs of both Keller and Mendelson depend on the two-dimensional nature of the problem.

For the three-dimensional case, i.e., a statistically homogeneous two-phase material when x, y and z are the principal directions of the effective conductivity, but when the phase boundaries are not aligned with the z axis, Eq. (1.1) is not valid as can be shown by considering counterexamples to the special case of statistical isotropy expressed by (1.2). Two such counterexamples will be presented in Sec. 3. We will show first of all that not only is Eq. (1.2) not valid in the three-dimensional case but also that no one-to-one relationship exists between $k^*(k_1, k_2)$ and $k^*(k_2, k_1)$, i.e., that knowledge of $k^*(k_1, k_2)$ is not sufficient to determine $k^*(k_2, k_1)$. We will show however that in the three-dimensional case an inequality can be written in place of Eq. (1.1). Specifically we find

$$k_x^*(k_1, k_2)k_y^*(k_2, k_1) \geq k_1k_2, \tag{1.3}$$

or for the statistically isotropic case

$$k^*(k_1, k_2)k^*(k_2, k_1) \geq k_1k_2. \tag{1.4}$$

2. NONEXISTENCE OF A KELLER-TYPE THEOREM IN THREE-DIMENSIONS

To disprove the existence of a Keller-type theorem in three-dimensions it is sufficient to point out one instance for which it cannot possibly be true. Consider a statistically homogeneous and isotropic two-phase material which in addition is statistically symmetric, so that an interchange of the two phase conductivities yields the same material. In general, if a Keller-type theorem exists then

$$k^*(k_2, k_1) = f(k^*(k_1, k_2), k_1, k_2) \tag{2.1}$$

where f is some definite function. But for symmetric materials

$$k^*(k_1, k_2) = k^*(k_2, k_1).$$

Hence for a symmetric material we conclude that

$$k^*(k_1, k_2) = g(k_1, k_2), \tag{2.2}$$

g being some definite function, i.e., $k^*(k_1, k_2)$ is a definite function of k_1 and k_2. Now the detailed phase geometry of a two-phase material is not completely determined by the requirement that the material be symmetric. Indeed, we will present an example of a class of symmetric materials for which different values of k^* are realizable. Hence Eq. (2.2) cannot hold true and no relationship of the sort suggested in Eq. (2.1) can exist.

The class of symmetric materials referred to above is constructed as follows. Let us make a laminate of thin slabs of materials 1 and 2, of equal thickness and stacked alternately. "Grains" of such a material whose dimensions are very large compared to slab thickness will behave as axially symmetric crystals whose conductivity perpendicular to the slabs (i.e., along the

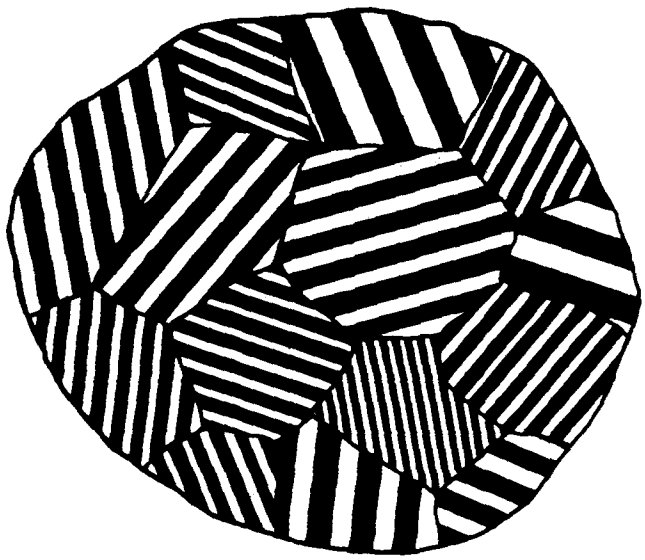


FIG. 1. Model of a statistically symmetric material.

crystal axis) is

$$k_{\parallel} = 2(1/k_1 + 1/k_2)^{-1},$$

and in the directions perpendicular to the crystal axis is

$$k_{\perp} = \tfrac{1}{2}(k_1 + k_2).$$

From these grains we now construct a statistically isotropic polycrystal (see Fig. 1). What has been described above is clearly a statistically symmetric two-phase material. It is known that a polycrystal constituted of axially symmetric crystals does not have a macroscopic conductivity uniquely determined by the principal conductivities of the constituting crystal. It was shown by Molyneux[5] who considered realizable correlation functions that at least a narrow range of effective conductivities is realizable. Schulgasser[6] demonstrated by presenting constructable polycrystal models that quite a large range of effective conductivities is realizable. If effective conductivity of the polycrystal is not uniquely determined by $k_{\parallel}$ and $k_{\perp}$, then k^* for the above described statistically symmetric material is not uniquely determined by k_1 and k_2 and hence no Keller-type principle can exist.

3. A THREE-DIMENSIONAL INEQUALITY

Consider a statistically homogeneous two-phase material for which the principle directions of the effective conductivity are x, y and z. Further, consider the functional U defined by the volume integral,

$$U = (1/2V)\int_V (\mathbf{q}\cdot\mathbf{q}/k)\,dV$$

$$= (1/2V)\int_V (1/k)(q_x q_x + q_y q_y + q_z q_z)\,dV,$$

where q is the heat flux vector and q_x, q_y and q_z are its components. For a statistically homogeneous material subjected to a macroscopic heat flux in the x direction

$$U = \tfrac{1}{2}\langle q_x\rangle^2/k_x^* \tag{3.1}$$

where $\langle\ \rangle$ denotes a space average. An appropriate set of boundary conditions is that on the external surface S of a large block of such a material, $\mathbf{n}\cdot\mathbf{q}$ have the same value as for a uniform flow $q_0 = \langle q_x\rangle$ along x. n is a unit vector normal to the surface. Then the boundary conditions can be written

$$lq_x + mq_y + nq_z = lq_0$$

where l, m, and n are the direction cosines of n. (3.1) serves as a definition of effective conductivity. Consider the classical variational principal

$$U \leq (1/2V)\int_V (1/k)(\tilde{q}_x\tilde{q}_x + \tilde{q}_y\tilde{q}_y + \tilde{q}_z\tilde{q}_z)\,dV, \tag{3.2}$$

where ~ denotes admissible trial heat flux fields. Admissibility requires that the condition

$$\nabla\cdot\tilde{\mathbf{q}} = 0 \tag{3.3}$$

be met and that at interfaces the normal heat flux must be continuous.

Now let us imagine a cut parallel to the x–y plane. Suppose that this cut, rather than being a cut through a statistically isotropic material, were a transverse cut through a fiber reinforced material (perpendicular to fiber direction). At each position on the z axis let us use the true heat flux field that would be obtained for a cut through a fiber reinforced material as the trial field in (3.2). This trial field clearly satisfies the condition (3.3) and the interface condition. We will denote such a trial field by the superscript 2D. For cuts perpendicular to the z axis at all positions on this axis we obtain statistically identical phase geometry and hence statistically identical trial heat flux fields. Noting that

$$q_z^{2D} = 0,$$

we write (3.2) as

$$U \leq (1/2V)\int_V (1/k)(q_x^{2D}q_x^{2D} + q_y^{2D}q_y^{2D})\,dV. \tag{3.4}$$

The trial field satisfies the boundary conditions for the three-dimensional problem since with $q_z = 0$ they can be written

$$[1/(1-n^2)^{1/2}]q_x + [m/(1-n^2)^{1/2}]q_y$$
$$= [1/(1-n^2)^{1/2}]q_0,$$

and the coefficients of q_x, q_y and q_0 are precisely the direction cosines appropriate for the boundary conditions of the two-dimensional case. Now the right hand side of (3.4) is simply

$$(1/2k_x^{*2D})\langle q_x\rangle^2,$$

where k_x^{*2D} denotes the effective conductivity in the x direction of a fiber reinforced material whose cross-section is statistically identical to the cut through the statistically isotropic material described above. Utilizing Eq. (3.1) we then have

$$(1/k_x^*) \leq (1/k_x^{*2D}), \tag{3.5}$$

or equivalently

$$k_x^* \geq k_x^{*2D}. \tag{3.6}$$

We could have looked at the above process from another point of view. If U is minimized under no constraints other than those required physically, and if

(3.1) is then used, the result is the correct $1/k_x^*$. If the minimization is carried out under additional constraints, the resulting value of $1/k_x^*$ can be higher than but not lower than the correct value. The constraint $q_z = 0$ leads to the result $1/k_x^{*2D}$; the inequality (3.5) follows immediately. Now (3.6) will be true if phases 1 and 2 have conductivities k_1 and k_2 respectively, or if the phase conductivities are interchanged throughout. Thus we can write both

$$k_x^*(k_1, k_2) \geq k_x^{*2D}(k_1, k_2) \tag{3.7}$$

and

$$k_x^*(k_2, k_1) \geq k_x^{*2D}(k_2, k_1). \tag{3.8}$$

Inequalities (3.7) and (3.8) are of course valid if the subscript x is replaced by y or z.

Applying Keller's theorem as expressed in Eq. (1.1) to the right hand side of (3.7) we get

$$k_x^*(k_1, k_2) \geq k_1 k_2 / k_y^{*2D}(k_2, k_1),$$

or

$$k_1 k_2 / k_x^*(k_1, k_2) \leq k_y^{*2D}(k_2, k_1). \tag{3.9}$$

Inequality (3.9) together with (3.8) then gives

$$k_1 k_2 / k_x^*(k_1, k_2) \leq k_y^{*2D}(k_2, k_1) \leq k_y^*(k_2, k_1),$$

or simply

$$k_y^*(k_2, k_1) \geq \frac{k_1 k_2}{k_x^*(k_1, k_2)}. \tag{3.10}$$

This is the sought after inequality (1.3), which for the case $k_x^* = k_y^*$ reduces to the inequality (1.4).

Before presenting examples for the statistically isotropic case which disprove the validity of (1.2) for the statistically isotropic case, but which do satisfy the inequality (1.4), several comments are in order. First of all the intermediate inequality (3.6) is valid not only for a two-phase material but for a material of any number of phases. Indeed it is true for a material with continuously varying properties, as long as the definition of k_x^{*2D} is suitably broadened. Secondly, one might expect that a second inequality, perhaps bounding $k_y^*(k_2, k_1)$ from above, might be obtainable from the classical variational principal complementary to (3.2). That this is not so is shown in the Appendix.

Beran and Molyneux[7] have calculated bounds on the quantity

$$Q = k^{*2}/k_1 k_2$$

for a symmetric two phase material. According to the inequality (1.4) when $k^*(k_1, k_2)$ is taken equal to $k^*(k_2, k_1)$ because of the statistical symmetry of the material, Q may be greater than 1, and need not necessarily be equal to 1 as implied by (1.2). Beran and Molyneux calculate upper and lower bounds on Q, but cannot show that they are realizable. For k_2/k_1 less than about 14 both bounds are greater than 1. Hence (1.2) cannot hold and (1.4) holds. For k_2/k_1 greater than 14 their lower bound is less than 1. Inequality (1.4) indicates that we can simply set the lower bound at 1.

As a second example consider the "spheres assemblage" model proposed by Hashin and Shtrikman[8] for which the exact conductivity can be calculated. For one alternative of their model (the other gives the same result with indices reversed)

$$k^*(k_1, k_2) = k_2 + 3v_1 k_2 (k_1 - k_2)/[3k_2 + v_2(k_1 - k_2)],$$

where v_1 and v_2 are the volume fractions of the phases. Then

$$k^*(k_2, k_1) = k_1 + 3v_1 k_1 (k_2 - k_1)/[3k_1 + v_2(k_2 - k_1)].$$

Using these expressions we find

$$\frac{k^*(k_1, k_2)k^*(k_2, k_1)}{k_1 k_2} = 1 + \frac{9v_1 v_2 (k_2 - k_1)^2}{(9 - 6v_2 + 2v_2^2)k_1 k_2 + (3v_2 - v_2^2)(k_1^2 + k_2^2)}$$

which is clearly always greater than 1 except for either $v_2 = 1$ or $v_1 = 1$ when it is exactly 1, as we would expect for this case of degeneracy to a one-phase material.

As a final interesting example of the inequality (1.3) and to emphasize its applicability to any statistically homogeneous material no matter what the nature of its anisotropy let us consider an extreme case. Take the inequality in the form

$$k_x^*(k_1, k_2) k_z^*(k_2, k_1) \geq k_1 k_2,$$

where the material we are considering has phase boundaries which are cylindrical surfaces parallel to the z axis. Then we know

$$k_z^*(k_1, k_2) = v_1 k_1 + v_2 k_2,$$

and by applying the inequality we find

$$k_x^*(k_1, k_2) \geq (v_1/k_1 + v_2/k_2)^{-1}.$$

This is the well known Wiener lower bound on the effective conductivity of two phase composite materials. k_x^* realizes the right hand side when the phase boundaries are planes perpendicular to the x axis, as it must, since for this case the problem is again two-dimensional.

APPENDIX

Consider a statistically homogeneous two-phase material for which the principal directions of the effective conductivity are x, y, and z. Further consider the functional U defined by

$$U = (1/2V) \int_V k \nabla T \cdot \nabla T \, dV$$

$$= (1/2V) \int_V k[(\partial T/\partial x)^2 + (\partial T/\partial y)^2 + (\partial T/\partial z)^2] dV,$$

where T is the temperature field in the composite material. For a statistically homogeneous material subjected to a temperature gradient in the x direction,

$$U = \tfrac{1}{2} k_x^* \langle \partial T/\partial x \rangle^2. \tag{A1}$$

(A1) serves as a definition of effective conductivity. Consider the classical variational principle

$$U \leq (1/2V) \int_V k[(\partial \tilde{T}/\partial x)^2 + (\partial \tilde{T}/\partial y)^2 + (\partial \tilde{T}/\partial z)^2] dV \tag{A2}$$

where ~ denotes admissible trial temperature gradients.

To be admissible it is required that T be continuous throughout the material.

Now let us imagine a cut in the x-y plane. Suppose this cut, rather than being a cut through a statistically isotropic material were a transverse cut through a fiber reinforced material (perpendicular to the fiber direction). Let us use the true temperature field that would be obtained for the cut through a fiber reinforced material as the field from which we derive the admissible temperature gradients in (A2). We will denote such a trial field by the superscript 2D. For cuts perpendicular to the z axis at all positions on this axis we obtain statistically identical phase geometry, and hence statistically identical trial temperature gradient fields. We note that

$$(\partial T/\partial z)^{2D} \neq 0$$

since for a given x—y coordinate position the temperature changes as we move in the z direction, by virtue of the fact that the phase geometry of two neighboring cuts changes. Then from (A1) and (A2) we can write only

$$\tfrac{1}{2}k_x^*\langle \partial T/\partial x\rangle^2$$

$$\leq \tfrac{1}{2}k_x^{*2D}\langle \partial T/\partial x\rangle^2 + (1/2V)\int_V [(\partial T/\partial z)^{2D}]^2\, dV.$$

No way of evaluating this last integral is apparent. Additionally it is possible to conceive of phase geometries for which the above proposed trial field is not even admissible, e.g., instances when phase boundaries are (locally) planes perpendicular to the z axis. For the case of continuously varying material properties the proposed trial field is clearly admissible.

[1]J. Keller, J. Math. Phys. 5, 548 (1964).
[2]We will use throughout the language of thermal conductivity, but by mathematical analogy the results to be obtained are also valid for electrical conductivity, electrical permittivity, and magnetic permeability.
[3]K. Mendelson, J. Appl. Phys. 46, 917 (1975).
[4]K. Schulgasser, J. Math. Phys. 17, 382 (1976).
[5]J. Molyneux, J. Math. Phys. 11, 1172 (1970).
[6]K. Schulgasser, "Relationship Between Single Crystal and Polycrystal Electrical Conductivity," Technical Report MED 16/75, Department of Mechanical Engineering, Ben Gurion University of the Negev, Israel, April 1975 (to be submitted).
[7]M. Beran and J. Molyneux, Q. Appl. Math. 24, 107 (1966).
[8]Z. Hashin and S. Shtrikman, J. Appl. Phys. 33, 3125 (1962).

On the conductivity of fiber reinforced materials

Kalman Schulgasser

Department of Mechanical Engineering, Ben Gurion University of the Negev, Beer Sheva, Israel
(Received 18 July 1975)

A two-phase material in which the phase boundaries are cylindrical surfaces is considered. A technique exists for finding upper and lower bounds on the effective thermal conductivity (or electrical conductivity, permittivity, or magnetic permeability) of the composite in the direction perpendicular to the generators of the phase boundaries in terms of two different three point correlation functions. It is shown how a phase interchange theorem can be introduced into these bounds enabling us to express them in terms of a single geometrical constant of phase geometry. We determine what range of values of this factor is realizable for real phase geometries, and we show that the bounds thus obtained span exactly all realizable effective conductivities for such composites. Finally, we show that the bounds as expressed here enable us to use a knowledge of the effective conductivity of a composite for one ratio of constituent conductivities to narrow the bounds for some other ratio.

1. INTRODUCTION

We consider here two-phase fiber reinforced materials. By this we mean any two-phase material whose boundaries are cylindrical surfaces, say parallel to the x_3 axis. No other restriction is made on the geometry of phase boundaries. Indeed, we do not even require that any one phase be identifiable as the matrix, the other being the inclusion (fiber). We are concerned with the problem of determining the effective thermal conductivity perpendicular to the fiber direction (i.e., perpendicular to the generators of the cylindrical phase boundaries) when the statistical properties of the phase geometry for cuts perpendicular to the fiber geometry are isotropic and homogeneous. The conductivity properties of each phase are assumed to be isotropic and homogeneous. Hashin[1,2] derived bounds for this effective conductivity in terms of the phase conductivities and volume fractions of the two phases. He was able to show that these are the best possible bounds obtainable if this is the only information available defining phase geometry. He showed this by presenting models of two-phase materials for which the effective conductivity could be exactly calculated and which exactly realized his upper and lower bound. For two-phase materials completely isotropic (statistically) in *three dimensions* Beran[3] has developed a statistical theory for bounding effective conductivity in terms of additional statistical information, viz. three-point correlation functions. Silnutzer[4] has applied Beran's technique to the problem under consideration here. It was only necessary to rewrite all of the results of Beran using two rather than three dimensions. Silnutzer found that k^*, the effective conductivity, is bounded as follows:

$$[\langle 1/k\rangle - \langle k'/k\rangle^2/4J\langle k\rangle^2]^{-1} \leq k^* \leq \langle k\rangle - \tfrac{1}{2}\left(\langle k'^2\rangle/\langle k\rangle\right)\Big/(1 + 2I\langle k\rangle/\langle k'^2\rangle), \tag{1.1}$$

where $\langle\ \rangle$ denotes an ensemble average which is assumed to be equal to the spatial average for a statistically homogeneous material, and $'$ denotes deviations from $\langle k\rangle$, i.e., if the two phases are labelled 1 and 2 then

$$k_1' = k_1 - \langle k\rangle \quad \text{and} \quad k_2' = k_2 - \langle k\rangle.$$

I and J are integrals of different three-point correlation functions and are given as follows:

$$I = \frac{1}{4\pi^2\langle k\rangle^2}\int_A\int_{A'}\frac{\partial^2}{\partial r_1 \partial s_1}\langle k'(0)k'(\mathbf{r})k'(\mathbf{s})\rangle\frac{r_i s_i}{r^2 s^2}\,d\mathbf{r}\,d\mathbf{s}, \tag{1.2}$$

$$J = \frac{1}{4\pi^2\langle k\rangle^2}\int_A\int_{A'}\frac{\partial^2}{\partial r_1 \partial s_1}\left\langle\frac{k'(\mathbf{r})k'(\mathbf{s})}{k(0)}\right\rangle\frac{r_i s_i}{r^2 s^2}\,d\mathbf{r}\,d\mathbf{s}. \tag{1.3}$$

$\mathbf{r}$ and $\mathbf{s}$ are vectors in the plane of the cut with components r_1, r_2 and s_1, s_2 respectively. The repeated index indicates summation. The integrations are over areas infinitely large compared to the scale of the phase geometry. Inequality (1.1) and the definitions (1.2) and (1.3) are the two-dimensional analog of Beran's three-dimensional results and are very nearly identical in form.

The usefulness of (1.1) and its three-dimensional counterpart depends on our ability to determine I and J. Little success has been achieved in this respect. In fact, short of actual measurement or calculation for a particular structure it has not been known what values of I and J are possible in real materials. For a certain class of cellular statistically isotropic three-dimensional materials Miller[5] has evaluated I and J in terms of cell shape. The analogous results for the two-dimensional case under consideration here have been presented by Beran and Silnutzer.[6] Both the extent of the applicability and the universality of these evaluations of I and J are not clear. Firstly, it is necessary to determine to what extent real materials fulfill the restriction on phase geometry imposed in these works, especially that of complete independence of conductivity properties between any two cells, and secondly because the I's and J's computed do not produce bounds which span the complete range of the Hashin bounds, a range which has been shown to be realizable.

A disturbing feature of the Silnutzer (and Beran) bounds is that the upper and lower bounds depend on different three-point correlation functions. We will show here, without putting any restrictions on phase geometry, that for the case of a fiber reinforced material both upper and lower bounds can be found in terms of *either* I or J. We will see that I can be written as

$$I = (k_2'^3/\langle k\rangle^2 v_1^3)I_1, \tag{1.4}$$

and that J can be written as

$$J=(k_2'^2/k_2\langle k\rangle^2)[J_1+(k_2/k_1)J_2], \tag{1.5}$$

where I_1, J_1 and J_2 are purely geometrical functions of the statistics of phase geometry. v_1 and v_2 are the volume fractions of the phases ($v_1+v_2=1$). Turning our attention particularly to I_1 we will find the limits on the values that I_1 can take for any given volume fractions of the phases and we will show that these limits are realizable, i.e., that these limiting values correspond to real materials. We will see that the Hashin bounds can be derived from those of Silnutzer. Further we will show that if for a material with a particular structure the actual k^* can be exactly (or approximately) calculated then the exact (or approximate) value of the corresponding factor I can be found. We will see that a knowledge of k^*, say from measurements, for a particular ratio k_2/k_1 permits us to bound I and that these bounds on I can in turn be used to bound k^* for different ratios k_2/k_1, the phase structure remaining unchanged. This last result is significant in view of the complete mathematical analogy between the problems of determining effective thermal conductivity, effective electrical conductivity, effective electrical permitivity and effective magnetic permeability of composite materials. That the analogy can in fact be extended to include the computation of effective elastic axial shear modulus of fiber reinforced materials was indicated by Springer and Tsai[7] for the case of rectangular periodic arrays, and it was noted that this analogy is generally valid by Hashin[2] who pointed out that the underlying mathematics is the same in both cases. Thus for example, a knowledge of the thermal conductivity of a composite permits us to place rigorous bounds on its electrical permittivity, assuming of course, that the phase conductivities and permittivities are known.

2. THE BOUNDS AND THEIR IMPLICATIONS

Consider the correlation function

$$P(\mathbf{r},\mathbf{s})=\langle k'(0)k'(\mathbf{r})k'(\mathbf{s})\rangle$$

which appears in the integrand of the integral defining I. Were we to perform experiments to determine $P(\mathbf{r},\mathbf{s})$ by randomly sampling for any given $\mathbf{r}$ and $\mathbf{s}$ we would be faced with eight possible outcomes. These are:

(1) $k_1'\cdot k_1'\cdot k_1'=k_2'^3(k_1'/k_2')^3$,

(2) $k_1'\cdot k_1'\cdot k_2'=k_2'^3(k_1'/k_2')^2$,

(3) $k_1'\circ k_2'\cdot k_1'=k_2'^3(k_1'/k_2')^2$,

(4) $k_2'\cdot k_1'\circ k_1'=k_2'^3(k_1'/k_2')^2$,

(5) $k_2'\circ k_2'\cdot k_1'=k_2'^3(k_1'/k_2')$,

(6) $k_2'\cdot k_1'\cdot k_2'=k_2'^3(k_1'/k_2')$,

(7) $k_1'\circ k_2'\cdot k_2'=k_2'^3(k_1'/k_2')$,

(8) $k_2'\cdot k_2'\cdot k_2'=k_2'^3(1)$.

In the second column $k_2'^3$ has been extracted. But using the definition of k_1' and k_2'

$$\frac{k_1'}{k_2'}=\frac{k_1-(v_1k_1+v_2k_2)}{k_2-(v_1k_1+v_2k_2)}=-\frac{v_2}{v_1},$$

i.e., each possible outcome of such an experiment yields the factor $k_2'^3$ times a quantity completely independent of phase conductivities, but rather dependent only on phase geometry. Then $P(\mathbf{r},\mathbf{s})$ can be written

$$P(\mathbf{r},\mathbf{s})=k_2'^3G(\mathbf{r},\mathbf{s}),$$

G depending purely on geometry. Putting this into the definition of I we obtain (1.4), viz.

$$I=(k_2'^3/\langle k\rangle^2v_1^3)I_1$$

where I_1 is a purely geometric factor for any particular phase geometry. The factor $1/v_1^3$ has been removed from I_1 for later convenience.

The possible outcomes for the same experiment performed to determine the correlation function in the integral of J are:

(1) $(k_1'\cdot k_1')/k_2=(k_2'^2/k_2)(k_1'/k_2')^2$,

(2) $(k_1'\cdot k_2')/k_2=(k_2'^2/k_2)(k_1'/k_2')$,

(3) $(k_2'\circ k_1')/k_2=(k_2'^2/k_2)(k_1'/k_2')$,

(4) $(k_2'\circ k_2')/k_2=(k_2'^2/k_2)(1)$,

(5) $(k_1'\cdot k_1')/k_1=(k_2'^2/k_2)(k_1'/k_2')^2(k_2/k_1)$,

(6) $(k_1'\cdot k_2')/k_1=(k_2'^2/k_2)(k_1'/k_2')(k_2/k_1)$,

(7) $(k_2'\circ k_1')/k_1=(k_2'^2/k_2)(k_1'/k_2')(k_2/k_1)$,

(8) $(k_2'\circ k_2')/k_1=(k_2'^2/k_2)(1)(k_2/k_1)$.

We see that all outcomes contain either of the factors $k_2'^2/k_2$ or $(k_2'^2/k_2)(k_2/k_1)$ times a quantity independent of phase conductivity. Hence it is clear that J can be written as in (1.5), viz.

$$J=(k_2'^2/k_2\langle k\rangle^2)[J_1+(k_2/k_1)J_2].$$

It is significant that two separate factors are necessary to determine J for a given geometry for any ratio k_2/k_1.

We will utilize a phase interchange theorem first proved by Keller[8] for ordered arrays of cylinders with certain symmetries embedded in a matrix, and which was later proved by Mendelson[9] to be valid for all two-phase materials whose phase boundaries are cylindrical surfaces, there being no furthere restriction on phase geometry. Let the conductivity of phase 1 be k_1 and of phase 2 be k_2. Then $k^*(k_1,k_2)$ is the effective conductivity perpendicular to the fiber direction. Keller showed that

$$k^*(k_1,k_2)k^*(k_2,k_1)=k_1k_2, \tag{2.1}$$

where $k^*(k_2,k_1)$ is the effective conductivity when the phase denoted by 1 is now given conductivity k_2 and the phase denoted by 2 is now given conductivity k_1. We have not changed interface geometry but only interchanged phase properties. (Flaherty and Keller[10] have also proved this theorem separately for the case of axial shear modulus.)

Now we note that the upper bound (right hand side) in (1.1) is a function of k_1, k_2, v_1, v_2 and I_1. Let us denote this function by $f_U(k_1,k_2,v_1,I_1)$. Then

$$k^*(k_1,k_2)\le f_U(k_1,k_2,v_1,I_1). \tag{2.2}$$

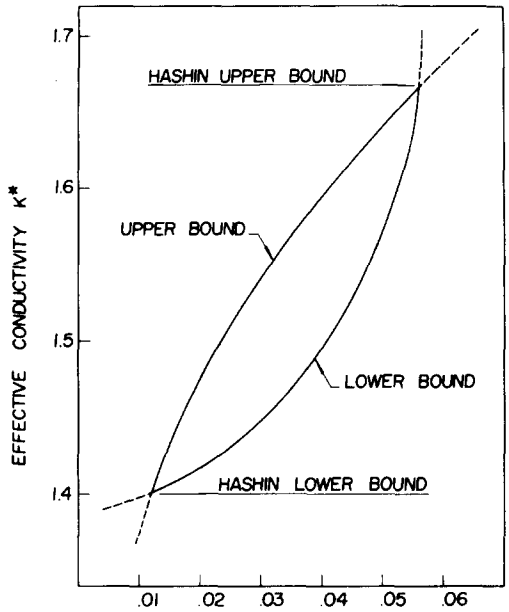


FIG. 1. Bounds on the effective conductivity k^* as a function of the geometrical factor I_1. $k_1=1$, $k_2=5$; $v_1=\frac{3}{4}$, $v_2=\frac{1}{4}$.

Similarly, for the same phase geometry, were the conductivity of 1 actually k_2 and that of 2, k_1 then

$$k^*(k_2, k_1) \leq f_U(k_2, k_1, v_1, I_1). \tag{2.3}$$

But from (2.1) we can write this last inequality as

$$k_1k_2/k^*(k_1, k_2) \leq f_U(k_2, k_1, v_1, I_1),$$

or

$$k^*(k_1, k_2) \geq k_1k_2/f_U(k_2, k_1, v_1, I_1). \tag{2.4}$$

(2.2) and (2.4) together give a set of bounds in terms of a single geometric factor I_1. Writing these out in detail we have

$$k_1k_2\left((k_1v_2+k_2v_1)-\frac{(v_1v_2/2)^2(k_2-k_1)^2}{(v_1v_2/2)(k_1v_2+k_2v_1)+(k_1-k_2)I_1}\right)^{-1}$$
$$\leq k^* \leq (k_1v_1+k_2v_2)-\frac{(v_1v_2/2)^2(k_1-k_2)^2}{(v_1v_2/2)(k_1v_1+k_2v_2)+(k_2-k_1)I_1}. \tag{2.5}$$

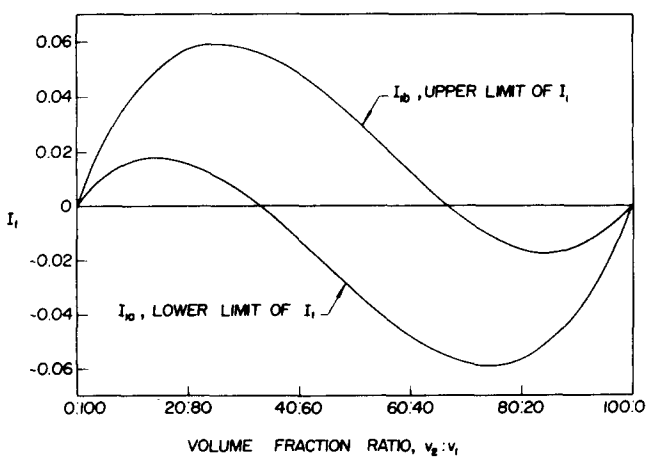


FIG. 2. Permissible values of the geometric factor I_1.

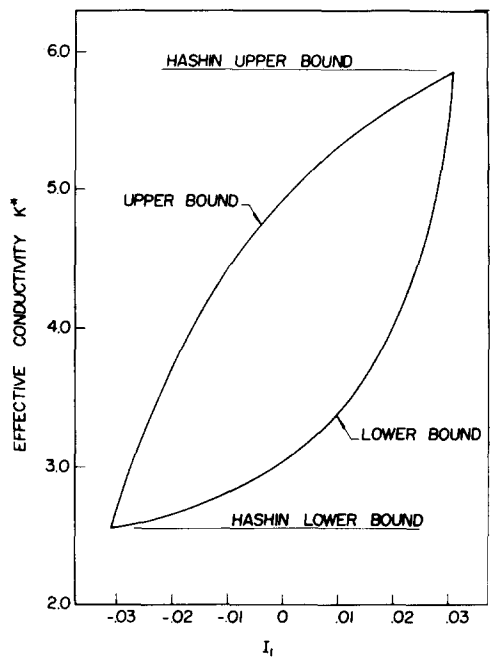


FIG. 3. Bounds on the effective conductivity k^* as a function of the geometrical factor I_1. $k_1=1$, $k_2=15$; $v_1=v_2=\frac{1}{2}$.

To obtain a visual image of the implications of these bounds we have in Fig. 1 plotted the bounds against I_1 for the case $v_1=\frac{3}{4}$, $v_2=\frac{1}{4}$ for $k_1=1$ and $k_2=5$. The maximum and minimum possible values of I_1 are at the intersections of the bounds, and the values of k^* at these crossing points are absolute bounds on the effective conductivity. To determine the crossing points we set

$$f_U(k_1, k_2, v_1, I_1) = k_1k_2/f_U(k_2, k_1, v_1, I_1)$$

and solve for I_1. We find that the values thus determined are independent of k_1 and k_2 and are given by

$$I_{1_a} = (v_1v_2/2)(v_1/2 - v_2) \tag{2.6}$$

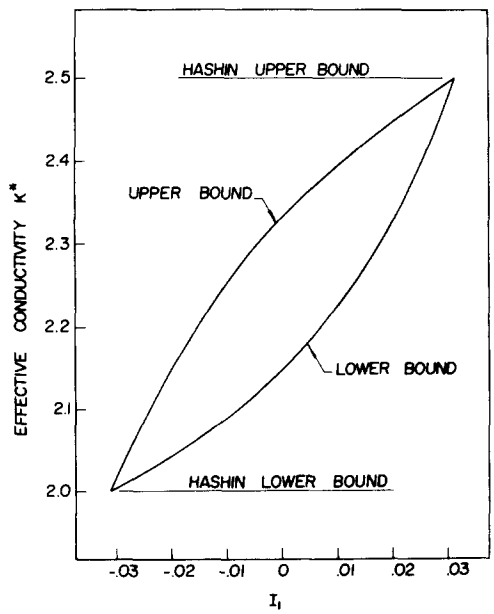


FIG. 4. Bounds on the effective conducitivity k^* as a function of the geometrical factor I_1. $k_1=1$, $k_2=5$; $v_1=v_2=\frac{1}{2}$.

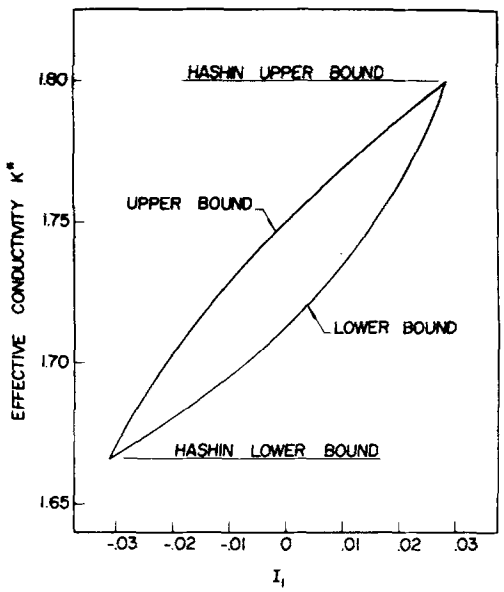


FIG. 5. Bounds on the effective conductivity k^* as a function of the geometrical factor I_1. $k_1=1$, $k_2=3$; $v_1=v_2=\frac{1}{2}$.

and

$$I_{1_b}=(v_1v_2/2)(v_1-v_2/2). \tag{2.7}$$

Further we find that the values of k^* corresponding to these are

$$k_a^*=k_1+v_2/\left(\frac{1}{k_2-k_1}+\frac{v_1}{2k_1}\right) \tag{2.8}$$

and

$$k_b^*=k_2+v_1/\left(\frac{1}{k_1-k_2}+\frac{v_2}{2k_2}\right) \tag{2.9}$$

respectively. These are *precisely* the Hashin bounds for effective conductivity which Hashin has shown are realizable. This was first shown by Hashin in the axial shear modulus case[1] and was later pointed out by him to be valid for the conductivity case.[2] It turns out that for $k_2>k_1$, k_a^* is the lower bound and k_b^* the upper; for $k_2<k_1$, k_a^* is the upper bound and k_b^* the lower. Hence I_{1a} and I_{1b} are not only bounds on I_1 but are also realizable. In Fig. 2 we show the permitted range of I_1 for all volume fraction ratios.

Another feature of the bounds can be appreciated by considering Figs. 3—6 successively. Here we have taken $v_1=v_2=\frac{1}{2}$ and consider ratios k_2/k_1 equal to 15, 5, 3 and 1.5. We see that the bounds narrow progressively at a rate much faster than that at which k_a^* approaches k_b^*. In fact using l'Hospital's rule we easily show that

$$\lim_{k_2\to k_1}\frac{f_U(k_1,k_2,v_1,I_1)-k_1k_2/f_U(k_2,k_1,v_1,I_1)}{k_a^*-k_b^*}=0$$

for all values of v_1 and I_1.

Referring to any of the figures we see how a knowledge of k^* (say from measurement) for particular values of k_1 and k_2 tells us the possible range of I_1 for a material of the structure for which this k^* is known, e.g., from Fig. 3 we see that if for a fifty-fifty mixture of materials with $k_1=1$ and $k_2=15$ the effective conductivity is 4, then I_1 must be between $-.0167$ and $.0199$. This range of I_1 represents nearly 60% of the total range of I_1. For $k_1=1$ and $k_2=1.5$ with $k^*=1.225$ we find that I_1 must lie between 0 and .00625. This range of I_1 represents slightly less than 10% of the total range of I_1 were nothing known about k^*, and hence we see how a rather precise determination of I_1 is possible if k^* can be determined for ratios of k_2/k_1 close to 1. Now for ratios of k_2/k_1 close to 1 the entire range between k_a^* and k_b^* is small, e.g., for $k_2/k_1=1.5$ we find that $(k_a^*-k_b^*)/k_a^*$ is less than 0.5%, so extremely accurate measurement techniques would be necessary to determine I_1 precisely. However, if we can exactly (or approximately) compute k^* for a given structure we could exactly (or approximately) compute I_1 by considering the case of $k_2\sim k_1$.

Prager[11] first posed and presented a solution to the problem of determining bounds on k^* for one set of phase conductivities when k^* is known for some other set of phase conductivities. His results, obtained from classical variational principles (his Eqs. 27 and 28), are written for a statistically isotropic three-dimensional two-phase material but the development is valid line by line also for the case under consideration here and the final results are unchanged. We see now how Prager's problem for the fiber reinforced case can be handled using the present bounds. The known value of k^* for one set of k_2, k_1 is used to find limits on I_1. These limits on I_1 are then used to find bounds on k^* for a different set of k_2, k_1. This technique will always produce bounds better than those of Hashin. In some instances Prager's solution gives narrower bounds, in some instances those obtained from the present technique are narrower. We give two illustrations. Consider a fifty-fifty mixture for which it is known that $k^*=2.2$ when $k_1=1$ and $k_2=5$. Then for $k_1=1$ and $k_2=3$ the various

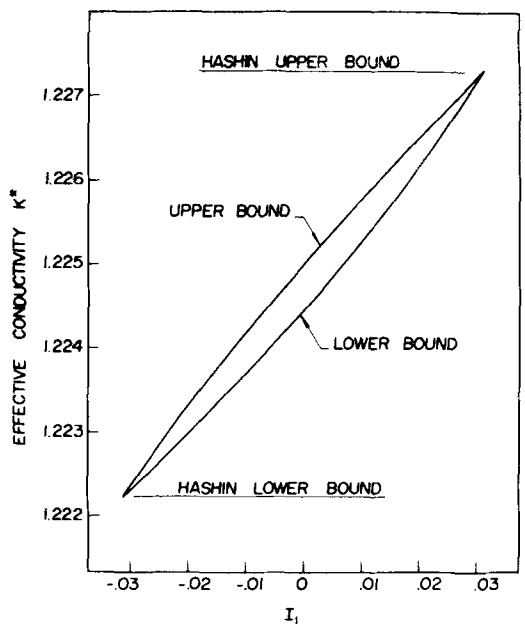


FIG. 6. Bounds on the effective conductivity k^* as a function of the geometrical factor I_1. $k_1=1$, $k_2=1.5$; $v_1=v_2=\frac{1}{2}$.

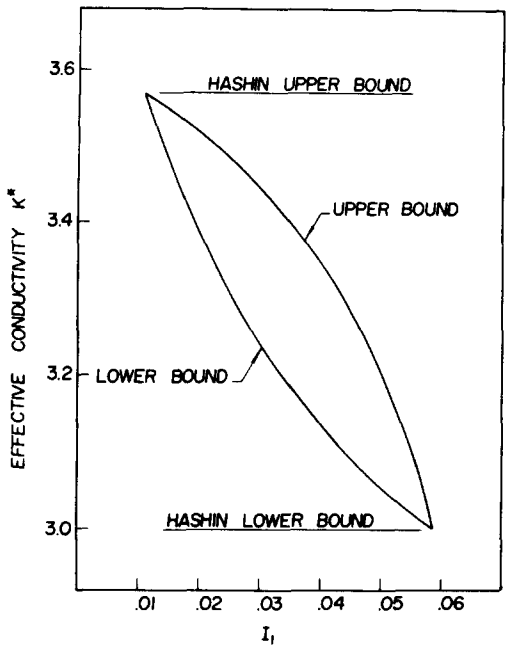


FIG. 7. Bounds on the effective conductivity k^* as a function of the geometrical factor I_1. $k_1=5$, $k_2=1$; $v_1=\frac{3}{4}$, $v_2=\frac{1}{4}$.

bounds obtained are as follows:

Hashin: $1.6667 \leq k^* \leq 1.8000$,

Prager: $1.6875 \leq k^* \leq 1.7500$,

Present work: $1.6875 \leq k^* \leq 1.7647$.

The Prager bounds are slightly better than those obtained by the present method. Consider now a fifty-fifty mixture for which it is known that $k^*=1.225$ when $k_1=1$ and $k_2=1.5$. Then for $k_1=1$ and $k_2=5$ the various bounds obtained are as follows:

Hashin: $2.000 \leq k^* \leq 2.500$,

Prager: $2.059 \leq k^* \leq 2.442$,

Present work: $2.143 \leq k^* \leq 2.375$.

The bounds obtained by the present method are somewhat better than those obtained from Prager's results.

When the effective conductivity is known for some ratio k_2/k_1 greater than 1 and sought for a different ratio k_2/k_1 less than 1 the Prager equations give bounds worse than those of Hashin. However, before the Prager equations are applied the Keller theorem can be used to invert the ratio for which effective conductivity is known, so a fair comparison is as above, i.e., both ratios k_2/k_1, either greater than 1 (or both less than 1). An interchange of k_2 and k_1 in the bounds found here produces a visually different set of bounds, but no new information. (Compare Fig. 7 with Fig. 1). This is to be expected since it was the Keller theorem which relates $k^*(k_1, k_2)$ to $k^*(k_2, k_1)$ that permitted the construction of the lower bounds in the first place.

Returning now to the Silnutzer bounds (1.1) we see that by applying the present approach to the left hand side we can obtain upper and lower bounds in terms of J. But since J is not expressible in terms of a single geometric factor, but rather in terms of two, J_1 and J_2, these bounds are of limited usefulness. Knowledge of k^* does not offer bounds on either J_1 or J_2 separately.

3. CLOSURE

Only the lower bounds that have been developed here in terms of I_1 are new; the upper bounds are those of Silnutzer rewritten in terms of the purely geometrical part of I. The attainment of this lower bound adds considerably to the usefulness of the Silnutzer result. Consider Fig. 1. Were only the upper bound known, then we would only be able to place a lower limit on I_1 (where the bound takes on the value of the Hashin lower bound), which we could not even say was realizable since it was not known whether the upper bound is the best possible in terms of I_1. We could only say that the upper limit on I_1 is at least as high as the crossing point of the upper bound with the Hashin upper bound. The addition of the lower bound and the fact of its crossing at the Hashin bounds (which are realizable) permits us to place precise limits on I_1 which are then realizable.

We do not know whether the bounds presented above are the best possible in terms of I_1, but it is not unlikely that they are because of the confluence of these bounds and those of Hashin at the extreme permissible values of I_1. In order to determine if indeed these bounds are best it would be necessary, given k_1, k_2 and v_1, to construct a set of models of two-phase materials for which k^* can be exactly calculated and all of which have the same value of I_1, but different values of k^*—and then to determine if the range between the bounds for each value of I_1 is spanned. The author is currently constructing such a set of models.

Another unanswered question is the relationship between I and J. If the bounds in terms of I are not best, does J contain information that permits narrowing the bounds for a specific structure? One indication that this is probably not so is as follows: For a completely symmetric material, i.e., one for which an interchange of k_1 and k_2 produces the same material ($v_1=v_2=\frac{1}{2}$ of course) we see from Keller's theorem (2.1) that

$$k^*=\sqrt{k_1k_2}. \tag{3.1}$$

(This was pointed out by Mendelson.[9]) In terms of the correlation function $P(\mathbf{r}, \mathbf{s})$ in the integrand of (1.2), symmetry implies that $P(\mathbf{r}, \mathbf{s})=0$ and hence $I_1=0$. Then the bounds (2.5) reduce to

$$[k_1k_2/\tfrac{1}{2}(k_1+k_2)][1-\tfrac{1}{2}(k_1-k_2)/(k_1+k_2)^2]^{-1}$$

$$\leq k^* \leq \tfrac{1}{2}(k_1+k_2)[1-\tfrac{1}{2}(k_1-k_2)/(k_1+k_2)^2]. \tag{3.2}$$

These are the same bounds as are obtained from the results of Beran and Silnutzer[6] when both of their cell shapes are taken to be the same, but *their* lower bound is obtained by placing the appropriate restrictions on the form of J, and not on I.

This leads to the more general question of the efficiency of three-point correlation functions in characterizing two-phase materials as to their effective conductivity. We have seen that the simple statement that a fiber reinforced material is symmetric completely de-

termines its effective conductivity. But bounds obtained through the three-point correlation functions are often quite wide (Fig. 3 with $I_1 = 0$). Symmetry implies $I_1 = 0$; $I_1 = 0$ does not necessarily imply symmetry. Hence the contribution of the higher order correlation functions must be very significant.

Prager's method characterizes a two-phase material by its effective conductivity for any given constituent conductivities. The present work can be interpreted as doing the same thing since, using the technique described at the end of the last section, we could have *explicity* written

$$G_L(k_1, k_2, v_1, k_1^\circ, k_2^\circ, k^{\circ *}) \leq k^* \leq G_U(k_1, k_2, v_1, k_1^\circ, k_2^\circ, k^{\circ *})$$

where the superscript $^\circ$ refers to the constituent conductivities for which effective conductivity is known. G_L and G_U are the appropriate functions. I_1 will not appear at all in the final result. As has been shown, neither Prager's method nor the present method consistently gives better results (narrower bounds) for all ratios k_1/k_2 and k_1°/k_2°. Hence it is to be expected that a better solution to the problem posed by Prager exists than either his solution or that obtained by the present method.

Hori and Yonezawa[12] have recently rederived the bounds (1.1) upon which the present work is based. In place of the nondimensional geometrical factor I_1 which we used in writing the right hand side of Silnutzer's bounds, they introduce a coefficient $A^{(3)}$ related to I_1 as follows

$$I_1 = A^{(3)} v_1 v_2 (v_1 - v_2).$$

They also show that a perturbation expansion for the case $k_2 \sim k_1$ gives

$$k^* = \langle k \rangle (1 - \tfrac{1}{2}\langle k'^2 \rangle / \langle k \rangle^2). \tag{3.3}$$

This is simply the right hand side of (1.1) with I (or I_1) taken as 0. However, referring to Fig. 2 we see that $I_1 = 0$ (and hence $I = 0$) is not possible for all volume fractions; specifically it is not possible for $v_1 < \frac{1}{3}$ or $v_1 > \frac{2}{3}$. Hence we are lead to believe that (3.3) is not the correct perturbation solution. Indeed if we write

$$k_2 = k_1(1 + \delta),$$

then expansion of the Hashin bounds (2.8) and (2.9) in terms of δ gives

$$k_a^*/k_1 = 1 + v_2\delta - \tfrac{1}{2}v_1v_2\delta^2 + (v_1)\tfrac{1}{4}v_1v_2\delta^3 + \cdots \tag{3.4}$$

$$k_b^*/k_1 = 1 + v_2\delta - \tfrac{1}{2}v_1v_2\delta^2 + (v_1 + 1)\tfrac{1}{4}v_1v_2\delta^3 + \cdots. \tag{3.5}$$

A similar expansion of the perturbation solution (3.3) gives

$$k^*/k_1 = 1 + v_2\delta - \tfrac{1}{2}v_1v_2\delta^2 + (2v_2)\tfrac{1}{4}v_1v_2\delta^3 + \cdots. \tag{3.6}$$

It is only in the third-order term that the expansions differ and we see that indeed, without any reference to the present work, the perturbation solution falls outside of the Hashin bounds for $v_1 < \frac{1}{3}$ or $v_1 > \frac{2}{3}$. It seems necessary then to question the validity of even considering a perturbation solution other than that obtained by taking only the terms of the Hashin bounds (up to the second-order) for which the bounds coincide. A higher order approximation, even if the coefficient of the third-order term were to fall between the coefficients in the expansions of the Hashin bounds, can only point to one particular value for k^*, not unique to third-order accuracy, since we expect that all k^*'s between the Hashin bounds are realizable. Indeed the method proposed in the last section for calculating I_1 for a particular structure for which we know how to calculate k^* simply requires calculating the third-order coefficient in a power series expansion in δ. Then since the bounds not only approach each other for $k_2 \sim k_1$ but also approach a straight line (see Fig. 6), the required I_1 is simply obtained by proportions as follows

$$\frac{I_1 - I_{1_a}}{I_{1_b} - I_{1_a}} = \frac{k^* - k_a^*}{k_b^* - k_a^*} = \frac{C^* - C_a}{C_b - C_a},$$

where C^* indicates the third-order coefficient of the material being considered and C_a and C_b are the third-order coefficients in (3.4) and (3.5). This reduces to

$$I_1 = C^* - v_1v_2^2/2.$$

Hence, the bounds (2.5) could be rewritten in terms of C^*, the coefficient of the third-order term in a power series expansion of k^* in δ. Now the perturbation solution (3.3) is the analog of the well-known three-dimensional perturbation solution [in this case the $\frac{1}{2}$ in (3.3) becomes $\frac{1}{3}$], and the processes of arriving at these results are identical. We must then question the validity of that perturbation solution also to terms of higher order than those which coincide in the bounds which Hashin and Shtrikman[13] have derived for the three-dimensional case, and which they also show are realizable.

We close by pointing out that the technique developed here is valid only for the two-dimensional problem of fiber reinforced materials (and not for the more general three-dimensional case), since it is only for this case that Keller's theorem is valid. The extension of Keller's theorem to the three-dimensional case has been considered by this author.[14]

ACKNOWLEDGMENT

The author would like to express his appreciation to Professor C. Rubenstein for his encouragement and for useful discussions during the course of this investigation.

[1] Z. Hashin, J. Mech. Phys. Solids 13, 119 (1965).
[2] Z. Hashin, *Theory of Composite Materials*, Conf. Mechanics of Composite Materials, edited by F. Wendt, H. Liebowitz, and S. Perrone (Pergamon, N.Y., 1970), p. 201.
[3] M. Beran, Nuovo Cimento 38, 771 (1965).
[4] N. Silnutzer, "Effective Constants of Statistically Homogeneous Materials," Ph.D. dissertation, University of Pennsylvania, Philadelphia (1972).
[5] M. Miller, J. Math. Phys. 10, 1988 (1969).
[6] M. Beran and N. Silnutzer, J. Compos. Mater. 5, 246 (1971).
[7] S. Springer and S. Tsai, J. Compos. Mater. 1, 166 (1967).
[8] J. B. Keller, J. Math. Phys. 5, 548 (1964).
[9] K. Mendelson, J. Appl. Phys. 46, 917 (1975).
[10] J. Flaherty and J. B. Keller, Commun. Pure Appl. Math. 26, 565 (1973).
[11] S. Prager, J. Chem. Phys. 50, 4305 (1969).
[12] M. Hori and F. Yonezawa, J. Math. Phys. 16, 365 (1975).
[13] Z. Hashin and S. Shtrikman, J. Appl. Phys. 33, 3125 (1962).
[14] K. Schulgasser, J. Math. Phys. 17, 378 (1976).

Long-wavelength normal mode vibrations of infinite, ionic crystal lattices. II

John A. Davies

Department of Physics, Clark University, Worcester, Massachusetts 01610
(Received 15 September 1975)

In an earlier paper [J. Math. Phys. **16**, 1156 (1975)] we presented a mathematical theory of the long-wavelength normal mode vibrations of infinite crystal lattices whose particles interact with Coulomb forces. (Retardation was neglected.) The paper showed how the eigenvalues and eigenvectors of the complete long-wavelength dynamical matrix are related to the eigenvalues and eigenvectors of the dynamical matrix obtained by neglecting the contribution of the macroscopic electric field. Rules were obtained for determining whether or not the various branches of the dispersion relations for a lattice approach definite frequencies in the long-wavelength limit. The paper was restricted to the rigid ion approximation. In this paper we show that the above treatment can be easily extended to include lattices with polarizable and deformable atoms.

I. INTRODUCTION

In a recent paper,[1] we presented a mathematical theory of the long-wavelength normal mode vibrations of infinite crystal lattices with Coulomb interactions when retardation is neglected. The paper was concerned with showing how eigenvalues and eigenvectors of the complete long-wavelength dynamical matrix $\mathbf{C}^0(\hat{\boldsymbol{\phi}})$ are related to those of the long-wavelength dynamical matrix $\mathbf{A}$ where the macroscopic electric field contribution is neglected. Further, theorems were developed for determining which dispersion relations for such lattices approach definite frequencies in the long-wavelength limit.

Our treatment in Ref. 1 was limited to the case of the rigid ion model. The purpose of this paper is to point out that the mathematical approach and substantive results of Ref. 1 also apply to lattices with polarizable and deformable atoms. In particular we show that the mathematical approach and results of Ref. 1 apply with only slight modifications to the phenomenological model for lattices with polarizable and deformable atoms as presented in Sec. VI. 5 of Maradudin, Montroll, Weiss, and Ipatova.[2]

A logical step by step development of the extension of our mathematical treatment to the case of polarizable, deformable atoms would be largely an unnecessary repetition of the rather lengthy presentation of Ref. 1. Thus, we merely outline to the reader already familiar with Ref. 1 how that treatment can be extended. The notation and definitions in this article are consistent with those of Ref. 1. Lemmas and theorems of that reference are referred to by number. Whenever an equation from Ref. 1 is referred to, the equation number will be followed by the Roman numeral I.

II. THE PHENOMENOLOGICAL THEORY OF CRYSTALS WITH POLARIZABLE AND DEFORMABLE ATOMS

We list those results from Sec. VI. 5 of Ref. 2 which are applicable to this work. Some changes in notation are made in order to make equations conform to the notation of Ref. 1. In particular all quantities are made dimensionless, the names of the f particles in a primitive cell are denoted by Greek letters and Cartesian coordinates are denoted by Latin letters.

According to the phenomenological theory, the long-wavelength normal mode vibrations of lattices with polarizable, deformable atoms (with retardation neglected) are governed by the eigenvalue equation

$$\mathbf{C}^0(\hat{\boldsymbol{\phi}})\boldsymbol{\Psi}^0(\hat{\boldsymbol{\phi}}) = \{A + [4\pi a^3/v_a \epsilon_L{}^\infty(\hat{\boldsymbol{\phi}})]\mathbf{N}'(\hat{\boldsymbol{\phi}})\}\boldsymbol{\Psi}^0(\hat{\boldsymbol{\phi}}) = \lambda^0(\hat{\boldsymbol{\phi}})\boldsymbol{\Psi}^0(\hat{\boldsymbol{\phi}}). \tag{1}$$

The above equation replaces Eq. (3I). Most of the above symbols are defined in Sec. II of Ref. 1. The new quantities are $\epsilon_L{}^\infty(\hat{\boldsymbol{\phi}})$ and $\mathbf{N}'(\hat{\boldsymbol{\phi}})$. The quantity $\epsilon_L{}^\infty(\hat{\boldsymbol{\phi}})$ is the the longitudinal, optical frequency dielectric constant. It is given by

$$\epsilon_L{}^\infty(\hat{\boldsymbol{\phi}}) = 1 + 4\pi\hat{\boldsymbol{\phi}}^\dagger\boldsymbol{\chi}^\infty\hat{\boldsymbol{\phi}}, \tag{2}$$

where the 3×3 matrix $\boldsymbol{\chi}^\infty$ with elements $\chi_{ij}{}^\infty$ is a susceptibility relating the macroscopic electric field to the electrostatic polarization of the lattice. The $3f\times3f$ matrix $\mathbf{N}'(\hat{\boldsymbol{\phi}})$ consists of 3×3 submatrices $\mathbf{N}'_{\kappa\nu}(\hat{\boldsymbol{\phi}})$ defined by

$$\mathbf{N}'_{\kappa\nu}(\hat{\boldsymbol{\phi}}) = \mathbf{f}_\kappa{}^t\mathbf{L}(\boldsymbol{\phi})\mathbf{f}_\nu/(\mu_\kappa\mu_\nu)^{1/2}. \tag{3}$$

The new quantities $\mathbf{f}_\nu$ are 3×3 matrices with elements $f_{ij}(\nu)$. The symbol $\mathbf{f}_\nu$ represents the (dimensionless) transverse effective charge of the νth atom in a primitive cell. In the rigid ion model $\mathbf{f}_\nu = Z_\nu\mathbf{I}$, where $\mathbf{I}$ is the 3×3 identity matrix, and $\mathbf{N}'(\hat{\boldsymbol{\phi}})$ reduces to the matrix $\mathbf{N}(\hat{\boldsymbol{\phi}})$ of Ref. 1.

We assume that $\hat{\boldsymbol{\phi}}$, $\boldsymbol{\chi}^\infty$, and f_ν are real quantities. Thus, we make no distinction between their adjoints † and their transposes ^{t}.

Some important properties of $\boldsymbol{\chi}^\infty$ and $\mathbf{f}_\nu$ are given in Ref. 2. The susceptibility $\boldsymbol{\chi}^\infty$ is symmetric; that is

$$\boldsymbol{\chi}^\infty = \boldsymbol{\chi}^{\infty t}. \tag{4}$$

Another property is charge neutrality of the primitive cell, expressed by

$$\sum_\nu \mathbf{f}_\nu = 0. \tag{5}$$

Let the point group of the space group of the lattice be G and let $\mathbf{R}\in G$ where $\mathbf{R}$ is a 3×3, real, orthogonal matrix. Designate by $F_0(\nu,\mathbf{R})$ the type of particle into which a particle of type ν is brought by a space group operation involving $\mathbf{R}$. Then,

$$\mathbf{f}_{F_0(\nu,\mathbf{R})}=\mathbf{R}\mathbf{f}_\nu\mathbf{R}^t \tag{6}$$

and

$$\chi^\infty=\mathbf{R}\chi^\infty\mathbf{R}^t. \tag{7}$$

The following relations, which can be immediately derived from Eq. (6), are useful in extending the treatment of Ref. 1 to polarizable, deformable ions: Let $\mathbf{R}\in G$, then

$$\sum_\nu f_\nu f_\nu{}^t/\mu_\nu=\mathbf{R}\sum_\nu(f_\nu f_\nu{}^t/\mu_\nu)\mathbf{R}^t, \tag{8}$$

and

$$(\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu)^{-1}=\mathbf{R}(\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu)^{-1}\mathbf{R}^t. \tag{9}$$

In Ref. 2, expressions are derived for the macroscopic electric field, polarization, and electric displacement field amplitudes produced by a normal mode vibration Ψ propagating in the direction $\hat{\phi}$. These expressions which replace Eqs. (25I), (29I), (30I), and (31I) are rewritten below in the dimensionless form employed in Ref. 1:

$$\mathbf{E}=-[4\pi a^3/v_a\epsilon_L{}^\infty(\hat{\phi})]\mathbf{L}(\hat{\phi})\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa, \tag{10}$$

$$\mathbf{P}=\{(a^3/v_a)\mathbf{I}-[4\pi a^3/v_a\epsilon_L{}^\infty(\hat{\phi})]\chi^\infty\mathbf{L}(\hat{\phi})\}\times\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa, \tag{11}$$

and

$$\mathbf{D}=(4\pi a^3/v_a)\sum_\kappa(f_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa - [4\pi a^3/v_a\epsilon_L{}^\infty(\hat{\phi})](\mathbf{I}+4\pi\chi^\infty)\mathbf{L}(\hat{\phi})\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa. \tag{12}$$

Next consider lattices for which the point group G belongs to the regular (cubic) system. In the remainder of this article we shall for the sake of brevity refer to such lattices as cubic lattices. For such lattices many of the above expressions reduce to simpler forms.

If a lattice is cubic (in the above sense), then Eq. (7) implies that

$$\chi^\infty=\chi^\infty\mathbf{I}, \tag{13}$$

where χ^∞ is a number. From Eq. (2), it follows that $\epsilon_L{}^\infty$ assumes the $\hat{\phi}$-independent value

$$\epsilon_L{}^\infty=1+4\pi\chi^\infty. \tag{14}$$

Further, we see from Eqs. (8) and (9) that for cubic lattices

$$\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu=\alpha\mathbf{I}, \tag{15}$$

where α is a number given by

$$\alpha=\tfrac{1}{3}\,\mathrm{Tr}\sum_\nu\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu. \tag{16}$$

Finally note that Eq. (12) reduces to

$$\mathbf{D}=(4\pi a^3/v_a)\mathbf{T}(\hat{\phi})\textstyle\sum_\kappa(f_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa, \tag{17}$$

where $\mathbf{T}(\hat{\phi})$ is defined by Eq. (18I).

III. A RESTRICTION ON THE TYPES OF LATTICES TO BE CONSIDERED

In this paper, we make one restriction on the types of three-dimensional, ionic lattices to be analyzed. We confine our treatment to lattices for which the real, Hermitian matrix $\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu$ has an inverse. This restriction is equivalent to assuming that

$$\det\sum_\nu\frac{f_\nu f_\nu{}^t}{\mu_\nu}\neq0. \tag{18}$$

The principal values of $\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu$ must be nonnegative. Thus, our restriction is also equivalent to insisting that all of the principal values are positive definite.

The physical meaning of the above limitation is clear. Imagine that we work with the principle axes of $\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu$ as coordinate axes. It is easily shown that the vanishing of the ith principal value of $\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu$ implies that only zeros occur in the ith row of $\mathbf{f}_\nu$. But referring to Eq. (10), we see that then no set of long-wavelength particle displacements in the direction of the ith principal axis will result in a macroscopic electric field. In effect, we have a three-dimensional lattice which exhibits ionic properties in fewer than three dimensions. An analysis of such lattices would be interesting. However, we exclude such lattices from our present treatment.

IV. EXTENSION OF THE ANALYSIS TO POLARIZABLE, DEFORMABLE ATOMS (PART I)

The starting point for our treatment of ionic lattices in Ref. 1 was to show that the matrix $\mathbf{N}(\hat{\phi})$ in Eq. (3I) has the $(3f-1)$-fold degenerate eigenvalue zero and only one nonvanishing eigenvalue. This fact is also true for the matrix $\mathbf{N}'(\hat{\phi})$ in Eq. (1). By direct calculation, we find that the normalized vector $\Psi^{In}(\hat{\phi})$ redefined by

$$\Psi_\nu{}^{In}(\hat{\phi})=[\mathrm{Tr}\sum_\omega\mathbf{f}_\omega{}^t\mathbf{L}(\hat{\phi})\mathbf{f}_\omega/\mu_\omega]^{-1/2}\mu_\nu{}^{-1/2}\mathbf{f}_\nu^t\hat{\phi}, \tag{19}$$

is an eigenvector of $\mathbf{N}'(\hat{\phi})$ corresponding to the eigenvalue

$$\lambda_N{}^{In}(\hat{\phi})=\mathrm{Tr}(\sum_\nu\mathbf{f}_\nu{}^t\mathbf{L}(\hat{\phi})\mathbf{f}_\nu/\mu_\nu) =\mathrm{Tr}[\sum_\nu(\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu)\mathbf{L}(\hat{\phi})]. \tag{20}$$

[If we set $\mathbf{f}_\nu=Z_\nu\mathbf{I}$, we regain the original definition of $\Psi^{In}(\hat{\phi})$ given in Ref. 1.] By working with the normal coordinates of $\sum_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu$ one can easily show that the restriction stipulated in Sec. III [Eq. (18)] is the necessary and sufficient condition that the factor of normalization in Eq. (19) be finite and that $\lambda_N{}^{In}(\hat{\phi})\neq0$ for all $\hat{\phi}$. In general $\lambda_N{}^{In}(\hat{\phi})$ is $\hat{\phi}$-dependent. However, Eqs. (15) and (16) show that in the case of cubic lattices (lattices whose point groups belong to the regular system) $\lambda_N{}^{In}$ is $\hat{\phi}$-independent and has the value given by

$$\lambda_N{}^{In}=\tfrac{1}{3}\,\mathrm{Tr}(\sum_\nu\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu). \tag{21}$$

A direct calculation also shows that any vector orthogonal to $\Psi^{In}(\hat{\phi})$ is an eigenvector of $\mathbf{N}'(\hat{\phi})$ corresponding to the eigenvalue zero. Since $\mathbf{N}'(\hat{\phi})$ is Hermitian, this eigenvalue is $(3f-1)$-fold degenerate. Further, from the theorem that eigenvectors of Hermitian matrices are orthogonal if they correspond to different eigenvalues,

it follows that the conditions $\mathbf{N}'(\hat{\phi})\Psi=0$ and $\mathbf{L}(\hat{\phi})\times\sum_\kappa \mathbf{f}_\kappa \mu_\kappa^{-1/2}\Psi_\kappa=0$ are equivalent.

The spaces and subspaces S_{3f}(total), $S_{3f-1}(\lambda_N=0;\hat{\phi})$, S_3(normal), and S_{3f-3}(zero) are now defined in complete analogy with their definitions in Ref. 1. The subspace S_3(normal) consists of vectors Ψ of the form

$$\Psi_\nu = k\mu_\nu^{-1/2}\mathbf{f}_\nu{}^t\psi, \tag{22}$$

where $k\neq 0$ is a number and ψ is an arbitrary three-dimensional vector. The restriction stipulated in Sec. III insures that $\Psi\neq 0$ if $\psi\neq 0$ and, therefore, that S_3 (normal) actually is three-dimensional.

In extending the work of Ref. 1 to include polarizable, deformable atoms, Lemmas I and II of Ref. 1 are to be left unchanged. Lemma III of Ref. 1 remains essentially the same except that now the branch of the dispersion relations whose long-wavelength eigenvectors are parallel to $\Psi^{in}(\phi)$ approaches a frequency corresponding to

$$\lambda=\lambda^a+[4\pi a^3/v_a\epsilon_L{}^\infty(\hat{\phi})]\,\mathrm{Tr}\sum{}_\nu \mathbf{f}_\nu{}^t\mathbf{L}(\hat{\phi})\mathbf{f}_\nu/\mu_\nu$$

as ϕ approaches zero. In general this frequency is $\hat{\phi}$-dependent. However, since Lemma III requires that $\mathbf{A}$ have at least a three-fold degenerate eigenvalue (non-acoustic), it is actually only applicable to cubic lattices (barring accidental degeneracies). For such lattices Eqs. (14)—(16) show that

$$\lambda=\lambda^a+(4\pi a^3/v_a\epsilon_L{}^\infty)\tfrac{1}{3}\,\mathrm{Tr}\sum{}_\nu \mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu,$$

which is $\hat{\phi}$-independent.

A necessary and sufficient condition that a single given branch of the dispersion relations approaches a definite frequency (independent of $\hat{\phi}$) in the long-wavelength limit can be determined by the same method applied to the rigid ion model in an earlier paper.[3] We find that a necessary and sufficient condition that λ^0 in Eq. (1) be independent of $\hat{\phi}$ is that for all $\hat{\phi}$, the $\Psi^0(\hat{\phi})$ for the branch obey either the equation

$$\mathbf{N}'(\hat{\phi})\Psi^0(\hat{\phi})=0, \tag{23}$$

or the equation

$$\begin{aligned}\sum_\kappa (\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa{}^0(\hat{\phi})&=[4\pi/\epsilon_L{}^\infty(\hat{\phi})]\chi^\infty\mathbf{L}(\hat{\phi})\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\\&\times\Psi_\kappa{}^0(\hat{\phi})+[\mathbf{L}(\hat{\phi})/\epsilon_L{}^\infty(\hat{\phi})]\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa{}^0(\hat{\phi}).\end{aligned} \tag{24}$$

Equation (24) is equivalent to the equation

$$\begin{aligned}\sum_\kappa \mathbf{f}_\kappa\Psi_\kappa/\mu_\kappa{}^{1/2}-[4\pi/\epsilon_L{}^\infty(\hat{\phi})]\chi^\infty\\\times\mathbf{L}(\hat{\phi})\sum_\kappa\mathbf{f}_\kappa\Psi_\kappa/\mu_\kappa{}^{1/2}=h(\hat{\phi})\hat{\phi},\end{aligned} \tag{25}$$

where $h(\hat{\phi})$ is an arbitrary number.[4]

Referring to Eqs. (10) and (12), we see that a branch will approach a definite frequency if and only if, for all $\hat{\phi}$, in the long-wavelength limit either $\mathbf{E}=0$ or $\mathbf{D}=0$. The same result was obtained in Ref. 1.

On the other hand, Lemma IV of Ref. 1 must be modified. The condition that $\Psi^0(\hat{\phi})\in S_{3f-2}(\lambda_M=0;\hat{\phi})$ stated in the lemma must be replaced by the condition that Eq. (24) be obeyed.

For cubic lattices, Eqs. (13) and (14) show that Eq. (24) reduces to

$$\mathbf{T}(\hat{\phi})\sum_\kappa(\mathbf{f}_\kappa/\mu_\kappa{}^{1/2})\Psi_\kappa{}^0(\hat{\phi})=0, \tag{26}$$

where $\mathbf{T}(\hat{\phi})$ is defined by Eq. (18I). Thus, for lattices whose point groups belong to the regular system, it is convenient to introduce a matrix analogous to the matrix $\mathbf{M}(\hat{\phi})$ defined in Ref. 1 by Eq. (17I). This is the Hermitian matrix $\mathbf{M}'(\hat{\phi})$ given by

$$\mathbf{M}'_{\kappa\nu}(\hat{\phi})=(\mathbf{f}_\kappa{}^t/\mu_\kappa{}^{1/2})\mathbf{T}(\hat{\phi})\mathbf{f}_\nu/\mu_\nu{}^{1/2}. \tag{27}$$

Using Eqs. (15) and (16), we find that, for cubic lattices, $\mathbf{M}'(\hat{\phi})$ has the two-fold degenerate eigenvalue

$$\lambda_M=\tfrac{1}{3}\,\mathrm{Tr}(\sum_\nu f_\nu f_\nu{}^t/\mu_\nu), \tag{28}$$

with corresponding normalized eigenvectors of the form

$$\Psi_\kappa=(\tfrac{1}{3}\,\mathrm{Tr}\sum_\omega \mathbf{f}_\omega\mathbf{f}_\omega^t/\mu_\omega)^{-1/2}(\mathbf{f}_\kappa{}^t/\mu_\kappa)\hat{\eta}(\hat{\phi}), \tag{29}$$

where $\hat{\phi}^\dagger\hat{\eta}(\hat{\phi})=0$. Equation (18) insures that $\lambda_M\neq 0$ and that the normalization factor in Eq. (29) is finite. Any vector orthogonal to a vector of the form given by Eq. (29) is also an eigenvector of $\mathbf{M}'(\hat{\phi})$ corresponding to the $(3f-2)$-fold degenerate eigenvalue zero. The conditions $\mathbf{M}'(\hat{\phi})\Psi=0$ and $\mathbf{T}(\hat{\phi})\sum_\nu \mu_\nu{}^{-1}\mathbf{f}_\nu\Psi_\nu=0$ are equivalent.

We now define the subspaces $S_{3f-2}(\lambda_M=0;\hat{\phi})$ and $S_2(\lambda_M=\frac{1}{3}\,\mathrm{Tr}\sum_\omega \mu_\omega{}^{-1}\mathbf{f}_\omega\mathbf{f}_\omega{}^t;\hat{\phi})$ in complete analogy with the corresponding subspaces introduced in Ref. 1, i.e., $S_{3f-2}(\lambda_M=0;\hat{\phi})$ and $S_2(\lambda_M=\sum_\kappa Z_\kappa{}^2/\mu_\kappa;\hat{\phi})$.

It is important to emphasize that the operator $\mathbf{M}'(\hat{\phi})$ and the subspaces $S_{3f-2}(\lambda_M=0;\hat{\phi})$ and $S_2(\lambda_M=\frac{1}{3}\,\mathrm{Tr}\sum_\omega \mu_\omega=\frac{1}{3}\,\mathrm{Tr}\sum_\omega \mu_\omega{}^{-1}\mathbf{f}_\omega\mathbf{f}_\omega{}^t;\hat{\phi})$ are to be employed only when dealing with cubic lattices.

Lemma V of Ref. 1 remains unchanged when the work is extended to polarizable, deformable atoms.

In Sec. IV of Ref. 1 an alternative form for writing $\mathbf{C}^0(\hat{\phi})$ was introduced. In dealing with polarizable, deformable atoms, we find this alternative form useful only when dealing with cubic lattices. For such lattices, we can write

$$\mathbf{C}^0(\hat{\phi})=\mathbf{A}'-(4\pi a^3/v_a\epsilon_L{}^\infty)\mathbf{M}'(\hat{\phi}), \tag{30}$$

where $\mathbf{A}'$ is the $\hat{\phi}$-independent matrix defined by

$$\mathbf{A}'=\mathbf{A}+(4\pi a^3/v_a\epsilon_L{}^\infty)[\mathbf{N}'(\hat{\phi})+\mathbf{M}'(\hat{\phi})]. \tag{31}$$

Using Eqs. (5) and (6) and the same general approach used in Sec. IV of Ref. 1, we find that for cubic lattices any properties of $\mathbf{A}$ derived by group theoretical methods apply with equal validity to $\mathbf{A}'$.[5]

Sec. V of Ref. 1 defines the term accidental degeneracy. No modification of this section is required.

V. EXTENSION OF THE ANALYSIS TO POLARIZABLE, DEFORMABLE ATOMS (PART II)

Equations (35I)—(41I) in Sec. VI of Ref. 1 list projection operators to various subspaces of S_{3f}(total). These expressions must be generalized when the work is extended to include lattices with polarizable, deformable atoms. The new expressions are listed be-

low (the notation conforming with that used in Ref. 1):

$$\mathbf{P}(\Psi^{In}(\hat{\phi})) = (\mathrm{Tr}\sum_{\nu} \mathbf{f}_\nu{}^t \mathbf{L}(\hat{\phi})\mathbf{f}_\nu/\mu_\nu)^{-1}\mathbf{N}'(\hat{\phi}), \tag{32}$$

$$\mathbf{P}(S_{3f-1}(\lambda_N = 0;\hat{\phi})) = \mathbf{I} - \mathbf{P}[\Psi^{In}(\hat{\phi})], \tag{33}$$

$$\mathbf{P}(S_3(\text{acoustic})) = (\sum_{\omega} \mu_\omega)^{-1}(\mu_\kappa \mu_\nu)^{1/2}\mathbf{I}, \tag{34}$$

$$\mathbf{P}_{\kappa\nu}(S_3(\text{normal})) = \frac{\mathbf{f}_\kappa{}^t}{\mu_\kappa}\left(\sum_{\omega}\frac{f_\omega f_\omega{}^t}{\mu_\omega}\right)^{-1}\frac{\mathbf{f}_\nu}{\mu_\nu{}^{1/2}}, \tag{35}$$

and

$$\mathbf{P}(S_{3f-3}(\text{zero})) = \mathbf{I} - \mathbf{P}(S_3(\text{normal})). \tag{36}$$

The following additional projection operators are employed when dealing with cubic lattices (lattices for which G belongs to the regular system):

$$\mathbf{P}(S_{3f-2}(\lambda_M = 0;\hat{\phi})) = \mathbf{I} - (\tfrac{1}{3}\,\mathrm{Tr}\sum_{\nu}\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu)^{-1}\mathbf{M}'(\hat{\phi}), \tag{37}$$

and

$$\mathbf{P}(S_2(\lambda_M = \tfrac{1}{3}\,\mathrm{Tr}\sum_{\nu}\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu;\hat{\phi})) = (\tfrac{1}{3}\,\mathrm{Tr}\sum_{\nu}\mathbf{f}_\nu\mathbf{f}_\nu{}^t/\mu_\nu)^{-1}\mathbf{M}'(\hat{\phi}). \tag{38}$$

The remaining development in Sec. VI requires little modification except in the details of the proofs of the lemmas. With the aid of Eq. (27), we find that Lemmas VI and VII remain true provided that it is understood that they apply only to cubic lattices and that the symbol $S_2(\lambda_M = \sum_\kappa Z_\kappa{}^2/\mu_\kappa;\hat{\phi})$ in Lemma VII is replaced by $S_2(\lambda_M = \frac{1}{3}\,\mathrm{Tr}\sum_\nu \mu_\nu{}^{-1}\mathbf{f}_\nu\mathbf{f}_\nu{}^t;\hat{\phi})$. The statements of Lemmas VII through XIII require no modification. Their proofs are similar to those given for the point ion case in Ref. 1. Equation (6) is helpful in overcoming the difficulty that the effective charges are now matrices instead of numbers.[6]

The subspaces S(polar) and S(nonpolar) are defined exactly as in Ref. 1.

Lemma XIV of Ref. 1 should be replaced by the following more general statement: Consider a branch of the dispersion relations all of whose long-wavelength eigenvectors $\Psi^0(\hat{\phi})$ lie in S(polar). If the $\Psi^0(\hat{\phi})$ obey Eq. (24) for each $\hat{\phi}$, then the $\Psi^0(\hat{\phi})$ span a subspace of S_{3f}(total) which is exactly three-dimensional.[7]

Lemma XV requires no modification. A generalization of Lemma XVI which holds true for lattices with polarizable and deformable atoms is the following: Consider a branch of the dispersion relations all of whose eigenvectors $\Psi^0(\hat{\phi})$ lie in S(polar). A necessary and sufficient condition that the branch approach a definite frequency in the long-wavelength limit is that either the $\Psi^0(\hat{\phi})$ lie in $S_{3f-1}(\lambda_N = 0;\hat{\phi})$ for all $\hat{\phi}$ or that the $\Psi^0(\hat{\phi})$ satisfy Eq. (24) for all $\hat{\phi}$. Further, for no $\hat{\phi}$ does $\Psi^0(\hat{\phi})$ both lie in $S_{3f-1}(\lambda_N = 0;\hat{\phi})$ and satisfy Eq. (24).

Some comments concerning the proof of the above statement are given in a footnote.[8]

All of the work in Sec. VII of Ref. 1 preceding the proof of Lemma XVII is based upon general symmetry arguments and not upon any particular model for a lattice. Thus, it applies to lattices with polarizable and deformable atoms as well as to the rigid ion model. The statements of Lemmas XVII and XVIII require no modification although some minor changes are required in the proof of Lemma XVII.[9]

We then find that Theorem I of Ref. 1 is correct as stated for the case of polarizable, deformable atoms. Eqs. (79I) and (80I), which are useful results for making calculations as well as decisive steps in the proof of Theorem I require no modification. An important step in the proof of Eq. (79I) for the point ion model was provided by Eqs. (74I) and (75I). These equations must be replaced by the following:

$$\sum_{\nu\approx\tau}\mathbf{f}_\nu\Psi_\nu{}^{\Gamma\mu i} = c_\tau{}^\Gamma(\mu)\begin{bmatrix}\delta_{i1}\\ \delta_{i2}\\ \delta_{i3}\end{bmatrix}, \tag{39}$$

where

$$c_\tau{}^\Gamma(\mu) = \sum_{\nu\approx\tau}(\mathbf{f}_\nu\Psi_\nu{}^{\Gamma\mu 3})_3. \tag{40}$$

In order to avoid confusion between effective charges and expansion coefficients, replace f_i in Eq. (76I) with the symbol h_i. Eq. (77I) is then replaced by the more general equation

$$\sum_{\tau}{}'(\mathbf{f}_\kappa{}^t c_\tau{}^\Gamma(\mu)/\mu_\kappa{}^{1/2}\mu_\tau{}^{1/2})\phi^{-1}\begin{bmatrix}\phi_1\\ \phi_2\\ \phi_3\end{bmatrix}(h^\dagger\hat{\phi}) = 0. \tag{41}$$

Having made the above generalizations, the reader familiar with Ref. 1 should have no difficulty in completing the proof of Theorem I for polarizable, deformable atoms using the same arguments as those in Sec. VIII of Ref. 1.

In Sec. IX of Ref. 1 a method is developed for determining the long-wavelength eigenvectors for a branch of the dispersion relations of a cubic lattice in the point ion approximation once an eigenvector for the branch is determined for just one value of $\hat{\phi}$. We find that the method remains unchanged for the case of polarizable and deformable atoms. Only some of the details of the proof of the method need be changed. [Here we take the opportunity to point out that a typographical error occurs in Eq. (86I). In that equation $\bar{\mathbf{Q}}(\mathbf{R})$ should be replaced with $\bar{\mathbf{Q}}^\dagger(\mathbf{R})$.]

Theorem II also remains true as stated for the case of polarizable, deformable ions. Equations (98I), (99I), (110I), and (111I) also remain true. In the proof of Eq. (98I), Eqs. (95I) and (96I) are to be replaced by

$$\sum_{\nu\approx\tau}\mathbf{f}_\nu\Psi_\nu{}^{(xy)i} = c_\tau{}^{(xy)}\begin{bmatrix}\delta_{i1}\\ \delta_{i2}\\ 0\end{bmatrix}, \tag{42}$$

where

$$c_\tau{}^{(xy)} = \sum_{\nu\approx\tau}(\mathbf{f}_\nu\Psi_\nu{}^{(xy)1})_1. \tag{43}$$

In the proof of Eq. (110I), replace Eqs. (103I) and (104I)

with

$$\sum_{\nu\approx\tau}\mathbf{f}_\nu\Psi_\nu^{(+)}=\sum_{\nu\approx\tau}\mathbf{f}_\nu\Psi_\nu^{(-)*}=c(\tau)\begin{bmatrix}1\\-i\\0\end{bmatrix}, \tag{44}$$

where

$$c(\tau)=\tfrac{1}{2}\sum_{\nu\approx\tau}[(\mathbf{f}_\nu\Psi_\nu^{(+)})_1+i(\mathbf{f}_\nu\Psi_\nu^{(+)})_2]. \tag{45}$$

Equation (105I), which defines the constant C appearing in Eqs. (110I) and (111I), should be replaced with

$$C=\sum_\tau{}' c(\tau)/\mu_\tau{}^{1/2}. \tag{46}$$

VI. CONCLUDING REMARKS. POLARITON MODES

We have shown that the treatment of Ref. 1 can be extended to lattices with polarizable and deformable atoms with no changes in such final results as Theorems I and II, the physical criterion that a branch of the dispersion relations approach a definite frequency (either **E** or **D** vanish), and the general forms of long-wavelength eigenvectors. Only small modifications in mathematical details are required.

It is important to emphasize that retardation is neglected in this work. The results therefore do not apply to the long-wavelength behavior of polariton modes. In fact, our results correspond to those obtained by letting the propagation vector go to infinite magnitude in treatments of long-wavelength polaritons (with spacial dispersion neglected).

The above statement can be simply illustrated by considering polaritons in an infinite crystal of rigid ions. In terms of the notation used in Ref. 1, such long-wavelength polariton modes are governed by the equation

$$\lambda^0(\phi)=\mathbf{A}\Psi^0(\phi)+(4\pi a^3/v_a)\times(\mathbf{N}(\hat\phi)-\mathbf{M}(\hat\phi)/\{[a\phi^2/a_0\lambda(\phi)]-1\})\Psi^0(\phi), \tag{47}$$

where $a_0=e^2/mc^2$, c is the speed of light, and e, m, and a are the electronic charge, a typical mass, and a typical cell dimension, respectively.[10] If we let ϕ approach infinity and assume $\lambda(\phi)$ approaches a finite value (thus ignoring modes which become purely electromagnetic), Eq. (47) reduces to Eq. (3I) of Ref. 1. However, consider the limit of the above equation as ϕ goes to zero. If $\lambda(\phi)$ does not go to zero, Eq. (47) becomes

$$\lambda^0\Psi^0=\{\mathbf{A}+(4\pi a^3/v_a)[\mathbf{N}(\hat\phi)+\mathbf{M}(\hat\phi)]\}\Psi^0=\mathbf{A}'\Psi^0. \tag{48}$$

Thus, each branch of the dispersion relations for the polariton modes either approaches zero or approaches some other eigenvalue of the $\hat\phi$-independent matrix $\mathbf{A}'$. Therefore, all such branches approach definite frequencies as ϕ approaches zero, regardless of the lattice symmetry.

[1] J. A. Davies and C. L. Mainville, J. Math. Phys. 16, 1156 (1975).

[2] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York and London, 1971).

[3] J. A. Davies, J. Math. Phys. 13, 1207 (1972). See Sec. IX.

[4] In both the derivation of the preceding result and of the corresponding result for rigid ions in Ref. 3 (Lemma IV of Ref. 1), we assume initially that $\Psi^0(\hat\phi)$ is real. Equations (23) and (24) are derived from perturbation theory using this assumption. However, the final lemma is correct even if $\Psi^0(\hat\phi)$ is complex; that is, $\Psi^0(\hat\phi)=\mathcal{R}\Psi^0(\hat\phi)+i\mathcal{I}\Psi^0(\hat\phi)$. The operator $\mathbf{C}^0(\hat\phi)$ is real and Hermitian. Thus, if $\Psi^0(\hat\phi)$ is an eigenvector of $\mathbf{C}^0(\hat\phi)$ corresponding to the eigenvalue λ^0, then $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$ are separately eigenvectors of $\mathbf{C}^0(\hat\phi)$ corresponding to the same eigenvalue. Thus, a necessary and sufficient condition that the branch approach a definite frequency is that the requirement of the lemma holds separately for $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$. Further, if $\Psi^0(\hat\phi)$ obeys Eq. (23) or (24), then both $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$ will obey the same equation. Thus, the condition that $\Psi^0(\hat\phi)$ obey either Eq. (23) or (24) is a sufficient condition. It is also a necessary condition. If $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$ are linearly dependent, then $\Psi^0(\hat\phi)$ is just a complex constant times a real vector and the proof of the lemma is the same as for real $\Psi^0(\hat\phi)$. If $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$ are linearly independent, then λ^0 is at least twofold degenerate. For general directions of $\hat\phi$ and barring accidental degeneracies, $\mathbf{C}^0(\hat\phi)$ can have such a degeneracy only if the degeneracy is required by Lemmas I, II, or III. In any of these cases, both $\mathcal{R}\Psi^0(\hat\phi)$ and $\mathcal{I}\Psi^0(\hat\phi)$ obey Eq. (23). Thus, $\Psi^0(\hat\phi)$ must obey Eq. (23).

[5] An important step in showing that **A** and **A'** have the same symmetry properties is to show that $\mathbf{T}(0,\mathrm{R})$ as defined by Eq. (13I) commutes with $\mathbf{N}'(\hat\phi)+M'(\hat\phi)$. Using Eqs. (3) and (27), we see that $[\mathbf{N}'(\hat\phi)+\mathbf{M}'(\hat\phi)]_{\kappa\nu}=\mathbf{f}_\kappa{}^\dagger\mathbf{f}_\nu(\mu_\kappa\mu_\nu)^{-1/2}$. Then making extensive use of Eq. (6), we obtain the result

$$\begin{aligned}
&\{\mathbf{T}(0,\mathrm{R})[\mathbf{N}'(\hat\phi)+\mathbf{M}'(\hat\phi)]\}_{\kappa\nu}\\
&\quad=\textstyle\sum_\lambda\mathrm{R}\delta[\kappa,F_0(\lambda,\mathrm{R})]\mathbf{f}_\lambda{}^\dagger\mathbf{f}_\nu(\mu_\lambda\mu_\nu)^{-1/2}\\
&\quad=\mathbf{R}\mathbf{f}^\dagger_{F_0^{-1}(\kappa,\mathrm{R})}\mathbf{f}_\nu(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\mathbf{f}_\kappa{}^\dagger\mathrm{R}\mathbf{f}_\nu(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\mathbf{f}_\kappa{}^\dagger\mathbf{f}_{F_0(\nu,\mathrm{R})}\mathrm{R}(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\textstyle\sum_\lambda\mathbf{f}_\kappa{}^\dagger\mathbf{f}_\lambda\delta[\lambda,F_0(\nu,\mathrm{R})]\mathrm{R}(\mu_\kappa\mu_\lambda)^{-1/2}\\
&\quad=\{[\mathbf{N}'(\hat\phi)+\mathbf{M}'(\hat\phi)]\mathbf{T}(0,\mathrm{R})\}_{\kappa\nu}.
\end{aligned}$$

[6] Lemma VIII follows directly from Eqs. (5), (34), and (35), and the procedure used in Ref. 1. An important step in the proof of Lemma IX is showing that $\mathbf{T}(0,\mathrm{R})$ and $\mathbf{P}[S_{3f-3}(\mathrm{zero})]$ commute. This statement will be ture if $\mathbf{T}(0,\mathrm{R})$ and $\mathbf{P}[S_3(\mathrm{normal})]$ commute. Using Eqs. (13I), (6), (9), and (35) we obtain

$$\begin{aligned}
&\{\mathbf{T}(0,\mathrm{R})\mathbf{P}[S_3(\mathrm{normal})]\}_{\kappa\nu}\\
&\quad=\textstyle\sum_\lambda\mathrm{R}\delta[\kappa,F_0(\lambda,\mathrm{R})]\mathbf{f}_\lambda{}^\dagger(\sum_\omega\mathbf{f}_\omega\mathbf{f}_\omega{}^\dagger/\mu_\omega)^{-1}\mathbf{f}_\nu(\mu_\lambda\mu_\nu)^{-1/2}\\
&\quad=\mathrm{R}\mathbf{f}^\dagger_{F_0^{-1}(\kappa,\mathrm{R})}(\textstyle\sum_\omega\mathbf{f}_\omega\mathbf{f}_\omega{}^\dagger/\mu_\omega)^{-1}\mathbf{f}_\nu(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\mathbf{f}_\kappa{}^\dagger(\textstyle\sum_\omega\mathbf{f}_\omega\mathbf{f}_\omega{}^\dagger/\mu_\omega)^{-1}\mathrm{R}\mathbf{f}_\nu(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\mathbf{f}_\kappa{}^\dagger(\textstyle\sum_\omega\mathbf{f}_\omega\mathbf{f}_\omega{}^\dagger/\mu_\omega)^{-1}\mathbf{f}_{F_0(\nu,\mathrm{R})}\mathrm{R}(\mu_\kappa\mu_\nu)^{-1/2}\\
&\quad=\textstyle\sum_\lambda\mathbf{f}_\kappa{}^\dagger(\sum_\omega\mathbf{f}_\omega\mathbf{f}_\omega{}^\dagger/\mu_\omega)^{-1}\mathbf{f}_\lambda(\mu_\lambda\mu_\kappa)^{-1/2}\mathrm{R}\delta(\lambda,F_0(\nu,\mathrm{R}))\\
&\quad=\{\mathbf{P}[S_3(\mathrm{normal})]\mathbf{T}(0,\mathrm{R})\}_{\kappa\nu}.
\end{aligned}$$

[7] The generalization of Lemma XIV is easily proved by writing out Eq. (25) in terms of a set of Cartesian coordinates relative to which χ^∞ is diagonal. One can then easily show that the vectors $\Psi^0(\hat\phi^{(1)})$, $\Psi^0(\hat\phi^{(2)})$, and $\Psi^0(\hat\phi^{(3)})$ are linearly independent, where $\hat\phi^{(i)}=(\delta_{i1},\delta_{i2},\delta_{i3})$.

[8] The proof is essentially the same as that given in Appendix B of Ref. 1. The chief problem is to rederive Eq. (B7I). Once this is done, the proof differs from that in Ref. 1 only in minor details. Equation (B7I) is rederived as follows. Suppose that the branch $\Psi^0(\hat\phi^{(a)})$ obeys Eq. (23) and $\Psi^0(\hat\phi^{(b)})$ obeys Eq. (24). Then $\mathbf{A}\Psi^0(\hat\phi^{(a)})=\lambda^0\Psi^0(\hat\phi^{(a)})$ and

$$\begin{aligned}
&[\mathbf{A}+(4\pi a^3/(v_a\epsilon_L{}^\infty(\hat\phi^{(b)}))\mathbf{N}'(\hat\phi^{(b)})]\Psi^0(\hat\phi^{(b)})\\
&\quad=\lambda^0\Psi^0(\hat\phi^{(b)}).
\end{aligned}$$

It follows that

$$\Psi^{0\dagger}(\hat{\phi}^{(a)})N'(\hat{\phi}^{(b)})\Psi^{0}(\hat{\phi}^{(b)}) = 0$$

or equivalently

$$[\sum_{\kappa} f_{\kappa}\mu_{\kappa}^{-1/2}\Psi_{\kappa}^{0}(\hat{\phi}^{(a)})]^{\dagger}L(\hat{\phi}^{(b)})\sum_{\nu} f_{\nu}\mu_{\nu}^{-1/2}\Psi_{\nu}(\hat{\phi}^{(b)}) = 0.$$

$\Psi^{0}(\hat{\phi}^{(a)})$ obeys the equation $\sum_{\kappa} f_{\kappa}\mu_{\kappa}^{-1/2}\Psi_{\kappa}^{0}(\hat{\phi}^{(a)}) = k\hat{\eta}(\hat{\phi}^{(a)})$ where $k \neq 0$. Using Eq. (25) and the identity $L(\hat{\phi})\chi^{\infty}L(\hat{\phi}) = \{[\epsilon_L{}^{\infty}(\hat{\phi}) - 1]/4\pi\}L(\hat{\phi})$, we find that

$$L(\hat{\phi}^{(b)})\sum_{\nu} f_{\nu}\mu_{\nu}^{-1/2}\Psi_{\nu}^{0}(\hat{\phi}^{(b)}) = \epsilon_L{}^{\infty}(\hat{\phi}^{(b)})h(\hat{\phi}^{(b)})\hat{\phi}^{(b)},$$

where $h(\hat{\phi}^{(b)}) \neq 0$. Equation (B7I) follows immediately.

[9]In order to prove Lemma XVII, use Eq. (6) to show that

$$\{P^{\mu}P[S_3(\text{normal})]\}_{\kappa\omega}$$

$$= f_{\kappa}{}^{\dagger}\mu_{\kappa}^{-1/2}(n_{\mu}/g)(\sum_{R}\chi^{(\mu)*}(R)R)\ (\sum_{\omega} f_{\omega}f_{\omega}{}^{\dagger}/\mu_{\omega})^{-1}f_{\nu}\mu_{\nu}^{-1/2}.$$

The above equation replaces Eq. (66I). The above expression vanishes if none of x, y, or z belongs to $D^{(\mu)}$. To show that the above expression does not vanish if any of x, y, or z belong to $D^{(\mu)}$, calculate $\text{Tr}\sum_{\kappa}\{P^{\mu}P[S_3(\text{normal})]\}_{\kappa\kappa}$. With the aid of Eq. (8), we find that $\text{Tr}\sum_{\kappa}\{P^{\mu}P(S_3(\text{normal})]\}_{\kappa\kappa}$ $= (n_{\mu}/g)\,\text{Tr}\sum_{R}\chi^{(\mu)*}(R)\mathbf{R}$. Equations (55I), (57I), (59I), and (64I) show that this quantity cannot vanish if any of x, y, or z belong to $D^{(\mu)}$.

[10]Equation (47) is obtained from Eq. (6.6.13) of Ref. 2.

Spectrum generating algebras and Lie groups in classical mechanics

M. Andrié

Fachhochschule Köln, D 5 Köln 1, Claudiusstr. 1, Germany
(Received 3 March 1975)

We give a general framework for a geometric foundation of time dependent classical mechanics. The theory is based on the concept of evolution space which is phase space extended by time. Lie algebras of constants of motion which may possess explicit time dependence are constructed, and general conditions for getting global Lie group actions from infinitesimal actions are derived. In a natural way these groups map solutions of the Hamiltonian equations of motion onto one another and act on the orbit space via symplectic transformations. The theory is applied to the nonrelativistic free particle, the harmonic and damped oscillator, nonstationary quadratic systems, and to the motion of a particle in constant electromagnetic fields.

I. INTRODUCTION

A lot of structures of classical mechanics are motivated by analogies between quantum theory and classical dynamics. The classification of elementary particles and quantum mechanical states of a physical problem is essentially based on group theoretical methods.[1–7] In this scheme an elementary relativistically invariant quantum mechanical system is described by irreducible representations of the inhomogeneous Lorentz group.[8] For a nonrelativistic particle the group action is given by the Galilean group.

Similar methods can be applied in classical physics. The basic objects are given by phase spaces (symplectic manifolds) and evolution spaces for time dependent problems. Invariance and noninvariance properties can be characterized by groups of transformations acting on these spaces.

General methods have been developed for getting global actions of symmetry groups from infinitesimal actions on phase spaces of conservative mechanics.[7,9–12]

In this paper we shall give a framework for time dependent mechanics by using evolution spaces[7,13] such that all relevant physical properties can be derived in a natural and stringent way. For the construction of group actions we use finite dimensional spectrum generating Lie algebras[2,3,7,14] of constants of motion of classical orbits. The infinitesimal actions of the corresponding groups are extended to global ones.

In special cases the set of orbits of different energies defines a manifold on which a group acts via symplectic transformations. The general theory is applied to several physical problems.

2. PRELIMINARIES[15–18]

In this section we set out the standard definitions and results of differential geometry. The relationship with the language the physicists are familiar with is indicated.

Let M be an n-dimensional C^∞-differentiable manifold. We denote by $C=C^\infty(M)$, the space of real valued C^∞-functions $f: M \to \mathbb{R}$ on M, and by M_p the tangent space to M at the point $p \in M$. An element X_p of M_p is called a tangent vector at p.

Let $u^1, \ldots, u^n$ be local coordinates in a neighborhood $U \subset M$ of p. Then the coordinate derivatives $\partial/\partial u^1|_p, \ldots, \partial/\partial u^1|_p, \ldots, \partial/\partial u^n|_p$ form a basis of M_p such that each tangent vector X_p can be expressed as a linear combination of these derivatives.

A differentiable contravariant vector field on M is an assignment of a tangent vector X_p to each point p of M such that for all $f \in C$ the function Xf given by $(Xf)(p) = X_p f$ is differentiable. On U the field X can be expressed by $X = X^k(u)\partial u^k$ with $X^k(u) = X(u^k)$ and $1 \leq k \leq n$. (We adopt the summation convention whereby a repeated index implies summation over all values of that index.)

The space D of all contravariant C^∞-vector fields on M is a real Lie algebra with Lie product defined by

$$[X, Y]f = X(Yf) - Y(Xf) \quad \text{for } X,\ Y \in D \text{ and } f \in C.$$

A covariant vector field or 1-*form* μ on M is a linear homogeneous mapping from the space D into the space of functions C. The value of μ at X is denoted by $\mu(X)$. Each function $f \in C$ defines the 1-form df by $df(X) = Xf$. With respect to the local coordinates $u^1, \ldots, u^n$ on $U \subset M$ we have $Xf = X^k(u)(\partial/\partial u^k)f$. df is called the *differential or gradient of* f. Furthermore we have $du^k(\partial/\partial u^j) = \partial u^k/\partial u^j = \delta^k_j$. Any 1-form μ on U may be expressed uniquely in the form

$$\mu = \mu_k(u)\, du^k \quad \text{with} \quad \mu_k(u) = \mu\left(\frac{\partial}{\partial u^k}\right)$$

A form field of degree r *or* r*-form on* M is a r-linear alternating map α from the r-fold product $D \times \cdots \times D$ into C, i.e., the function

$$\alpha: (X_1, \ldots, X_r) \mapsto \alpha(X_1, \ldots, X_r) \quad \text{with } X_1, \ldots, X_r \in D$$

is linear in each X_j and skew symmetric in the arguments.

The values $\alpha_{k_1 \ldots k_r} = \alpha(\partial/\partial u^{k_1}, \ldots, \partial/\partial u^{k_r})$ where $1 \leq k_j \leq n$, $1 \leq j \leq r$ with respect to the basis field $\partial/\partial u^1, \ldots, \partial/\partial u^n$ on $U \subset M$ are called the (covariant) components of α.

The wedge product or exterior product $\alpha \wedge \beta$ of the r-form α and s-form β is defined by

$$\alpha \wedge \beta(X_1, \ldots, X_{r+s}) = \frac{1}{r!\,s!} \sum_\pi \operatorname{sgn}(\pi)\, \alpha(X_{\pi 1}, \ldots, X_{\pi r}) - \beta(X_{\pi(r+1)}, \ldots, X_{\pi(r+s)})$$

for $X_1, \ldots, X_{r+s} \in D$ and where the sum has to be taken over all permutations π of $1, \ldots, r+s$. $\alpha \wedge \beta$ is a form field of degree $r+s$. In terms of local coordinates on U an r-form α can be expanded as

$$\alpha = (1/r!)\, \alpha_{k_1 \cdots k_r}\, du^{k_1} \wedge \cdots \wedge du^{k_r}.$$

The components of $\alpha \wedge \beta$ are given by

$$(\alpha \wedge \beta)_{k_1 \cdots k_{r+s}} = \alpha_{[k_1 \cdots k_r} \beta_{k_{r+1} \cdots k_{r+s}]},$$

where [ ] denotes the antisymmetrization operator used in physics. $\alpha_{k_1 \cdots k_r}$ and $\beta_{k_1 \cdots k_s}$ are the components of α and β. An r-form α can be contracted with a fixed vector field X to give an $(r-1)$-form $i(X)\alpha$ defined by the *interior multiplication* $i(X)$ with

$$(i(X)\,\alpha)(X_2, \ldots, X_r) = \alpha(X, X_2, \ldots, X_r).$$

$i(X)\,\alpha$ is obtained by inserting into the function α of r vector variables the fixed vector X as the first argument. A straightforward calculation gives

$$(i(X)\,\alpha)_{k_2 \cdots k_r} = (i(X)\,\alpha)\left(\frac{\partial}{\partial u^{k_2}}, \ldots, \frac{\partial}{\partial u^{k_r}}\right) = X^k \alpha_{k k_2 \cdots k_r}$$

with respect to the coordinate system $u^1, \ldots, u^n$ on $U \subset M$.

The interior multiplication is an antiderivation in the sense that

$$i(X)(\alpha \wedge \beta) = (i(X)\,\alpha) \wedge \beta + (-1)^r \alpha \wedge (i(X)\,\beta),$$

where α is an r-form and β is a s-form.

Locally the 1-form df of a function f is given by $df = \partial f/\partial u^k [du^k]$. The operator d sends 0-forms (functions) into 1-forms. This operation can be extended to form fields of arbitrary degree. For a given r-form $\alpha = (1/r!)\, \alpha_{k_1 \cdots k_r}\, du^{k_1} \wedge \cdots \wedge du^{k_r}$ we define the *exterior derivative* $d\alpha$ by ($r \geq 1$)

$$d\alpha = (1/r!)\, d\alpha_{k_1 \cdots k_r} \wedge du^{k_1} \wedge \cdots \wedge du^{k_r};$$

$d\alpha$ is a $(r+1)$-form. The components are given by

$$(d\alpha)_{k_1 \cdots k_{r+1}} = (-1)^r (r+1)\, \alpha_{[k_1 \cdots k_r, k_{r+1}]}.$$

$\alpha_{[k_1, \ldots, k_{r+1}]}$ denotes the antisymmetrization of the derivative $(\partial/\partial u^{k_{r+1}})\, \alpha_{k_1 \cdots k_r}$. One shows that $d\alpha$ is independent of the coordinate system. The exterior derivative is a $\mathbb{R}$-linear map with

$$d(d\alpha) = 0, \quad \text{and} \quad d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^r \alpha \wedge d\beta,$$

where α is a form of degree r and β a form of degree s.

Let $\phi: M \to N$, $p \mapsto \phi(p)$, be a differentiable transformation from the manifold M into the manifold N. If f is a function on N, the mapping ϕ defines the function $\phi^* f$ on M as the function whose value at the point $p \in M$ is the value of f at $\phi(p)$, i.e.,

$$\phi^* f(p) = f(\phi(p)).$$

The mapping ϕ induces a linear mapping of the tangent space M_p into $N_{\phi(p)}$ defined by

$$(\phi_* X_p) f\big|_{\phi(p)} = X_p(\phi^* f)\big|_p \text{ with } X_p \in M_p, \quad f \in C^\infty(N).$$

Introducing local coordinates $u^1, \ldots, u^n$ around p in the manifold M of dimension n and $v^1, \ldots, v^l$ around $\phi(p)$ in the manifold N of dimension l we have for $X_p = X^k (\partial/\partial u^k)\big|_p$,

$$\phi_* X_p = Y^j \frac{\partial}{\partial v^j}\bigg|_{\phi(p)} \text{ with } Y^j = \frac{\partial(\phi^* v^j)}{\partial u^i}\bigg|_p X^i.$$

The transformation ϕ induces *tranformations of forms* in the following way: Let α be a r-form on N. We can *pull back* this form from N to M by the definition

$$(\phi^* \alpha)(X_1, \ldots, X_r)\big|_p = \alpha(\phi_* X_1, \ldots, \phi_* X_r)\big|_{\phi(p)},$$

for $X_i \in M_p$ $(i = 1, \ldots, r)$. With respect to the coordinate systems one obtains

$$\phi^* \alpha = (1/r!)\, \alpha_{k_1 \cdots k_r} \frac{\partial(\phi^* v^{k_1})}{\partial u^{j_1}} \cdots \frac{\partial(\phi^* v^{k_r})}{\partial u^{j_r}}\, du^{j_1} \wedge \cdots \wedge du^{j_r}$$

The pull back operation for forms commutes with the exterior derivative and the operations addition and exterior multiplication. For two transformations ϕ and ψ we have $(\phi \circ \psi)^* = \psi^* \circ \phi^*$.

A curve $t \mapsto \phi_t$ on M is a map of an interval of the real line $\mathbb{R}$ into M. A differentiable curve $t \mapsto \phi_t$ on M is said to be an integral curve of a contravariant vector field X on M with initial point p if

$$\frac{d}{dt} f(\phi_t(p)) = (Xf)(\phi_t(p)), \quad \phi_0(p) = p$$

for all $f \in C^\infty(M)$. The vector field X is said to be *complete* if, for each $p \in M$, X has an integral curve $t \mapsto \phi_t$ with initial point p and parameter domain $-\infty < t < \infty$.

The *Lie derivative* $L_X \alpha$ of an r-form α with respect to X is defined by the derivative of α along the integral curve $t \mapsto \phi_t$ of X, i.e.,

$$L_X \alpha = \lim_{t \to 0} (\phi_t^* \alpha - \alpha)/t.$$

The three operators' interior product $i(X)$, exterior derivative d, and the Lie operator L_X are related by

$$L_X = i(X)d + di(X).$$

This formula proves to be very useful for applications.

3. SYMPLECTIC MANIFOLDS [7, 9, 13, 15, 16]

Let M be a real manifold of even dimension $2n$. A *symplectic form on* M is a 2-form Ω satisfying:

(a) $d\Omega = 0$, i.e., Ω is closed,

(b) Ω is nondegenerate, i.e., for each $p \in M$, $\Omega(X_p, Y_p)\big|_p = 0$ for all $X_p \in M_p$ only if $Y_p \in M_p$ is zero.

Ω defines a nondegenerate skew-symmetric covariant tensor field of degree two. The pair (M, Ω) is called a *symplectic manifold*.

Theorem 1. (*Darboux*)[15,18]: Let (M, Ω) be a $2n$-dimensional symplectic manifold. Then, for each point of M there exist an open set $U \subset M$ containing the point and a local coordinate system $x = (x^k) = (q^a, p_a) = (q^1, \ldots, q^n, p_1, \ldots, p_n)$ on U such that $(k = 1, \ldots, 2n;\ a = 1, \ldots, n)$

$$\Omega = \tfrac{1}{2} \Omega_{kj}\, dx^k \wedge dx^j = dp_a \wedge dq^a$$

with

$$(\Omega_{kj}) = \begin{pmatrix} 0 & -\delta_{ab} \\ \delta_{ab} & 0 \end{pmatrix}; \quad a, b = 1, \ldots, n$$

on this open set. $x=(k^k)$ is called a *canonical coordinate system*.

For each $f\in C^\infty(M)$ the differential df defines a covariant vector field. By using Ω as a "metric" tensor we may pass to the contravariant (canonical) vector field which we denote by X_f uniquely defined by[15,17]

$$[i(X_f)\Omega](Y)=\Omega(X_f, Y)=-df(Y)$$

for all contravariant vector fields Y on M. By using the fact that $i(X_f)$ is an antiderivation, the vector field $X_f=Q^a(\partial/\partial q^a)+P^a(\partial/\partial p_a)$ with respect to canonical coordinates $(x^k)=(q^a, p_a)$ can be written in the form

$$X_f=\frac{\partial f}{\partial p_a}\frac{\partial}{\partial q^a}-\frac{\partial f}{\partial q^a}\frac{\partial}{\partial p_a},$$

$$-df=-\left(\frac{\partial f}{\partial q^a}\,dq^a+\frac{\partial f}{\partial p_a}\,dp_a\right)=i(X_f)\,\Omega=i(X_f)\,dp_a\wedge dq^a$$

$$=[i(X_f)\,dp_a]\,dq^a-[i(X_f)\,dq^a]\,dp_a$$

$$=(X_f p_a)\,dq^a-(X_f q^a)\,dp_a=P^a\,dq^a-Q^a\,dp_a.$$

A transformation $\phi: M\to M$ which leaves the 2-form Ω invariant, $\phi^*\Omega=\Omega$, is said to be *symplectic*. Since

$$L_{X_f}\Omega=d(i(X_f)\,\Omega)+i(X_f)\,d\Omega=-d(df)=0$$

the integral curve $t\mapsto\phi_t$ of the vector field X_f preserves Ω, i.e., $\phi_t^*\Omega=\Omega$.

The *Poisson bracket* $\{f, g\}$ of two functions $f,g\in C^\infty(M)$ is defined by

$$\{f, g\}=X_g f.$$

In terms of canonical coordinates we have

$$\{f, g\}(q, p)=\frac{\partial f}{\partial q^a}\cdot\frac{\partial g}{\partial p_a}-\frac{\partial f}{\partial p_a}\cdot\frac{\partial g}{\partial q^a}.$$

The space of functions $C^\infty(M)$ equipped with the Poisson bracket defines a Lie algebra.

4. EVOLUTION SPACES[7, 13]

Let (M, Ω) be a $2n$-dimensional symplectic manifold. We consider the direct product

$$W=M\times\mathbb{R}$$

which is a $(2n+1)$-dimensional manifold locally described by a system of coordinate functions

$$(x, t)=(q^a, p_a, t).$$

Let $f: M\times\mathbb{R}\to\mathbb{R}$ be a C^∞-function. For any $t\in\mathbb{R}$ define the function

$$f_t: M\to\mathbb{R} \text{ by } f_t(x)=f(x, t).$$

Then the Poisson bracket $[f, g]$ of two functions $f, g: W\to\mathbb{R}$ is given by

$$[f, g](x, t)=\{f_t, g_t\}(x).$$

The space of functions $C^\infty(W)$ defines an infinite dimensional Lie algebra with Poisson bracket as Lie bracket.

Now let $h: W\to\mathbb{R}$ with $dh\neq 0$ be a C^∞-function on W called Hamiltonian. Then a closed 2-form Ω_h on W is defined by

$$\Omega_h=\pi^*\Omega-dh\wedge dt$$

where $\pi: W\to M$ is the projection from W onto M and $\pi^*\Omega$ denotes the pull back of Ω. Ω_h has rank $2n$ such that the restriction of Ω_h,

$$\Omega_h|M_t \text{ with } M_t=M\times\{t\} \text{ for all } t\in\mathbb{R},$$

is nondegenerate. (W, Ω_h) *is said to be an evolution space.*

Ω_h is degenerate on W and we shall show that there exists a vector field Z on W such that

$$(i(Z))\,\Omega_h(Y)=\Omega_h(Z, Y)=0$$

for all contravariant vector fields Y on W. A vector field Z with this property can be defined in a natural way: Let $h_t: M\to\mathbb{R}$ be the function with $h_t(x)=h(x, t)$. Then X_{h_t} is the canonical vector field on M given by

$$i(X_{h_t})\,\Omega=-dh_t.$$

Then X_h with $X_h(x, t)=(X_{h_t}(x), 0)$ defines a vector field on W.

Theorem 2[13,15,16]: Let $h: W\to\mathbb{R}$ be a Hamiltonian with $dh\neq 0$ and define the vector field Z_h on W by

$$Z_h=X_h+\frac{\partial}{\partial t}.$$

Then

$$i(Z_h)\,\Omega_h=0.$$

In terms of local coordinates Z_h is expressed by

$$Z_h=\frac{\partial h}{\partial p_a}\cdot\frac{\partial}{\partial q^a}-\frac{\partial h}{\partial q^a}\cdot\frac{\partial}{\partial p_a}+\frac{\partial}{\partial t}.$$

The integral curve of Z_h through the point $(q_0, p_0, 0)\in W$ is given by

$$t\mapsto(q(t), p(t), t) \text{ with } (q(0), p(0), 0)=(q_0, p_0, 0)$$

where

$$\dot q^a(t)=\frac{\partial h}{\partial p_a}(q(t), p(t), t),\quad \dot p_a(t)=-\frac{\partial h}{\partial q^a}(q(t), p(t), t).$$

Since $i(Z_h)\,\Omega_h=0$ the Lie derivative of Ω_h is zero $(d\Omega_h=0)$,

$$L_{Z_h}\Omega_h=i(Z_h)\,d\Omega_h+d(i(Z_h)\,\Omega_h)=0,$$

such that the integral curve of Z_h preserves the 2-form Ω_h.

5. GROUP ACTIONS

We shall prove a basic theorem concerning group actions which are generated by a system of constants of motion $I: W\to\mathbb{R}$ defined by the solutions of the equation

$$Z_h I=[I, h]+\frac{\partial}{\partial t}I=\frac{dI}{dt}=0.$$

The constants of motion define a Lie algebra with the Poisson bracket as Lie bracket. In the following we consider a k-dimensional Lie algebra of constants of motion given on a $(2n+1)$-dimensional evolution space W where $k\geq 2n$.

Let $I^1, \ldots, I^k$ be a basis of the Lie algebra. We associate vector fields X_{I^j} $(j=1, \ldots, k)$ on W in the following way: For I_t^j defined by $I_t^j(x)=I^j(x, t)$, with $(x, t)\in M\times\mathbb{R}$ we have the vector fields $X_{I_t^j}$ with

$$i(X_{I_t^j})\,\Omega = -\,dI_t^j.$$

Then the vector fields X_{I^j} with $X_{I^j}(x,t) = (X_{I_t^j}(x), 0)$ are defined on W. The map $I^j \mapsto X_{I^j}$ generates a Lie algebra homomorphism.

We are interested in Lie algebras of complete vector fields which generate Lie groups of transformations. We shall say that a *Lie group G acts on a manifold N* if there exists a homomorphism

$$\phi : G \to \mathrm{Diff}N,\ g \mapsto \phi_g$$

from the group G into the group Diff$N$ of diffeomorphisms ϕ_g on N. Sometimes G is called a *tranformation group*.

Now let h with $dh \neq 0$ be a (Hamiltonian) function defined on N. If the transformations of the group G do not leave invariant the function h, the Lie group G is said to be a *noninvariance group with respect to h*. The corresponding Lie algebra is called a *spectrum generating algebra*.

In the following we shall construct spectrum generating Lie algebras of constants of motion and the corresponding Lie groups acting on evolution spaces and orbit spaces.

Theorem 3: Let $(W = M\times\mathbb{R}, \Omega_h)$ be a $(2n+1)$-dimensional evolution space. We assume that the vector field $Z_h = X_h + \partial/\partial t$ is complete. Let L be a finite dimensional Lie subalgebra of $C^\infty(W)$ with a basis $I^1, \ldots, I^k$ of L $(k \geq 2n)$ such that

(i) $$[I^j, h] + \frac{\partial I^j}{\partial t} = 0,\quad j = 1, \ldots, k,$$

(ii) $dI^1, \ldots, dI^k$ span a $2n$-dimensional vector space at each point of $W = M\times\mathbb{R}$,

(iii) each I^j generates a complete vector field X_{I^j} on W. Then

(a) a connected Lie group G with Lie algebra isomorphic to L acts on the evolution space. The 2-form Ω_h is left invariant by the group actions;

(b) the group action maps an orbit onto an orbit of Z_h.

Proof: (a) The map $I^j \mapsto X_{I^j}$ defines a Lie algebra homomorphism. From a result of Palais[19] it follows that for a given Lie algebra of contravariant vector fields X_{I^j} there exists a connected Lie group G which acts on the evolution space. Since $L_{X_{I^j}}\Omega_h = 0$ for all X_{I^j} the 2-form Ω_h is left invariant by the corresponding group transformations.

(b) It can be shown that[14] ($[Z_h, X_{I^j}]$ Lie bracket of vector fields)

$$[Z_h, X_{I^j}] = X_{[I^j, h] + \partial I^j/\partial t}.$$

From condition (a) it follows that $[Z_h, X_{I^j}] = 0$. Therefore the group actions commute with the orbits. G maps a trajectory of Z_h onto a trajectory of Z_h.

We show that there are intrinsic relations between symplectic transformations on the manifold (M, Ω) and transformation on W. From $X_{I^j}t = 0$ it follows that the group actions on W preserve the time t. Therefore the manifold $M_t = M\times\{t\} \subset W$ with $t \in \mathbb{R}$ is left invariant. On the other hand, group actions on W are induced by symplectic transformations on M, i.e., there is a Lie algebra homomorphism from the Lie algebra of functions $I_t^j : M \to \mathbb{R}$ into the Lie algebra of contravariant vector fields $X_{I_t^j}$ defined on M with $i(X_{I_t^j})\,\Omega = -\,dI_t^j$ for all $t \in \mathbb{R}$. Assuming that all vector fields are complete there exists a connected Lie group G acting on M. Since $L_{X_{I_t^j}}\Omega = 0$ for each $t \in \mathbb{R}$ the group G acts on M by symplectic transformations which may be given by

$$\phi_g^{(t)} : M \to M,\quad x \mapsto \phi_g^{(t)}(x),\quad \text{for all } g \in G.$$

This transformation on M can be extended to a transformation on W,

$$\Psi_g : W \to W \quad \text{with} \quad \Psi_g(x,t) = (\phi_g^{(t)}(x), t).$$

The space $M_t \subset W$ is left invariant by this action.

On the other hand we have $Z_h t = 1$, i.e., the time is not left invariant with respect to the integral curves of the complete vector field Z_h. Therefore any trajectory of Z_h is transversal to the manifold M_t for $t \in \mathbb{R}$. The trajectories commute with the action of the Lie group G which leaves each M_t invariant. From this it follows that the integral curves of Z_h interpreted as a 1-parameter group of transformations on W connect the manifolds M_t for different values of t.

It is interesting to note that for $t = 0$ the problem has been discussed by Dirac[20] in the case of special relativistic dynamics. In this connection we have transformations induced by the inhomogeneous Lorentz group. The preceding results show that even in the general theory it does make sense to discuss the problem for $t = 0$.

6. ORBIT SPACES

Let $(W = M\times\mathbb{R}, \Omega_h)$ be a $(2n+1)$-dimensional evolution space. We assume that Z_h is complete on W. Let $\overline{W}$ be the set of orbits of Z_h and $\sigma : W \to \overline{W}$ the map assigning to each point of W the trajectory of Z_h through the point $w \in W$. Since the projection of $W = M\times\mathbb{R}$ onto $\mathbb{R}$ maps $Z_h = X_h + \partial/\partial t$ onto $\partial/\partial t$ every trajectory passes through one and only one point of $M_t = M\times\{t\}$ for any given $t \in \mathbb{R}$. Therefore there exists a manifold structure on $\overline{W}$ such that for any $t \in \mathbb{R}$ the restriction

$$\sigma|_{M_t} : M_t \to \overline{W}$$

is a diffeomorphism. *$\overline{W}$ is called the orbit space of the evolution space W.*

One shows that the 2-form Ω_h given on W induces a symplectic 2-form $\overline{\Omega}$ on $\overline{W}$ defined by ($\sigma : W \to \overline{W}$)[9,13,25]

$$\overline{\Omega}(\sigma_* X_I, \sigma_* X_J)(\sigma(w)) = \Omega_h(X_I, X_J)(w)$$

for $w \in W$ and all functions I and J with $Z_h I = Z_h J = 0$. Therefore $(\overline{W}, \overline{\Omega})$ defines a symplectic manifold where $\sigma^*\overline{\Omega} = \Omega_h$.

Theorem 3: Let the conditions of Theorem 2 hold. Furthermore we assume that the set of orbits of Z_h defines a C^∞-manifold $\overline{W}$. Then

(a) The Lie group G with Lie algebra isomorphic to L acting on the evolution space W induces a unique group action on the orbit space $\overline{W}$.

(b) The group G acts on $\bar{W}$ by symplectic transformations.

Proof: (a) The group G maps a trajectory of Z_h onto a trajectory of Z_h. From a theorem of Palais[19] it follows that there exists a unique group action

$$\bar{\phi}_g : \bar{W} \to \bar{W}, \quad \bar{w} \mapsto \bar{\phi}_g(\bar{w})$$

such that

$$\sigma(\phi_g(w)) = \bar{\phi}_g(\sigma(w)) \quad \text{or} \quad \sigma \circ \phi_g = \bar{\phi}_g \circ \sigma$$

with $w \in W$ and the projection $\sigma : W \to \bar{W}$ for all $g \in G$.

(b) We have to show that $\bar{\Omega}$ leaves invariant with respect to the group action $\bar{\phi}_g$, i.e., $\bar{\phi}_g^* \bar{\Omega} = \bar{\Omega}$. Since $\phi_g^* \Omega_h = \Omega_h$ and $\Omega_h = \sigma^* \bar{\Omega}$ we have

$$\sigma^* \bar{\Omega} = \phi_g^*(\sigma^* \bar{\Omega}) = (\sigma \circ \phi_g)^* \bar{\Omega} = (\bar{\phi}_g \circ \sigma)^* \bar{\Omega} = \sigma^*(\bar{\phi}_g^* \bar{\Omega})$$

such that $\bar{\Omega} = \bar{\phi}_g^* \bar{\Omega}$, where we have used $\sigma \circ \phi_g = \bar{\phi}_g \circ \sigma$ and $\phi_g^* \circ \sigma^* = (\sigma \circ \phi_g)^*$.

7. PHYSICAL APPLICATIONS

The preceding theory shall be applied to physical problems which can be described by Lie algebras of constants of motion I^j with $Z_h I^j = 0$ for a given Hamiltonian function h.

The space of constants of motion can be identified with the space of functions $C^\infty(\bar{W})$ on the symplectic orbit space $(\bar{W}, \bar{\Omega})$. $C^\infty(\bar{W})$ defines a Lie algebra under Poisson bracket on $\bar{W}$. In the following we study finite dimensional Lie algebras of functions generating Lie group actions on orbit spaces.

For special problems we shall give solutions I^j $(j = 1, \ldots, 2n)$ of

$$Z_h I^j = 0 \quad \text{or} \quad \frac{\partial I^j}{\partial t} = -X_h I^j,$$

with initial conditions $I^j(x, 0) = x^j$ calculated in terms of Lie series,[21] i.e.,

$$I^j(x,t) = \exp(-tX) \cdot x^j = \left(\sum_{m=0}^{\infty} [(-t)^m/m!] X_h^m \right) \cdot x^j$$

$$= x^j - t[x^j, h] + (t^2/2!)[[x^j, h], h] - \cdots .$$

Furthermore we shall construct the map

$$\sigma : W \to \bar{W} \quad \text{with} \quad \Omega_h = \sigma^* \bar{\Omega}.$$

For a given system of coordinate functions I^j defined in a neighborhood U of point $w \in W$ there exist coordinate functions $\bar{I}^j$ defined on $\sigma(U)$ such that for all $w \in U$ we have

$$\bar{I}^j(\sigma(w)) = I^j(w), \quad \bar{I}^j \circ \sigma = I^j,$$

respectively.

7A. THE NONRELATIVISTIC FREE PARTICLE

In the case of a free nonrelativistic particle with time independent Hamiltonian

$$h : W \to \mathbb{R}, \quad h(q,p,t) = p^2/2m, \quad p \in \mathbb{R}^3 \setminus \{0\},$$

and the vector field Z_h is given by $(a = 1, 2, 3,)$

$$Z_h = \frac{\partial}{\partial t} + \sum_{a=1}^{3} \frac{p_a}{m} \frac{\partial}{\partial q^a}.$$

The solutions $I^j : W \to \mathbb{R}$ of $Z_h I^j = 0$ with $I^j(x, 0) = x^j$ are given by

$$I^a(q,p,t) = \exp(-tX) \cdot q^a - t[q^a, h] + (t^2/2!)[[q^a, h], h] - \cdots$$

$$= q^a - \frac{p_a}{m} t,$$

$$I^{a+3}(q,p,t) = p_a.$$

The differentials dI^j at each point of W have rank 6. There exist coordinate functions Q_a and P_a on $\bar{W}$ such that $(\sigma : W \to \bar{W})$

$$I^a = Q_a \circ \sigma, \; I^{a+3} = P_a \circ \sigma.$$

$\bar{\Omega}$ is given by

$$\bar{\Omega} = \sum_{a=1}^{3} dP_a \wedge dQ_a$$

since

$$\sigma^* \bar{\Omega} = \sigma^* \left(\sum_{a=1}^{3} dP_a \wedge dQ_a \right) = \sum_{a=1}^{3} d(P_a \circ \sigma) \wedge d(Q_a \circ \sigma)$$

$$= \sum_{a=1}^{3} dp_a \wedge d(q^a - (p_a/m)\, t) = \sum_{a=1}^{3} dp_a \wedge dq^a - dh \wedge dt = \Omega_h.$$

Therefore we are given a system of canonical coordinates on the symplectic manifold $(\bar{W}, \bar{\Omega})$.

One immediately verifies that the functions P_a, G_a, J_a, and h with

$$G_a = -mQ_a, \quad J_a = \sum_{b,c=1}^{3} \epsilon_{abc} Q_b P_c, \quad P_a \text{ and } h = p^2/2m$$

define a realization of the Poisson algebra of the Galilean group[21,22] if one uses the Poisson bracket defined on the orbit space as Lie bracket. The functions G_a, J_a, P_a, and h generate complete canonical vector fields on $\bar{W}$ which have rank 6 since the differentials of the functions have the same rank and $\bar{\Omega}$ is nondegenerate. Therefore the group generated by the canonical vector fields is globally defined and acts on $\bar{W}$ via canonical transformations. The Galilean group can be interpreted as a noninvariance group of the free nonrelativistic particle.

7B. THE HARMONIC OSCILLATOR

In the case of the three-dimensional harmonic oscillator with Hamiltonian

$$h(q,p,t) = p^2/2 + q^2/2, \quad q, p \in \mathbb{R}^3 \setminus \{0\},$$

coordinate functions Q_a, P_a on the orbit space $\bar{W}$ are given by

$$Q_a \circ \sigma(q,p,t) = q^a \cos t - p_a \sin t,$$

$$P_a \circ \sigma(q,p,t) = q^a \sin t + p_a \cos t.$$

If we introduce the complex variables

$$z_a = (1/\sqrt{2})(Q_a + iP_a), \quad \bar{z}_a = (1/\sqrt{2})(Q_a - iP_a)$$

we have

$$\sum_{a=1}^{3} dP_a \wedge dQ_a = i \sum_{a=1}^{3} d\bar{z}_a \wedge dz_a.$$

With respect to the corresponding Poisson bracket given on $\bar{W}$ the functions

$$A_{ab} = \bar{z}_a z_b, \quad A_{a4} = i z_a \left(\sum_{b=1}^{3} \bar{z}_b z_b + 1 \right)^{1/2},$$

$$A_{4a} = -\bar{A}_{a4}, \quad A_{44} = -\sum_{a=1}^{3} \bar{z}_a z_a - 1$$

define a realization of the Lie algebra of the group $U(3,1)$[23] which is a noninvariance group of the three-dimensional harmonic oscillator.

7C. A CHARGED PARTICLE IN A CONSTANT ELECTROMAGNETIC FIELD

For a particle with mass m and negative charge $-e$ moving in a constant magnetic field $B = (B_1, B_2, B_3)$ the vector potential A may be given by ($B_3 = |B|$, $q \in \mathbb{R}^3$)

$$A = \tfrac{1}{2}[B, q] = (-\tfrac{1}{2}B_3 q_2, \tfrac{1}{2}B_3 q_1, 0).$$

Then the Hamiltonian h is defined by ($\omega = eB_3/mc$)

$$h(q,p,t) = (1/2m)(p_1 - \tfrac{1}{2}m\omega q_2)^2 + (1/2m)(p_2 + \tfrac{1}{2}m\omega q_1)^2 + (1/2m)p_3^2,$$

with

$$\pi_1 = p_1 - \tfrac{1}{2}m\omega q_2, \quad \pi_2 = p_2 + \tfrac{1}{2}m\omega q_1.$$

six constants of motion are

$$I^1(q,p,t) = x_0 - (\pi_1/m\omega)\sin\omega t + (\pi_2/m\omega)\cos\omega t,$$

$$I^2(q,p,t) = y_0 - (\pi_1/m\omega)\cos\omega t - (\pi_2/m\omega)\sin\omega t,$$

$$I^3(q,p,t) = q_3 - (p_3/m)\,t,$$

$$I^4(q,p,t) = \tfrac{1}{2}m\omega y_0 + (\pi_1/2)\cos\omega t + (\pi_2/2)\sin\omega t,$$

$$I^5(q,p,t) = -\tfrac{1}{2}m\omega x_0 - (\pi_1/2)\sin\omega t + (\pi_2/2)\cos\omega t,$$

$$I^6(q,p,t) = p_3,$$

where

$$x_0 = \tfrac{1}{2}q_1 - p_2/m\omega, \quad y_0 = \tfrac{1}{2}q_2 + p_1/m\omega.$$

By choosing these constants of motion as a coordinate system one immediately sees that the particle moves on a cylinder with constant velocity in the direction of q_3.

On the orbit space $\overline{W}$ a system of canonical coordinates is given by the functions ($\sigma : W \to \overline{W}$)

$$Q_1, Q_2, Q_3, P_1, P_2, P_3$$

with

$$Q_1 \circ \sigma = I^1 - x_0, \quad Q_2 \circ \sigma = x_0, \quad Q_3 \circ \sigma = I^3$$

$$P_1 \circ \sigma = 2(I^4 - \tfrac{1}{2}m\omega y_0), \quad P_2 \circ \sigma = m\omega y_0, \quad P_3 \circ \sigma = I^6.$$

With respect to these coordinate functions one verifies

$$\sigma^* \overline{\Omega} = \Omega_h.$$

If there is an additional electric field then in a similar way one can calculate the constants of motion and construct the orbit space.

7D. TIME-DEPENDENT QUADRATIC SYSTEMS

We consider a Hamiltonian h with $[(x^k) = (q^a, p_a)]$

$$h(x,t) = \sum_{k,j=1}^{2n} A_{kj}(t)\, x^k x^j + \sum_{k=1}^{2n} B_k(t)\, x^k$$

and assume that the problem can be described by $2n$ constants of motion I^k with

$$I^k(x,t) = \sum_{j=1}^{2n} a_{kj}(t)\, x^j + b_k(t)$$

which are solutions of

$$Z_h I^k = [I^k, h] + \frac{\partial I^k}{\partial t} = 0 \quad \text{with} \quad I^k(x,0) = x^k.$$

Then the coefficients a_{kj} and b_k must satisfy the differential equations

$$\dot{b}_k = \frac{db_k}{dt} = \sum_{j,l=1}^{2n} a_{kj}\,\Omega_{jl}\,B_l$$

and

$$\dot{a}_{kj} = \frac{da_{kj}}{dt} = \sum_{l,m=1}^{2n} a_{kl}\,\Omega_{lm}(A_{mj} + A_{jm})$$

where

$$(\Omega_{kj}) = \begin{pmatrix} 0 & -\delta_{ab} \\ \delta_{ab} & 0 \end{pmatrix} \quad \text{and} \quad [x^k, x^j] = -\Omega_{kj}.$$

Furthermore

$$[I^k, I^j](x,t) = [I^k, I^j](x,0) = [x^k, x^j] = -\Omega_{kj}$$

such that

$$\sum_{l,m=1}^{2n} a_{kl}\,\Omega_{lm}\,a_{jm} = \Omega_{kj}.$$

Now introducing the complex variables

$$z_a = (1/\sqrt{2})[I^a(x,t) + iI^{a+n}(x,t)],$$

$$\bar{z}_a = (1/\sqrt{2})[I^a(x,t) - iI^{a+n}(x,t)],$$

we can define a realization of the Lie algebra $U(n,1)$ by setting ($1 \le a, b \le n$)

$$A_{ab} = \bar{z}_a z_b, \quad A_{a\,n+1} = iz_a\left(\sum_{b=1}^{n} \bar{z}_b z_b + 1\right)^{1/2},$$

$$A_{n+1\,a} = -\bar{A}_{a\,n+1}, \quad A_{n+1\,n+1} = -1 - \sum_{b=1}^{n} \bar{z}_b z_b.$$

We shall apply this result to the problem of a damped oscillator: The problem may be described by the equation

$$m\ddot{q} + \beta\dot{q} + kq = 0 \quad (q \in \mathbb{R}^3).$$

The corresponding Hamiltonian is defined by[24]

$$h(q,p,t) = (p^2/2m)\cdot\exp[-(\beta/m)\,t] + (k/2)\,q^2\cdot\exp[(\beta/m)\,t].$$

Using the equations given above we have for instance ($d = 1, 2, 3$)

$$\dot{a}_{dd} = k\exp[(\beta/m)\,t]\circ a_{d\,3+d}$$

and

$$\dot{a}_{d\,3+d} = -(1/m)\exp[-(\beta/m)\,t]\circ a_{dd}$$

such that

$$\ddot{a}_{dd} - (\beta/m)\,\dot{a}_{dd} + (k/m)\,a_{dd} = 0.$$

The general solutions lead to the result

$$[\Delta = (\beta^2 - 4km)^{1/2},\ I^a(x,0) = q^a,\ I^{a+3}(x,0) = p_a]$$

$$I^a(x,t) = (1/2\Delta)\exp(\beta/2m)\,t\ \{(\Delta - \beta)\exp[(\Delta/2m)\,t] + (\beta + \Delta)\exp[-(\Delta/2m)\,t]\}\circ q^a$$

$$-(1/\Delta)\exp[-(\beta/2m)\,t]\{\exp[(\Delta/2m)\,t] - \exp[-(\Delta/2m)\,t]\}\cdot p_a,$$

$$I^{a+3}(x,t) = (km/\Delta)\exp[(\beta/2m)t]\{\exp[(\Delta/2m)t] - \exp[-(\Delta/2m)t]\}\cdot q^a + (1/2\Delta)\exp[-(\beta/2m)t]\{(\beta+\Delta)\exp[(\Delta/2m)t] - (\beta-\Delta)\exp[-(\Delta/2m)t]\}\cdot p_a.$$

The constants of motion generate a Lie algebra of the group U(3, 1) which is a noninvariance group of this problem as in the case of the stationary three-dimensional harmonic oscillator. Finally the constants of motion can be used as a system of canonical coordinates on the orbit space.

[1]M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 (1966).
[2]Y. Dothan, Phys. Rev. D 2, 1974 (1970).
[3]V.I. Man'ko, "Invariants and Symmetries of Dynamical Systems," Proceedings of the Second International Colloquium on Group Theoretical Methods in Physics, Nijmegen, 1973.
[4]M. Perrin, "Time Dependent Invariance Algebras in the Schrödinger Picture," Proceedings of the Third International Colloquim on Group Theoretical Methods in Physics, Marseille, 1974.
[5]P. Renouard, "Variétés symplectiques et quantification," Thèse, Orsay (1969).
[6]D. Simms, Zeitschrift für Naturforschung 28a, 538 (1972).
[7]J. M. Souriau, *Structures des Systèmes Dynamiques* (Dunod, Paris, 1970).
[8]H. Rollnik, Teilchenphysik, BI 706/759 (1971).
[9]M. Andrié and D. Simms, J. Math. Phys. 13, 331 (1972).
[10]R. Arens, J. Math. Phys. 12, 2415 (1971).
[11]H. Bacry, H. Ruegg, and J. Souriau, Commun. Math. Phys. 3, 323 (1966).
[12]J. Moser, Commun. Pure Appl. Math. 23 (1970).
[13]B. Kostant, "Line Bundles and the Prequantized Schrödinger Equation," talk presented at the International Colloquium on New Methods in Physics, Bonn, 1973.
[14]R. Hermann, *Lectures in Mathematical Physics* (Benjamin, New York, 1970) Vol. II; J. Math. Phys. 13, 833 (1972).
[15]R. Abraham, *Foundations of Mechanics* (Benjamin, New York, 1967).
[16]C. Godbillon, *Géométrie Différentielle et Mécanique Analytique* (Hermann, Paris, 1969).
[17]N. J. Hicks, *Notes on Differential Geometry* (Princeton, New Jersey, 1965).
[18]C.W. Kilmister, *Hamiltonian Dynamics* (Longmans Green, N. Y., 1964).
[19]R. S. Palais, Mem. Am. Math. Soc. 22, 95 (1957).
[20]P. A. M. Dirac, Rev. Mod. Phys. 21, 3 (1949).
[21]P. Mittelstaedt, BI Hochschulskripten, 500/500a (1970).
[22]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[23]J. Hajdu, G. Györgyi, and Th. Kahan, in *Théorie des groupes, t.* 2, edited by Th. Kahan (Dunod, Paris, 1971).
[24]H. Goenner, private communication, University of Göttingen.
[25]M. Andrié, BI Hochschulskripten 726, 98 (1972).

Dynamical quantization of the Kepler manifold

Enrico Onofri

Istituto di Fisica dell'Università di Parma, Italy
and Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy
(Received 29 April 1975)

The dynamical quantization of the "Kepler manifold" in any number of degrees of freedom is constructed. The Kepler manifold is the phase space of the regularized Kepler motion and is shown to be an SO(n,2)-homogeneous symplectic manifold, corresponding to an extremely singular orbit in the co-adjoint representation; the quantization is obtained by "approximating" this orbit by more regular ones, which are equivalent to homogeneous bounded domains. The most relevant result is that the usual quantum-mechanical "hydrogen atom" model is recovered in the particular representation introduced by Fock in 1935 [SO(n)-homogeneous integral equation in momentum space].

INTRODUCTION

"Geometric quantization" is generally accepted today by mathematicians as a powerful technique which gives almost all unitary irreducible representations for a wide class of Lie groups.[1,2,3] There has been little interest in the theory among theoretical physicists, however; the reason may be found in the narrow limits of applicability of the theory, i.e., harmonic oscillators,[4] free relativistic and de Sitter particles,[5] and energy levels of the hydrogen atom.[6] The aim of the present paper is to give an example which should be of interest also for theoretical physicists. The problem of applying the theory of geometric quantization to the Kepler problem attracted the attention of several people. Simms[6] considered each energy level separately and showed that the usual level structure could be obtained; a certain deviation from original Kostant theory was however necessary (this will be discussed in Sec. 2). Onofri and Pauri[7] suggested applying the theory to a certain SO(4, 2)-homogeneous symplectic manifold which represents the phase space of the (regularized) Kepler motion with negative energy. It was shown, however,[8] that there does not exist any invariant polarization in this case, and Kostant theory does not apply. Recent developments of the theory allow us to deal with non-invariant polarizations as well[3]; but there does not seem to exist a simple recipe to construct the so-called Kostant—Blattner—Sternberg kernel. In this paper we shall adopt an alternate procedure. We shall show that *a correct quantization of the "Kepler manifold" can be obtained through a limit of quantizable manifolds, admitting an invariant (complex) polarization.* Whether the same result can be obtained by means of the Kostant—Blattner—Sternberg approach is still to be investigated. Although our procedure may appear rather heuristic, it is *a posteriori* justified, since we obtain the usual quantum-mechanical model of the hydrogen atom (for negative energy). In particular we obtain the representation found by Fock[9] in terms of functions defined on a hypersphere in momentum space.

The material is organized as follows: In the first section we study some singular orbits O_{ω_l} with respect to the co-adjoint representation of SO(n, 2). The lowest-dimensional one O_{ω_0} is shown to represent the phase space of Kepler motion with negative energy in $n-1$ dimensions; to make this identification one must first regularize the Hamiltonian flow as shown by Moser[10] [see also Refs. 4 and 11]. In the second section, we construct the quantization of the orbits O_{ω_l}; the corresponding unitary irreducible representation of SO(n, 2) is studied in some detail, in particular the "coherent state" basis is constructed. In the third section we study the limit $l \to 0$. The result is that (i) a natural realization of the carrier space of the representation of SO(n, 2) in this limit is given by functions on the sphere S^{n-1}, and (ii) the energy eigenvectors are solutions of the integral equation

$$\Psi_s(x) = \frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{n/2}}\,(s+\tfrac{1}{2}n-1)\int_{S^{n-1}} \frac{\Psi_S(y)}{\|x-y\|^{n-2}}\,\mathring{y},$$

where $(s+\frac{1}{2}n-1)$ is the eigenvalue of $mk/(-2mH)^{1/2}$. This equation is identical with the usual one as reported in Ref. 12.

From a mathematical point of view, the idea of obtaining unitary representations through an analytic continuation in the invariants goes back to Knapp and Okamoto[13]; it was suggested by the present author for SO(n, 2), with an eye toward the Kepler problem, in Ref. 14. The mathematical ideas are essentially all contained in Harish-Chandra's works; in particular the singular case is considered in Ref. 15. We included a lot of well-known results, the aim being to make the paper readable for a wider group of theoretical physicists.

1. A FAMILY OF SINGULAR SO(n, 2)-orbits

Let us consider the group of linear transformations of R^{n+2} onto itself which leave the pseudo-Euclidean form $\mathbf{g}$ invariant, where $g_{11}=g_{22}=-g_{33}=\cdots=-g_{n+2\ n+2}=1$, and $g_{ij}=0$ for $i\neq j$. We consider only transformations which are connected to the identity. This group is denoted by SO(n, 2); $n=3,4$ correspond to the de Sitter and conformal group, respectively. We shall deal in general with $n\geq 3$. Let $g=\exp X$ be the exponential mapping, X being an element in the Lie algebra $\mathbf{so}(n,2)$.
A basis in $\mathbf{so}(n,2)$ is given by X_{ij}, $\exp(\varphi X_{ij})$ being a special transformation in the (i,j) plane. Of course $X_{ij}=-X_{ji}$ and

$$[X_{ij},X_{hk}]=g_{ih}X_{jk}+g_{jk}X_{ih}-g_{jh}X_{ik}-g_{ik}X_{jh}, \tag{1}$$

or equivalently $[X_{ij}, X_{ik}] = g_{ii} X_{jk}$ (two equal indices), or zero if all indices are different.

It will be useful, in the sequel, to have special symbols for the generators, namely:

$$\left.\begin{aligned} &X_{12} = S, \\ &X_{\mu+2\nu+2} = M_{\mu\nu}, \end{aligned}\right\} \quad \text{maximal compact subalgebra } \mathbf{so}(n) \oplus \mathbf{so}(2)$$
$$X_{1\mu+2} = Z_\mu,$$
$$X_{2\mu+2} = W_\mu, \tag{2}$$

Greek indices run from 1 to n and are "tensorial" with respect to the compact subgroup $SO(n)$.

The dual space of $\mathbf{so}(n,2)$ is denoted by $\mathbf{so}(n,2)^*$. The Killing form $B(X,Y) = (1/2n)\,\mathrm{Tr}(\mathrm{ad}X\,\mathrm{ad}Y)$ is nonsingular; then $X_i^\flat$ provide a basis in $\mathbf{so}(n,2)^*$; here $X^\flat \in \mathbf{so}(n,2)^*$ is defined by $\langle X^\flat, Y\rangle \equiv B(X,Y)$ for every $Y \in \mathbf{so}(n,2)$. Explicitly we have $B(X_{ij}, X_{hk}) = g_{ik}g_{jh} - g_{ih}g_{jk}$. A generic point $\omega \in \mathbf{so}(n,2)^*$ is then given by

$$\omega = s\,S^\flat + \sum_{\mu<\nu} m_{\mu\nu} M_{\mu\nu}^\flat + \sum_\mu (z_\mu Z_\mu^\flat + w_\mu W_\mu^\flat), \tag{3}$$

and we shall simply write $\omega = (s,\ m_{\mu\nu},\ z_\mu,\ w_\mu)$.

The orbits in $\mathbf{so}(n,2)^*$ with respect to the co-adjoint representation constitute the most general model of homogeneous symplectic manifold with respect to the group.[1,2] We shall denote by $\mathcal{O}_\omega$ the orbit through ω. We shall limit ourselves to some "singular" orbits, namely $\mathcal{O}_{\omega_l}$ $(l \geq 0)$ defined by

$$\begin{aligned} &\omega_l = lS^\flat, && (l>0), \\ &\omega_0 = S^\flat + Z_1^\flat + W_2^\flat + M_{12}^\flat, && (l=0). \end{aligned} \tag{4}$$

As submanifolds of $\mathbf{so}(n,2)^*$, these orbits are characterized as follows:

$$\text{(i)}\ \mathcal{O}_{\omega_l} = \Big\{\omega = (s, m_{\mu\nu}, z_\mu, w_\mu)\ \Big|\ s^2 + \sum_{\mu<\nu} m_{\mu\nu}^2 - \sum_\mu (z_\mu^2 + w_\mu^2) = l^2;\ s m_{\mu\nu} = z_\mu w_\nu - z_\nu w_\mu \Big\}\ (l>0).$$

Since

$$\mathrm{ad}^*X(Y^\flat) = [X,Y]^\flat \tag{5}$$

the stability subgroup of ω_l is given by the commutant of S which is $SO(2)\otimes SO(n)$, the maximal compact subgroup. It follows that $\mathcal{O}_{\omega_l}$ is equivalent to the n-dimensional complex bounded domain $\mathcal{D}^{(n)} \subset C^n$ of type IV.[16] Said otherwise, it is possible to parametrize $\mathcal{O}_{\omega_l}$ by complex coordinates $(\zeta_1, \zeta_2, \ldots, \zeta_n)$ in such a way that the action of $SO(n,2)$ on $\mathcal{O}_{\omega_l}$ is given by holomorphic transformations. This will be shown in Sec. 2.

$$\text{(ii)}\ \mathcal{O}_{\omega_0} = \Big\{\omega = (s, m_{\mu\nu}, z_\mu, w_\mu)\ \Big|\ s^2 + \sum_{\mu<\nu} m_{\mu\nu}^2 - \sum_\mu (z_\mu^2 + w_\mu^2) = 0;\ s m_{\mu\nu} = z_\mu w_\nu - z_\nu w_\mu;\ \sum_\mu z_\mu^2 = \sum_\mu w_\mu^2;\ \sum_\mu z_\mu w_\mu = 0 \Big\}.$$

It follows that $\sum_\mu z_\mu^2 = \sum_\mu w_\mu^2 = s^2 = \sum_{\mu<\nu} m_{\mu\nu}^2$ and the global structure of $\mathcal{O}_{\omega_0}$ is given by[7] $\mathcal{O}_{\omega_0} \approx R \times SO(n)/SO(n-2)$. This is the same as the manifold introduced by Moser[10] $T_0^* S^{n-1}$ (cotangent bundle on S^{n-1} with the zero-section deleted) and called the "Kepler manifold" by Souriau.[11]

The identification of $\mathcal{O}_{\omega_0}$ with the phase space of Kepler motion can be seen by introducing a local chart of canonical coordinates. Let $(\mathbf{x}, \mathbf{p})$ be canonical coordinates in $T^*(R^{n-1} - \{0\})$, $H = p^2/2m - k/r$ is the Hamiltonian, where $p^2 = \mathbf{p}\cdot\mathbf{p}$, $r = (\mathbf{x}\cdot\mathbf{x})^{1/2}$. Let $\mathbf{z} = (z_1, \ldots, z_{n-1})$, $\mathbf{w} = (w_1, \ldots, w_{n-1})$, $\rho = (m_{1n}, \ldots, m_{n-1n})$. Then the following transformations are canonical (locally):

$$\begin{aligned} &s = mk/(-2mH)^{1/2}, \\ &\boldsymbol{\sigma} \equiv \mathbf{z} - i\mathbf{w} = \left(r\,\mathbf{p} + i\,\frac{mk\,\mathbf{x}/r - (\mathbf{x}\cdot\mathbf{p})\mathbf{p}}{(-2mH)^{1/2}}\right) \\ &\quad \times \exp\{i(-2mH)^{1/2}\mathbf{x}\cdot\mathbf{p}/mk\}, \\ &\sigma_n \equiv z_n - iw_n = \left(\frac{rp^2 - mk}{(-2mH)^{1/2}} + i\,\mathbf{x}\cdot\mathbf{p}\right) \\ &\quad \times \exp\{i\,(-2mH)^{1/2}\mathbf{x}\cdot\mathbf{p}/mk\}, \\ &m_{\mu\nu} = x_\mu p_\nu - x_\mu p_\mu, \quad \mu,\nu = 1,2,\ldots,n-1, \\ &\rho = \frac{p^2\mathbf{x} - (\mathbf{x}\cdot\mathbf{p})\,\mathbf{p} - mk\,\mathbf{x}/r}{(-2mH)^{1/2}} \quad \text{(Runge–Lenz vector)}. \end{aligned} \tag{6}$$

Let us briefly comment on these formulas. The relation between s and H is such that the rotation angle τ in $(1,2)$-plane coincides with $2\pi t/T$, where t is the time in Hamilton's equations and T is the period of motion. The expression of $m_{\mu\nu}$ and ρ is obvious; that of z_μ and w_μ is not so obvious, yet well-known (Bacry–Gyorgyi parameters[17]). A simple proof is given below.

Let us take for granted that the expression of s and $m_{\mu\nu}$ is correct. We have to determine $(\mathbf{z}, z_n)$ and $(\mathbf{w}, w_n)$ as functions of $\mathbf{x}$ and $\mathbf{p}$. Let us put

$$\begin{aligned} &\mathbf{z} = \phi_1\mathbf{x} + \phi_2\mathbf{p}, \quad z_n = \phi, \\ &w = \psi_1\mathbf{x} + \psi_2\mathbf{p}, \quad w_n = \psi, \end{aligned} \tag{7}$$

where the ϕ's and ψ's may depend only on r, p and $\mathbf{x}\cdot\mathbf{p}$. Let us fix these unknown functions by requiring that $(s,\ m_{\mu\nu},\ z_\mu,\ w_\mu)$ belong to $\mathcal{O}_{\omega_0}$, i.e.,

$$\begin{aligned} &\mathbf{z}\cdot\mathbf{w} + z_n w_n = 0 \\ &\quad\Rightarrow\quad \phi_1\psi_1 r^2 + (\phi_1\psi_2 + \psi_1\phi_2)\mathbf{x}\cdot\mathbf{p} + \phi_2\psi_2 p^2 + \phi\psi = 0, \\ &z^2 + z_n^2 = s^2 \\ &\quad\Rightarrow\quad \phi_1^2 r^2 + 2\phi_1\phi_2\,\mathbf{x}\cdot\mathbf{p} + \phi_2^2 p^2 + \phi^2 = s^2, \\ &w^2 + w_n^2 = s^2 \\ &\quad\Rightarrow\quad \psi_1^2 r^2 + 2\psi_1\psi_2\,\mathbf{x}\cdot\mathbf{p} + \psi_2^2 p^2 + \psi^2 = s^2, \\ &z_\mu w_\nu - z_\nu w_\mu = s\,m_{\mu\nu} \\ &\quad\Rightarrow\quad \begin{cases} \phi_1\psi_2 - \psi_1\phi_2 = s, \\ \psi\phi_1 - \phi\psi_1 = \dfrac{s^2}{mk}\left(p^2 - \dfrac{mk}{r}\right), \\ \psi\phi_2 - \phi\psi_2 = -\dfrac{s^2}{mk}\,\mathbf{x}\cdot\mathbf{p}. \end{cases} \end{aligned} \tag{8}$$

It follows that

$$\phi = \frac{s}{mk}\left[\mathbf{x}\cdot\mathbf{p}\,\phi_1 + \left(p^2 - \frac{mk}{r}\right)\phi_2\right] \equiv \lambda_1\phi_1 + \lambda_2\phi_2,$$
$$\psi = \frac{s}{mk}\left[\mathbf{x}\cdot\mathbf{p}\,\psi_1 + \left(p^2 - \frac{mk}{r}\right)\psi_2\right] \equiv \lambda_1\psi_1 + \lambda_2\psi_2, \tag{9}$$

and

$$(\phi_1\phi_2)\begin{pmatrix} A & C \\ C & B\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = s^2,$$
$$(\psi_1\psi_2)\begin{pmatrix} A & C \\ C & B\end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = s^2,$$
$$(\phi_1\phi_2)\begin{pmatrix} A & C \\ C & B\end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = 0, \tag{10}$$
$$(\phi_1\phi_2)\begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix}\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = s,$$

where $A = r^2 + \lambda_1^2$, $B = p^2 + \lambda_2^2$, $C = \mathbf{x}\cdot\mathbf{p} + \lambda_1\lambda_2$, $AB - C^2 = s^2$.

If we have a solution of Eq. (10), then we can obtain another solution by applying a transformation T of the form

$$T = \begin{pmatrix} \cos\delta - (C/s)\sin\delta & -(B/s)\sin\delta \\ (A/s)\sin\delta & \cos\delta + (C/s)\sin\delta\end{pmatrix}. \tag{11}$$

Let us determine a particular solution by imposing $\phi_1 = 0$. Then we have $\phi_2 = s/B^{1/2} = r$, $\psi_1 = -s/r$, $\psi_2 = Cs/Br = [(\mathbf{x}\cdot\mathbf{p})/mk]s$. The general solution is then

$$\phi_1 = -\frac{s}{r}\sin\delta,$$
$$\phi_2 = r\cos\delta + \frac{s}{mk}\,\mathbf{x}\cdot\mathbf{p}\,\sin\delta,$$
$$\psi_1 = -\frac{s}{r}\cos\delta, \tag{12}$$
$$\psi_2 = -r\sin\delta + \frac{\mathbf{x}\cdot\mathbf{p}}{mk}\,s\,\cos\delta,$$

and from Eq. (9)

$$\phi = \frac{s}{mk}(rp^2 - mk)\cos\delta - \mathbf{x}\cdot\mathbf{p}\,\sin\delta,$$
$$\psi = -\mathbf{x}\cdot\mathbf{p}\,\cos\delta - \frac{s}{mk}(rp^2 - mk)\sin\delta. \tag{13}$$

We can now impose the condition that $\mathbf{x}$ and $\mathbf{p}$ be canonical and we obtain, as a particular solution,

$$\delta = -\mathbf{x}\cdot\mathbf{p}/s. \tag{14}$$

The solution corresponding to $\delta = 0$, still noncanonical, is interesting, since it provides the variables which regularize the Hamiltonian flow, namely

$$\Xi \equiv (\sqrt{-2mH}\;r\mathbf{p}/mk,\; rp^2/mk - 1) \in S^{n-1},$$
$$\Gamma \equiv \left(\frac{(\mathbf{x}\cdot\mathbf{p})\mathbf{p} - mk\mathbf{x}/r}{\sqrt{-2mH}},\; -\mathbf{x}\cdot\mathbf{p}\right) \in T^*_{\Xi} S^{n-1}. \tag{15}$$

It is a nice feature of this group-theoretical approach that the explicit form of these variables is straightforwardly obtained by simple algebraic considerations.

2. THE QUANTIZATION OF O_{ω_l}

The homogeneous complex structure of O_{ω_l} $(l > 0)$ provides an invariant complex polarization for the prequantized O_{ω_l}. We shall then construct a unitary irreducible representation of $SO(n, 2)$ in a Hilbert space H_l of holomorphic functions on O_{ω_l}. The exposition will be rather sketchy, since the general theory of quantization is well-known for the homogeneous Kaehler case.[14] A premise is necessary, however, on the "quantization rule" $\omega_l \mapsto \lambda$ which provides the representation e^{λ} of the stability subgroup. According to the original Kostant–Souriau theory, the prescription is simply $\lambda = i\omega_l$ which gives, in our case,

$$e^{\lambda}\left[\exp\left(\varphi S + \sum_{\mu<\nu}\varphi_{\mu\nu}M_{\mu\nu}\right)\right] \equiv \exp\left[i\left\langle\omega_l, \left(\varphi S + \sum_{\mu<\nu}\varphi_{\mu\nu}M_{\mu\nu}\right)\right\rangle\right] = \exp(-il\varphi). \tag{16}$$

It has been pointed out by several authors that this choice for λ does not always lead to reasonable results.[6,14] In particular, let w be an element in the Weyl group W_K of $\mathbf{so}(2)^* \oplus \mathbf{so}(n)^*$ with respect to a fixed Cartan subalgebra $\mathbf{h}$. Then $w(\omega_l) \in O_{\omega_l}$ but the representations induced by $\exp(i\omega_l)$ and $\exp[iw(\omega_l)]$ are *not* equivalent, in general. A W_K-invariant quantization rule is given by

$$\omega_l \mapsto \lambda = i\omega_l - \rho, \tag{17}$$

ρ being half the sum of positive roots (this fact follows from results of Schmid on Langland's conjecture[18]). Any other rule, such as

$$\omega_l \mapsto \lambda = i\omega_l - \rho + w^{-1}(\rho_c) + \lambda_0, \tag{18}$$

is again W_K-invariant, provided that $w \in W$ [Weyl group of $\mathbf{so}(n, 2)_c$] is such that $w(i\omega_l) + \rho_c$ belongs to the highest Weyl chamber and $W_K\lambda_0 = \lambda_0$; here ρ_c is half the sum of positive *compact* roots. We prefer Eq. (18) to the former one for the following reason, if the group under consideration is compact, then $\lambda_0 \equiv 0$, $\rho_c \equiv \rho$ and we have $\lambda = i\omega - \rho + w^{-1}(\rho)$, which is the correct quantization rule in the compact case. In our specific problem we obtain

$$\lambda = i(l + n/2 + l_0)S^{\flat}, \tag{19}$$

where l_0 is still to be fixed [see the appendix for the proof of Eq. (19)]. We make the choice $l_0 = -1$ (*independent of n*); whether the introduction of symplectic spinors or some other geometric structure[3,6,19] could account for this is still unknown. It must also be stressed that we are also going to consider noninteger values for $l + \frac{1}{2}n - 1$; this means that we shall obtain a projective representation of $SO(n, 2)$, in general.

Let us now exhibit the homogeneous complex structure of O_{ω_l}. Let $\zeta \in C^n$ be represented by a one-row matrix $(\zeta_1, \ldots, \zeta_n)$; let ζ' be its transpose, so that $\zeta\zeta' = \sum_\mu \zeta_\mu^2$. In the following we shall adhere to the notation of Ref. (16) as closely as possible. Let $\sigma = (\sigma_1, \ldots, \sigma_n)$

be defined as in Sec. 1 by $\sigma_\mu = z_\mu - i w_\mu$. The map $\mu_l : O_{\omega_l} \to C^n$, defined by

$$\zeta = \frac{i}{2s}\left(\sigma + \frac{\sigma\sigma'}{2s(s+l) - \sigma\bar{\sigma}'}\bar{\sigma}\right), \tag{20}$$

is differentiable and nonsingular everywhere (s is a well-defined function of σ, on O_{ω_l}). The following holds:

$$\begin{aligned} \zeta\bar{\zeta}' &= 1 - \frac{2l(s+l)}{2ls + [4l^2s^2 + |\sigma\sigma'|^2]^{1/2}}, \\ 1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2 &= \frac{4l^2}{2ls + [4l^2s^2 + |\sigma\sigma'|^2]^{1/2}}. \end{aligned} \tag{21}$$

The range of μ_l is then the *homogeneous bounded domain* of type IV [16] (Cartan domain),

$$\mathcal{D}^{(n)} = \{\zeta \in C^n \mid \zeta\bar{\zeta}' < 1, \quad 1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2 > 0\}. \tag{22}$$

O_{ω_l} inherits from $\mathcal{D}^{(n)}$ the structure of a $SO(n,2)$-homogeneous Kaehler manifold. The inverse mapping $\mu_l^{-1} : \mathcal{D}^{(n)} \to \mathfrak{so}(n,2)^*$ is given by

$$\begin{aligned} \sigma &= -2il\,\frac{\zeta - \zeta\zeta'\bar{\zeta}}{1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2}, \\ s &= l\,\frac{1 - |\zeta\zeta'|^2}{1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2}, \\ m_{\mu\nu} &= 2il\,\frac{\bar{\zeta}_\mu\zeta_\nu - \zeta_\mu\bar{\zeta}_\nu}{1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2}. \end{aligned} \tag{23}$$

μ_l^{-1} is precisely the "moment" of the symplectic action of $SO(n,2)$ on $\mathcal{D}^{(n)}$ which is given by

$$g\zeta = \frac{(1,-i)C' + 2\zeta D' + (1,i)C'\zeta\zeta'}{\Delta(A') + 2\zeta B'(\tfrac{1}{i}) + \Delta^c(A')\zeta\zeta'}, \tag{24}$$

where

$$g \equiv \begin{pmatrix} \overbrace{A}^{2} & \overbrace{B}^{n} \\ C & D \end{pmatrix}\begin{matrix}\}2\\ \}n\end{matrix},$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I^{(2)} & 0 \\ 0 & -I^{(n)} \end{pmatrix}\begin{pmatrix} A' & C' \\ B' & D' \end{pmatrix} = \begin{pmatrix} I^{(2)} & 0 \\ 0 & -I^{(n)} \end{pmatrix}, \tag{25}$$

$$\Delta(A') \equiv (1,-i)A'(\tfrac{1}{i}), \quad \Delta^c(A') \equiv (1,i)A'(\tfrac{1}{i}).$$

By holomorphic induction from $\exp[i(l + n/2 - 1)S^b]$, we obtain the following representation of $SO(n,2)$,

$$(U_g\Psi)(\zeta) = (j_g{}^{-(l+n/2-1)/n}\Psi)(g^{-1}\zeta) \equiv \mu(g,\zeta)\,\Psi(g^{-1}\zeta). \tag{26}$$

where

$$\begin{aligned} j_g(g^{-1}\zeta) &= \det\left\| \frac{\partial g\xi}{\partial \xi} \right\|_{\xi = g^{-1}\zeta} \\ &= [\tfrac{1}{2}\Delta(A) - \zeta C(\tfrac{1}{i}) + \tfrac{1}{2}\Delta^c(A)\zeta\zeta']^n \end{aligned} \tag{27}$$

[the proof of Eq. (27) is given in the appendix]. The invariant inner product is given by

$$\begin{aligned} \langle\Psi_1 \mid \Psi_2\rangle &= \mathcal{N}\int_{\mathcal{D}^{(n)}} \overline{\Psi_1(\zeta)}\Psi_2(\zeta) \\ &\quad \times (1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2)^{l-n/2-1}\,\mathring{\zeta} \end{aligned}$$

with

$$\mathring{\zeta} = \prod_{\mu=1}^{n} \frac{\overline{d\zeta_\mu} \wedge d\zeta_\mu}{2i}.$$

$\mathcal{N}$ is a normalization factor which can be chosen in such a way that the constant function $\Psi_0(\zeta) \equiv 1$ has norm 1, i.e.,

$$\begin{aligned} \mathcal{N} &= \left[\int_{\mathcal{D}^{(n)}} (1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2)^{l-n/2-1}\,\mathring{\zeta}\right]^{-1} \\ &= \frac{2^n}{\pi^n}\,\frac{(l-1)\,\Gamma(l + n/2 - 1)}{\Gamma(l - n/2)}. \end{aligned} \tag{29}$$

It follows from a theorem of Harish-Chandra [20] that the Hilbert space $\mathcal{H}_l$ of holomorphic functions $\Psi(\zeta)$ such that $\|\Psi\| < \infty$, *is nontrivial if and only if* $l > \frac{1}{2}n$. The representation of $SO(n,2)$ in $\mathcal{H}_l$ is unitary and irreducible; it is the quantization of O_{ω_l}. Since we are interested in the limit $l \to 0$, we shall need a definition of $\mathcal{H}_l$ which make sense also for $l \leq \frac{1}{2}n$; this will be done in Sec. 3. For the moment, let us examine some properties of the representation for $l > \frac{1}{2}n$.

As is well-known, there exists an overcomplete basis $\{|\zeta\rangle \in \mathcal{H}_l\}$ in a one-to-one correspondence with the points of $\mathcal{D}^{(n)}$. The vectors $|\zeta\rangle$ are usually called "coherent states" or "principal vectors" [4,14,21,22] and provide the most simple way of taking the "classical limit" of expectation values and probability distributions. [23,24] The properties of coherent states in $\mathcal{H}_l$ can be summarized as follows:

$$\begin{aligned} &\text{(i)}\ \langle\zeta\mid\Psi\rangle = \mathcal{N}\int_{\mathcal{D}^{(n)}} \overline{\xi(\xi)}\,\Psi(\xi) \\ &\qquad \times (1 - 2\xi\bar{\xi}' + |\xi\xi'|^2)^{l-n/2-1}\,\mathring{\xi} \equiv \Psi(\zeta), \\ &\text{(ii)}\ \xi(\xi) = \langle\xi\mid\zeta\rangle = (1 - 2\xi\bar{\xi}' + \xi\xi'\overline{\zeta\zeta}')^{-l-n/2+1}, \\ &\qquad \text{(reproducing kernel)}, \\ &\text{(iii)}\ \mathcal{N}\int_{\mathcal{D}^{(n)}} |\zeta\rangle\,\langle\zeta|/\langle\zeta|\zeta\rangle\,\mathring{\zeta} = \mathbb{1}, \quad \text{(completeness)}, \\ &\text{(iv)}\ U_g|\zeta\rangle = \overline{\mu(g^{-1},\zeta)}\,|g\zeta\rangle \quad [\text{transformation} \\ &\qquad\qquad\qquad \text{property under } SO(n,2)]. \end{aligned} \tag{30}$$

It follows from Eq. 30 (ii) and (iv) that

$$|g\circ 0\rangle = \frac{U_g|0\rangle}{\langle 0|U_g|0\rangle}, \tag{31}$$

which shows the connection with the general definition of coherent states given in Ref. 22 [see also Ref. (25)].

The state $|\zeta\rangle$ corresponds to a probability distribution ρ_ζ on O_{ω_l} given by $\rho_\zeta = \rho_0(g^{-1}\sigma)$, with $\zeta = g\circ 0$ and

$$\begin{aligned} \rho_0(\sigma) &= \mathcal{N}\left(\frac{4l^2}{2s(s+l) - \sigma\bar{\sigma}'}\right)^{l+n/2-1} \\ &\equiv \mathcal{N}\left(\frac{4l^2}{(s+l)^2 - \sum_{\mu<\nu} m_{\mu\nu}^2}\right)^{l+n/2-1} \end{aligned} \tag{32}$$

Let us note that in our units $\hbar = 1$; Planck's constant would appear at the exponent in Eq. (32) in such a way that in the limit $\hbar \to 0$, ρ_ζ would converge to a Dirac distribution centered on $(s, m_{\mu\nu}, z_\mu, w_\mu)$.

The reproducing kernel contains all the properties of

the representation. Let $\{\varphi_i(\xi) \mid i \in I\}$ be a sequence of linearly independent holomorphic functions in $\mathcal{D}^{(n)}$ such that

$$\langle \xi \mid \xi \rangle = \sum_{i \in I} \varphi_i(\xi)\, \overline{\varphi_i(\xi)}.$$

Then it follows that the set $\{\varphi_i \mid i \in I\}$ is *orthonormal* and *complete*. This fact will be used in Sec. 3.

The representation $X \mapsto \hat{X}$ of the Lie algebra so$(n, 2)$ is easily obtained by differentiating Eq. (26),

$$\hat{S} = \left(l + \frac{n}{2} - 1\right) + \sum_{\mu} \xi_\mu \partial_\mu ,$$

$$\hat{Z}_\mu = i\left(l + \frac{n}{2} - 1\right)\xi_\mu + i\sum_{\nu}\left(\xi_\mu\xi_\nu - \frac{1+\xi\xi'}{2}\,\delta_{\mu\nu}\right)\partial_\nu ,$$

$$\hat{W}_\mu = -\left(l + \frac{n}{2} - 1\right)\xi_\mu - \sum_{\nu}\left(\xi_\mu\xi_\nu + \frac{1-\xi\xi'}{2}\,\delta_{\mu\nu}\right)\partial_\nu ,$$

$$\hat{M}_{\mu\nu} = -i(\xi_\mu\partial_\nu - \xi_\nu\partial_\mu) ,$$

$$\left(\partial_\mu \equiv \frac{\partial}{\partial \xi_\mu}\right).$$

We know that in general $\langle \xi \mid \hat{X} \mid \xi \rangle / \langle \xi \mid \xi \rangle$ coincides with the classical generating function $x(\xi, \bar{\xi})$ [14]; this can be easily verified here, except for a factor $(l + \frac{1}{2}n - 1)/l$ due to the new quantization rule Eq. (19).

The representation of so$(n, 2)$ is found to be identical to that of Ref. 26 if we are allowed to take $l = 0$ and real ξ!

3. THE QUANTIZATION OF THE KEPLER MANIFOLD

We now come to the main result of this paper, namely that it is possible to take the continuation of the representation of SO$(n, 2)$ for $l \leq \frac{1}{2}n$ and the limit $l \to 0$ yields the quantization of the Kepler manifold. We shall study the limit $l \to 0$ of $\mathcal{O}_{\omega_l}$, of μ_l, and finally of the representation.

It is clear from the definition that $\tilde{\mathcal{O}}_0 \equiv \lim_{l \to 0} \mathcal{O}_{\omega_l}$ *contains* $\mathcal{O}_{\omega_0}$. What happens is that $\tilde{\mathcal{O}}_0$ is not homogeneous; it splits into $\{0\} \equiv \{\sigma = 0\}$, $\mathcal{O}_{\omega_0} \equiv \{\sigma\sigma' = 0\}$ and a $2n$-dimensional orbit $\mathcal{O}_\vartheta = \{\sigma\sigma' \neq 0\}$. Let us note that for any point in $\tilde{\mathcal{O}}_0$, the following holds,

$$m^2 \equiv \tfrac{1}{2}\sum_{\mu\nu} m_{\mu\nu}^2 = s^2 - |\sigma\sigma'| . \tag{35}$$

$\mathcal{O}_{\omega_0}$ is then the homogeneous component of $\tilde{\mathcal{O}}_0$ where m^2 [the canonical invariant of SO(n)] *is a function of* s *alone*. This corresponds to the fact that the submanifolds $\Sigma_s = \{s = \text{constant}\}$ are SO(n)-homogeneous. A similar property will hold for the quantization of SO$(n, 2)$.

When $l \to 0$ the map μ_l becomes singular. From Eq. (21) it follows that:

$$\begin{aligned} &\text{(i)}\ \ \sigma = 0 \Rightarrow \xi = 0, \\ &\text{(ii)}\ \ |\sigma\sigma'| \to L \neq 0 \Rightarrow \xi\bar{\xi}' \to 1,\ \ 1 - 2\xi\bar{\xi}' + |\xi\xi'|^2 \to 0, \\ &\text{(iii)}\ \ \sigma\sigma' \to 0 \Rightarrow 1 - 2\xi\bar{\xi}' + |\xi\xi'|^2 \to 0, \end{aligned} \tag{36}$$

but $\xi\bar{\xi}'$ has no limit.

Eq. (36 (ii)) means that a limit map can be defined on $\mathcal{O}_\vartheta$ and its range is the *Bergman–Šilov* (B–$\check{S}$) *boundary of* $\mathcal{D}^{(n)}$,

$$\xi = \frac{i}{2s}\left(\sigma + \frac{\sigma\sigma'}{|\sigma\sigma'|}\,\bar{\sigma}\right) \in S^{n-1} \times S^1, \tag{37}$$

i.e.,

$$\xi = x\exp(i\omega), \quad \omega = \tfrac{1}{2}\arg\{\sigma\sigma'\} + \frac{\pi}{2} ,$$

$$x = \mathrm{Re}\left\{\exp[-(i/2)\arg\{\sigma\sigma'\}]\,\frac{\sigma}{s}\right\}.$$

In case (iii) all that we can define is a map $\nu_0 : \partial\mathcal{D}^{(n)} \to \Sigma_s$ (the energy surface). From Eq. (23) we obtain

$$\nu_0(\xi) = \frac{\sigma}{2s} = \frac{1}{i}\,\frac{\xi - (\xi\xi')\bar{\xi}}{1 - |\xi\xi'|^2} , \tag{38}$$

where ξ now belongs to the boundary $\partial\mathcal{D}^{(n)}$. Let $\mathcal{C}_0 \equiv \{\xi \in \partial\mathcal{D}^{(n)} \mid \xi\xi' = 0\}$, then all points $\xi = \xi_0 + \lambda\bar{\xi}_0$ with $\xi_0 \in \mathcal{C}_0$, $|\lambda| < 1$ belong to $\partial\mathcal{D}^{(n)}$ and are mapped into the same point $\nu_0(\xi) = \sigma/2s = -i\xi_0$. The boundary $\partial\mathcal{D}^{(n)}$ (subtracted from the B–$\check{S}$ boundary) is fibered with base space $\mathcal{C}_0$ and fiber $\mathcal{D}^{(1)}$. $\mathcal{C}_0$ *is diffeomorphic* to Σ_s. By restricting to $\mathcal{C}_0$, we can associate a function on $\mathcal{O}_{\omega_0}$ to any function on $\mathcal{D}^{(n)}$ which extends to the boundary; the converse is not true, however, since there does not exist any Cauchy formula for $\mathcal{C}_0$, analogous to that for the B–$\check{S}$ boundary.

Let us now consider the limit of the representation. We know that $\mathcal{H}_l$ is nontrivial only for $l > \frac{1}{2}n$. *But* $\mathcal{H}_l$ *can be equivalently defined in a way that preserves its meaning also for* $l < \frac{1}{2}n$. This can be done as follows: $\mathcal{H}_l$ is a space of holomorphic functions on $\mathcal{D}^{(n)}$; it is the linear span of $\{U_g\Psi_0 \mid g \in \mathrm{SO}(n, 2),\ \Psi_0(\xi) \equiv 1\}$ (completed according to the topology given by uniform convergence), explicitly,

$$(U_g\Psi_0)(\xi) = j_g(g^{-1}\xi)^{-(1/n)(l+n/2-1)}. \tag{39}$$

The norm is implicitly defined by *requiring* that the kernel $K(\xi, \bar{\xi}) = (1 - 2\xi\bar{\xi}' + \xi\xi'\overline{\xi\xi'})^{-(l+n/2-1)}$ *be reproducing*, i.e., $K_\xi(\xi) = K(\xi, \bar{\xi}) \in \mathcal{H}_l$ and $\langle K_\xi \mid \Psi \rangle = \Psi(\xi)$. The action of SO$(n, 2)$ is unchanged. The point is that this new definition is equivalent to the previous one for $l > \frac{1}{2}n$ but makes sense also for $l \leq \frac{1}{2}n$ (the difference being that the representation is square-integrable for $l > \frac{1}{2}n$ while it is not for $l \leq \frac{1}{2}n$). [27]

Now, let us examine the case $l = 0$,

$$\begin{aligned} K(\xi, \bar{\xi}) &= (1 - 2\xi\bar{\xi}' + \xi\xi'\overline{\xi\xi'})^{-n/2+1} \\ &= \sum_{i \in I} \varphi_i(\xi)\,\overline{\varphi_i(\xi)}. \end{aligned} \tag{40}$$

The holomorphic functions $\varphi_i(\xi)$ are uniquely characterized by their values on the B–$\check{S}$ boundary; then it is sufficient to consider the following expansion, [28]

$$\begin{aligned} &K(x\exp(i\vartheta),\ y\exp(-i\omega)) \\ &= (1 - 2xy'\exp[i(\vartheta - \omega)] + \exp[2i(\vartheta - \omega)])^{-n/2+1} \\ &= \frac{4\pi^{n/2}}{\Gamma(n/2 - 1)} \sum_{s=0}^{\infty} \frac{\exp[is(\vartheta - \omega)]}{2s + n - 2} \sum_{\alpha} Y_{s\alpha}(x)\,\overline{Y_{s\alpha}(y)}. \end{aligned} \tag{41}$$

$\{Y_{s\alpha}(x)\}$ is an orthonormal set of spherical harmonics of degree s on S^{n-1}, $\alpha=(k_1,\ldots,k_{n-2})$ with $s\geq k_1\geq k_2\geq\cdots\geq|k_{n-2}|\geq 0$ (see Ref. 29, Sec. 3.6); the number of linearly independent $Y_{s\alpha}$ is

$$N(n,s)=(2s+n-2)\Gamma(s+n-2)/\Gamma(n-1)\Gamma(s+1). \quad (42)$$

By analytic continuation we find an orthonormal basis in H_0,

$$\varphi_{s\alpha}(\xi)=\left[\frac{2\pi^{n/2}}{\Gamma(\frac{1}{2}n-1)(s+\frac{1}{2}n-1)}\right]^{1/2} \times(\xi\xi')^{s/2}Y_{s\alpha}\left(\frac{\xi}{\sqrt{\xi\xi'}}\right). \quad (43)$$

The functions $\varphi_{s\alpha}$ are homogeneous polynomials of degree s; then

$$\hat{S}\varphi_{s\alpha}=(\tfrac{1}{2}n-1)+s\varphi_{s\alpha}. \quad (44)$$

For a fixed value s, we have a $N(n,s)$-dimensional subspace where an irreducible representation of $SO(n)$ is defined; the Casimir invariant $\hat{M}^2=\frac{1}{2}\sum_{\mu\nu}\hat{M}_{\mu\nu}^2$ is simply related to the Laplace operator on S^{n-1} and

$$\hat{M}^2\varphi_{s\alpha}(\xi)=s(s+n-2)\varphi_{s\alpha}. \quad (45)$$

We find that $\hat{S}$ has the same spectrum and multiplicity as the operator $mk/(-2mH)^{1/2}$, where H is the usual quantum-mechanical Hamiltonian for the $(n-1)$-dimensional Hydrogen atom (negative energy only).

The explicit form of $\varphi_{s\alpha}$ shows that a *particular realization* of our Hilbert space H_0 is the following. H_0 is given by functions defined on S^{n-1} such that

$$\langle\Psi_1|\Psi_2\rangle=\frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{n/2}} \times\int_{S^{n-1}}\overline{\Psi_1(x)}\,[-\Delta+(\tfrac{1}{2}n-1)^2]^{1/2}\Psi_2(x)\,\dot{x}. \quad (46)$$

Given such a $\Psi(x)$, a holomorphic function on $D^{(n)}$ is constructed as follows,

$$\Psi(x)=\sum_{s\alpha}\Psi_{s\alpha}Y_{s\alpha}(x)$$

$$\Rightarrow\Psi(\xi)=\sum_{s\alpha}\Psi_{s\alpha}(\xi\xi')^{s/2}Y_{s\alpha}\left(\frac{\xi}{\sqrt{\xi\xi'}}\right).$$

The reproducing property of $K(\xi,\xi)$ is then

$$\Psi(x)=\langle K_x|\Psi\rangle=\frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{n/2}} \times\int_{S^{n-1}}\frac{[-\Delta+(\frac{1}{2}n-1)^2]^{1/2}\Psi(y)}{\|x-y\|^{n-2}}\,\mathring{y}. \quad (47)$$

where $\|x-y\|=\sqrt{(x-y)(x-y)'}$. In particular for any eigenfunction of S belonging to the eigenvalue $s+\frac{1}{2}n-1$ it holds that

$$\Psi_s(x)=\frac{\Gamma(\frac{1}{2}n-1)}{2\pi^{n/2}}(s+\tfrac{1}{2}n-1) \times\int_{S^{n-1}}\frac{\Psi_s(y)}{\|x-y\|^{n-2}}\,\mathring{y}. \quad (48)$$

which is exactly the $SO(n)$-invariant integral equation discovered by Fock for $n=4$ (see Ref. 12 for the general case).

The realization of H_0 on the sphere is not the only one at our disposal. We can represent our functions $\varphi_{s\alpha}$ by means of their restriction $\mathring{\varphi}_{s\alpha}$ to C_0, that is to say to the energy surface Σ_s, through ν_0 [Eq. (38)]. The functions $\mathring{\varphi}_{s\alpha}$ are simply related to the matrix elements of the representation $D^{(n)}$ of $SO(n)$. Since C_0 is $SO(n)$-homogeneous, we have

$$\mathring{\varphi}_{s\alpha}(R\xi_0)=(U_{R^{-1}}\varphi_{s\alpha})(\xi_0) =\sum_{\alpha'}\varphi_{s\alpha'}(\xi_0)D^{(s)}_{\alpha'\alpha}(R^{-1}), \quad (49)$$

where $\xi_0=(\frac{1}{2}i,\frac{1}{2},0,\ldots,0)$. Now $\mathring{\varphi}_{s\alpha}(\xi_0)=0$ unless $\alpha=(s,s,\ldots,s)$ (highest weight). Then $\mathring{\varphi}_{s\alpha}(\xi)$ is proportional to the "first column" of the representation $D^{(s)}$:

$$\mathring{\varphi}_{s\alpha}(\xi)\propto D^{(s)}_{s\alpha}(R^{-1}_{\xi_0\to\xi})\equiv\langle ss\,|\,U^{\dagger}_{R_{\xi_0\to\xi}}|\,s\alpha\rangle, \quad (50)$$

which shows that the realization on C_0 is "coherent" with respect to $SO(n)$.[22,25] As a matter of fact, the realization on S^{n-1} is coherent too, in the generalized sense of Ref. 22,

$$Y_{s\alpha}(x)\propto D^{(s)}_{o\alpha}(R^{-1}_{x_0\to x})\equiv\langle so\,|\,U^{\dagger}_{R_{x_0\to x}}|\,s\alpha\rangle, \quad (51)$$

$0=(o,o,\ldots,o)$ and $x_0=(o,o,\ldots,1)\in S^{n-1}$ (see Ref. 29). The stability subgroups of the rays $|so\rangle$ and $|ss\rangle$ are $SO(n-1)$ and $SO(n-2)$ respectively, and correspondingly we have functions on S^{n-1} and C_0, respectively.

Let us note that, as a by-product, we obtain that the two columns "O" and "s" of the representation are connected by an analytic continuation! Explicitly we have (by Cauchy's formula),

$$\varphi_{s\alpha}(\xi)\propto\int_{S^{n-1}}\dot{x} \times\int_0^{2\pi}d\omega\,\frac{Y_{s\alpha}(x)\exp(is\omega)}{[1-2x\xi'\exp(-i\omega)+\xi\xi'\exp(-2i\omega)]^{n/2}}$$

$$\propto(\xi\xi')^{s/2}\int_{S^{n-1}}\dot{x}\,Y_{s\alpha}(x)\,C_s^{n/2}\left(\frac{x\xi'}{\sqrt{\xi\xi'}}\right). \quad (52)$$

Taking into account the explicit form of Gegenbauer polynomials and restricting to C_0 ($\xi\xi'=0$) we have[30]

$$\mathring{\varphi}_{s\alpha}(\xi)\propto\int_{S^{n-1}}\dot{x}\,Y_{s\alpha}(x)\,(x\xi')^s \quad (53)$$

(a similar formula can be found in Ref. 31).

ACKNOWLEDGMENTS

The main result of the paper was announced at the C.N.R.S. Conference held at Aix-en-Provence in June '74; I thank warmly Professor J.M. Souriau for his kind hospitality at the Congress. The work was completed at Bedford College, University of London, where I benefited from a N.A.T.O. Senior Fellowship from Italian Consiglio Nazionale delle Ricerche. I thank warmly Professor R.F. Streater for his kind hospitality. Thanks are also due to Professor F. Duimio and Professor M. Pauri for reading the manuscript and for useful comments.

APPENDIX TO SEC. 2

A. Root spaces for SO (n,2) = quantization rule

(i) $n=2r-1$, $so(n,2)\approx B_r$, $i\omega_l = le_1 = ilS^\flat$.

Roots: $\begin{cases} \pm e_k \quad (k=1,2,\ldots,r), \\ \pm e_k \pm e_h \quad (1\leq k<h\leq r), \end{cases}$

Positive roots: $-e_k$, $-e_k \pm e_h$,

Positive noncompact roots:

$-e_1$, $-e_1 \pm e_h$,

$\rho - \rho_c = -(r-\frac{1}{2})e_1$,

$i\omega_l - \rho + \rho_c = (l+r-\frac{1}{2})e_1 = (l+n/2)e_1$.

(ii) $n=2r-2$, $so(n,2)\approx D_r$.

Roots: $\pm e_k \pm e_h$ $(1\leq k<h\leq r)$.

Positive roots: $-e_k \pm e_h$.

Positive noncompact roots:

$-e_1 \pm e_h$, $\rho - \rho_c = -(r-1)e_1$,

$i\omega_l - \rho + \rho_c = (l+r-1)e_1 = (l+n/2)e_1$.

B. The multiplier $\mu(g,\zeta)$

The most convenient method to calculate $\mu(g,\zeta) = j_g(g^{-1}\zeta)^{-\lambda/n}$ is the following, based on the existence of the basis $|\zeta\rangle$,

(i) $\langle\zeta|U_g|0\rangle = \mu(g,\zeta)\langle g^{-1}\zeta|0\rangle = \overline{\mu(g^{-1},0)}\langle\zeta|g\circ 0\rangle$,

(ii) $\langle 0|U_g|0\rangle = \mu(g,0) = \overline{\mu(g^{-1},0)}$,

(i) + (ii) $\mu(g,\zeta) = \mu(g,0)\langle\zeta|g\circ 0\rangle$

$= j_{g^{-1}}(0)^{(l+n/2-1)/n}\langle\zeta|g\circ 0\rangle$

We are left to calculate $j_{g^{-1}}(0)$. From Eq. (24),

$$\left.\frac{\partial(g\zeta)_\mu}{\partial\zeta_\nu}\right|_{\zeta=0} = \frac{2D_{\mu\nu}\Delta(A') - 2(C_{\mu 1} - iC_{\mu 2})(B_{1\nu}+iB_{2\nu})}{\Delta(A')^2}$$

$$= \frac{2D_{\mu\sigma}}{\Delta(A')}\left[\delta_{\sigma\nu} - \frac{D^{-1}_{\sigma\rho}(C_{\rho 1}-iC_{\rho 2})(B_{1\nu}+iB_{2\nu})}{\Delta(A')}\right],$$

$$j_g(0) = \left(\frac{\Delta(A')}{2}\right)^{-n} \det D\left(1 - \frac{\Delta[(BD^{-1}C)']}{\Delta(A')}\right).$$

Taking into account that $BD^{-1}C = BB'A'^{-1} = A - A'^{-1}$ and that $\det D = \det A$, we obtain

$$j_g(0) = \left(\frac{\Delta(A')}{2}\right)^{-n} \det A\,\frac{\Delta(A'^{-1})}{\Delta(A')} = \left(\frac{\Delta(A')}{2}\right)^{-n}.$$

Finally

$$\mu(g,\zeta) = \left(\frac{\Delta(A)}{2}\right)^{-(l+n/2-1)}\left[1 - 2\frac{\sum\zeta_\mu(C_{\mu 1}+iC_{\mu 2})}{\Delta(A)} + \zeta\zeta'\frac{\sum(C_{\mu 1}+iC_{\mu 2})^2}{\Delta(A)^2}\right]^{-l-n/2+1}$$

$$= \left[\frac{\Delta(A)}{2} - \zeta C(\tfrac{1}{i}) + \frac{\Delta^c(A)}{2}\zeta\zeta'\right]^{-l-n/2+1}.$$

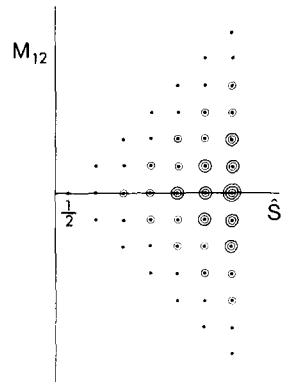


FIG. 1. The level structure (weight diagram) for $n=3$, $l>0$.

APPENDIX TO SEC. 3: DETAILS ON THE CASE $n=3$

Let us examine the representation corresponding to $l\geq 0$, $n=3$. We have

$$(1 - 2\zeta\bar{\zeta}' + |\zeta\zeta'|^2)^{-(l+1/2)} = \sum_{s=0}^{\infty}\sum_{m=-s}^{s}{\sum_k}' |\varphi_{smk}(\zeta)|^2, \tag{54}$$

where

$$\varphi_{smk}(\zeta) = \frac{2^{2l+k-1}}{\sqrt{\pi\Gamma(l)}}\,\Gamma(l+\tfrac{1}{2}+k)\left(\frac{(s-k)!\,2(l+k)}{\Gamma(s+k+2l+1)}\right)^{1/2} \times \frac{\Gamma[l+(k+m)/2]\,\Gamma[l+(k-m)/2]}{[(k+m)/2]!\,[(k-m)/2]!} \times(\zeta_1+i\zeta_2)^{(k+m)/2}(\zeta_1-i\zeta_2)^{(k-m)/2} \times(\zeta\zeta')^{(s-k)/2}C^{l+k+1/2}_{s-k}\left(\frac{\zeta_3}{\sqrt{\zeta\zeta'}}\right). \tag{55}$$

k takes on only the values such that $(k-|m|)/2 = 0,1,2,\ldots,[(s-|m|)/2]$ and $C^\lambda_n(x)$ are Gegenbauer polynomials.

φ_{smk} are eigenfunctions of S and M_{12}; k labels the vectors belonging to the same weight. This means that the subspace corresponding to a fixed value s is reducible under the action of SO(3) as follows (see Fig. 1),

$$\Sigma_s = D^{(s)}\oplus D^{(s-2)}\oplus\cdots\oplus D^{(\epsilon)},\quad \epsilon = s - 2\left[\frac{s}{2}\right]. \tag{56}$$

At this point we take the limit $l\to 0$. We find that all normalization factors vanish except for $k=|m|$; correspondingly,

$$\varphi_{sm|m|}(\zeta)\to\sqrt{4\pi/(2s+1)}\;Y^m_s\left(\frac{\zeta}{\sqrt{\zeta\zeta'}}\right)\quad(\zeta\zeta')^{s/2}, \tag{57}$$

and we are left with simple weights (s,m) and irreducible energy levels [with respect to SO(3)]. The representation of SO(3,2) that we obtain is well-known as one of the "Majorana" representations.

The explicit form of φ_{sm} when restricted to the submanifold C_0, is the following,

$$\mathring{\varphi}_{sm}(\zeta) = \frac{(2s)!}{2^s s!\sqrt{(s-m)!\,(s+m)!}} \times(\zeta_1+i\zeta_2)^{(m+|m|)/2}(-\zeta_1+i\zeta_2)^{(|m|-m)/2} \tag{58}$$

Since $C_0 \approx SO(3) \approx SU(2)/Z_2$, we can introduce complex coordinates a_1, a_2 defined by

$$\begin{aligned} &\zeta_1 + i\zeta_2 = a_1^2, \qquad |a_1|^2 + |a_2|^2 = 1, \\ &-\zeta_1 + i\zeta_2 = a_2^2, \quad (a_1, a_2) \equiv (-a_1, -a_2), \\ &\zeta_3 = a_1 a_2. \end{aligned} \tag{59}$$

In terms of a_i we have the more familiar formulas:

$$\mathring{\varphi}_{sm}(a_1, a_2) = \frac{(2s)!}{2^s s! \sqrt{(s-m)!\,(s+m)!}}\, a_1^{s+m} a_2^{s-m}. \tag{60}$$

We can easily check Eq. (50) in this particular case:

$$\mathring{\varphi}_{sm}(R\zeta_0) = \frac{\sqrt{(2s)!}\; i^s}{2^s s!}\, D^{(s)}_{sm}(R^{-1}). \tag{61}$$

The expansion of the kernel

$$\begin{aligned} (1 - 2\xi\bar{\xi}' + \zeta\zeta'\bar{\xi}\bar{\xi}')^{-1/2} &= [1 - (a_1\bar{b}_1 + a_2\bar{b}_2)^2]^{-1/2} \\ &= \textstyle\sum_{sm} \mathring{\varphi}_{sm}(a_1, a_2)\, \overline{\mathring{\varphi}_{sm}(b_1, b_2)} \end{aligned}$$

is simply given by the binomial theorem.

[1]J. M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1970).
[2]B. Kostant, *Quantization and Unitary representations*, *Lecture Notes in Mathematics*, edited by C. T. Taam (Springer, New York, 1970), Vol. 170.
[3]R. J. Blattner, Proc. Symp. Pure Math. 26 (Amer. Math. Soc. Providence, Rhode Island, 1973).
[4]E. Onofri and M. Pauri, Lett. Nuovo Cimento 2(3), 35 (1972).
[5]J. Rawnsley, *DeSitter symplectic spaces and their quantizations* (Mathematical Institute, Oxford, Preprint 1973); *Orbits for inhomogeneous orthogonal groups* I & II (Napoli, Preprint 1974); *Representations of a semidirect product by quantization* (Napoli, Preprint 1974).
[6]D. J. Simms, "Geometric Quantization of Energy Levels in the Kepler Problem," in *Symposia Mathematica* (Academic, London, 1974), Vol. XIV; "Metalinear Structures and a Geometric Quantization of the Harmonic Oscillator" (Int. Colloq. Symplectic Geometry Math. Phys., Aix-en-Provence, 1974).
[7]E. Onofri and M. Pauri, J. Math. Phys. 13, 533 (1972).
[8]J. Elhadad, "Sur l'interpretation en géometrie symplectique des etats quantiques de l'atome d'hydrogène," in *Symposia Mathematica* (Academic, London, 1974), Vol. XIV.
[9]V. Fock, Z. Physik 98, 145 (1935).
[10]J. Moser. Commun. Pure Appl. Math. 23, 609 (1970).
[11]J. M. Souriau, "Sur la variété de Kepler," in *Symposia Mathematica* (Academic, London, 1974), Vol. XIV.
[12]M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 (1966).
[13]A. Knapp, in *Symmetric Spaces*, edited by W. M. Boothby and G. L. Weiss (Dekker, New York, 1972).
[14]E. Onofri, "Quantization Theory for Homogeneous Kaehler Manifolds," University of Parma (1974).
[15]F. Bruhat, "Travaux de Harish-Chandra," Sém. Bourbaki 143 (1957).
[16]L. K. Hua, *Harmonic Analysis* (Amer. Math. Soc., Transl. Math. Monographs, 1963), Vol. 6.
[17]G. Gyorgyi, Acta Phys. Acad. Sci. Hung. 27, 435 (1969).
[18]W. Schmid, Ann. Math. 93, 1 (1971).
[19]B. Kostant, "Symplectic Spinors", in *Symposia Mathematica* (Academic, London, 1974), Vol. XIV.
[20]Harish-Chandra, Am. J. Math. 77, 743 (1955); 78, 1 (1956).
[21]V. Bargmann, Commun. Pure Appl. Math. 14, 354 (1961).
[22]A. M. Peremolov, Commun. Math. Phys. 26, 222 (1972); "Coherent states and symmetric spaces" (ITEP-17, Moscow, 1974).
[23]K. Hepp, Commun. Math. Phys. 35, 265 (1974).
[24]E. H. Lieb, Commun. Math. Phys. 31, 327 (1973).
[25]E. Onofri, J. Math. Phys. 16, 1087 (1975).
[26]I. A. Malkin and V. I. Man'ko, JETP Lett. 2, 146 (1965).
[27]A. L. Carey, "Induced representations, reproducing kernels and the conformal group," Mathematical Institute, Oxford, Preprint (1974).
[28]C. Muller, *Spherical Harmonics* (Springer-Verlag, Berlin, 1966) Lecture Notes in Mathematics, Vol. 17. See also H. Hochstadt, *The functions of Mathematical Physics* (Wiley, N. Y., 1971).
[29]N. Ja. Vilenkin, *Fonctions spéciales et théorie de la représentation des groupes* (Dunod, Paris, 1969).
[30]Equation 53 can also be derived from orthogonality properties of $D^{(s)}$, once it is known that $\hat{\phi}_{s\alpha} \propto D^{(s)}_{s\alpha}$.
[31]R. S. Strichartz, J. Funct. Anal. 12, 341 (1973).

On the equality of S operators corresponding to unitarily equivalent Hamiltonians in single channel scattering

A. W. Sáenz and W. W. Zachary

Naval Research Laboratory, Washington, D.C. 20375
(Received 12 September 1975)

Some years ago, conditions in order that unitarily equivalent Hamiltonians H and W^*HW yield the same S operator were investigated by Ekstein and by Coester and his collaborators in the framework of nonrelativistic time-dependent single-channel quantum scattering theory. This subject has turned out to be of considerable importance for nuclear physics, since it constitutes the foundation of a widely used method for constructing phase-equivalent nucleon–nucleon potentials. In the present paper we derive a rigorous and concise necessary and sufficient condition for a pair of unitarily equivalent Hamiltonians, governing the dynamics of a pair of nonrelativistic particles in the center-of-mass system, to yield the same S operator. Our theory, based on a time-dependent approach, applies to very general types of short-range potentials, with or without hard cores, and to an extensive class of long-range potentials. Our necessary and sufficient condition simplifies when certain strong limits $W_\pm$, related to W, are unitary and when $W_+ = W_-$. Requirements sufficient for each of these properties to hold are determined. Various examples of operators W such that H and W^*HW have the same S operator are discussed.

1. INTRODUCTION

Using time-dependent methods, Ekstein[1] and Coester and his collaborators[2] determined, some years ago, conditions sufficient to guarantee that unitarily equivalent Hamiltonians lead to the same S operator in the context of nonrelativistic single-channel scattering. This type of transformation of the Hamiltonian has recently been employed in many nuclear physics investigations, in particular, for generating phase-equivalent potentials to study off-shell effects in nuclear few-body problems and nuclear matter.[2,3] However, as of this writing, a systematic study of this transformation method was not available in the literature.

We propose to initiate such a study in the present paper by a rigorous time-dependent scattering theory approach. The Hamiltonians dealt with here are nonrelativistic ones governing the behavior of a pair of interacting particles in the center-of-mass frame. Our methods can also be used in the case of multichannel scattering, as we shall show in a separate publication.

In Sec. 2, we consider a pair of unitarily equivalent Hamiltonians H and W^*HW, supposing that H contains only a short-range potential, which is allowed to be of a very general kind. Some of the topics discussed in this section are as follows. We prove a concise necessary and sufficient condition for the above two Hamiltonians to yield the same S operator. A special version of the sufficiency aspect of this condition was given previously.[1,2] Our necessary and sufficient condition involves certain strong limits $W_\pm$, related to W, which are shown to exist under very general circumstances. The condition simplifies if $W_\pm$ are unitary and if $W_+ = W_-$. Sufficient conditions for these two properties to hold are obtained. It is found that, when they exist, $W_\pm$ are unitary if W has a gap in its spectrum or if $W_\pm$ commute with appropriate observables. The relation $W_+ = W_-^*$ is shown to follow if W is assumed to be rotationally and time-reversal invariant. It is proved that $W_+ = W_-$ is a necessary and sufficient condition for the equality of the S operators pertaining to H and W^*HW, provided, in particular, that the appropriate operators commute with suitable observables and that the former S operator is unitary.

In Secs. 3 and 4, we generalize most of the theoretical developments of Sec. 2 to apply to Hamiltonians with long-range potentials without hard cores and short-range potentials with hard cores, respectively, under very weak conditions on these two types of potentials. The Coulombic potentials and spherically symmetric hard-core potentials included in this generalization have special relevance for nuclear physics applications. The pertinent results in Sec. 3, when specialized to the former potentials, provide a rigorous foundation for the procedure employed recently by Sauer[3] to construct phase-equivalent potentials for proton–proton scattering.

In Sec. 5, we discuss examples of unitary operators W which are such that the S operators corresponding to a pair of Hamiltonians H and W^*HW of suitable type coincide.

Appendices A and B are devoted to the proof of certain mathematical results connected with subjects discussed in Sec. 2–5.

2. FORMULATION AND RESULTS FOR SHORT-RANGE POTENTIALS WITHOUT HARD CORES

In this section, we first formulate our problem in a time-dependent manner for a pair of nonrelativistic spinless particles, assuming that the pertinent potentials are of the above type. We then establish a number of theorems concerning the scattering properties of two-particle systems of this kind having unitarily equivalent Hamiltonians.

We will work in the center-of-mass system of the pair of particles of interest. All operators considered in this and the next section have domains dense in the Hilbert space $\mathcal{H} = L^2(R^3)$ and ranges in $\mathcal{H}$. Our units are such that the kinetic energy operator H_0 of the latter pair is the unique self-adjoint extension of the negative

Laplacian $-\Delta$ on $C_0^\infty(R^3)$. The self-adjoint Hamiltonian operator governing the dynamics of the pair in the above system will be denoted by H. In general, we will not find it necessary to specify whether H is a self-adjoint extension of $-\Delta+V$, where V denotes the potential describing the interactions of the pair, or whether it is to be interpreted as the sum of the quadratic forms of H_0 and V in certain cases in which this self-adjoint extension does not exist.[4] The potential V is not required to be local.

Letting

$$U_t=\exp(-itH_0), \quad V_t=\exp(-itH) \tag{2.1}$$

for each real t, we define the Møller wave operators by

$$\Omega_\pm=\text{s-}\lim_{t\to\pm\infty}\Omega_t \tag{2.2}$$

in the present section, whenever these strong limits exist. Here

$$\Omega_t=V_t^*U_t \tag{2.3}$$

over the range of t just stated.[5]

There have been many investigations of the self-adjointness of H, the existence of the limits (2.2), the unitarity of the corresponding S operator, etc. for "nonsingular" as well as "singular" short-range potentials.[6]

The work of the present section applies to short-range potentials of nonsingular or singular type, provided that the respective requirements stated explicitly below are fulfilled. As was stated in the Introduction, generalizations to long-range potentials and to those with hard cores will be made in later sections.

We now consider a "transformed" system, whose kinetic energy operator is again H_0 and whose Hamiltonian is

$$\tilde{H}=W^*HW, \tag{2.4}$$

where W denotes a unitary operator. Hence it is natural to set

$$\tilde{V}_t=\exp(-it\tilde{H})=W^*V_tW \tag{2.5}$$

and to define new wave operators by

$$\tilde{\Omega}_\pm=\text{s-}\lim_{t\to\pm\infty}\tilde{\Omega}_t, \tag{2.6}$$

whenever these limits exist. Here

$$\tilde{\Omega}_t=\tilde{V}_t^*U_t=W^*V_t^*WU_t. \tag{2.7}$$

We shall be mainly interested in the case when W is such that the scattering operators

$$S=\Omega_+^*\Omega_-, \quad \tilde{S}=\tilde{\Omega}_+^*\tilde{\Omega}_- \tag{2.8}$$

pertaining to H and $\tilde{H}$, respectively, are equal.

It is convenient to define the operators

$$W_\pm=\text{s-}\lim_{t\to\pm\infty}W_t \tag{2.9}$$

when these limits exist, where

$$W_t=U_t^*WU_t. \tag{2.10}$$

The special case when $W_\pm=\lambda I$, where λ is a unimodular complex number, and I the identity operator on $\mathcal{H}$, can evidently be reduced to the one in which $W_\pm=I$ by considering the unitary operator $\bar{\lambda}W$ instead of W. Only the case $W_\pm=I$ is considered in Refs. 1 and 2, where in the former reference $W_\pm$ corresponds to an operator H_0 different from the one defined here. In order for $W_\pm=I$ to hold, it is clearly necessary and sufficient that

$$\text{s-}\lim_{t\to\pm\infty}(W-I)U_t=0. \tag{2.11}$$

A familiar sufficient condition for (2.11) to hold is that $W-I$ be compact, but this condition is not necessary.[7] One easily shows that the requirement $W_\pm=\lambda I$ is sufficient for the equality

$$\tilde{S}=S \tag{2.12}$$

to obtain when S exists. However, it is not necessary in order for (2.12) to hold, contrary to a statement in Ref. 2. In Sec. 5 we will discuss an example illustrating this last remark.

Symmetry properties play an important role in the present paper. Thus, we shall have occasion to consider the invariance of appropriate linear operators T, which may be either bounded or unbounded and self-adjoint, with respect to

(a) spatial rotations: T commutes with the three self-adjoint operators L_i $(i=1, 2, 3)$ corresponding to the angular momentum components,

(b) time reversal: T commutes with the operator θ defined by

$$(\theta f)(\mathbf{x})=\overline{f(\mathbf{x})}, \quad f\in\mathcal{H},$$

where the bar denotes complex conjugation.[8,9]

It is not necessary to invoke translational invariance explicitly because the pertinent operators act in the center-of-mass frame of the two-particle system. Furthermore, we do not gain anything by postulating invariance under spatial inversion because we impose rotational invariance and consider only spinless particles. Similarly, in the case of time-reversal invariance, because all particles are spinless, we need only consider the case in which θ is a conjugation.

Our first result provides a simple criterion for the equality of $\tilde{S}$ and S.

Theorem 2.1: If Ω_+ and W_+ (Ω_- and W_-) exist, then $\tilde{\Omega}_+$ ($\tilde{\Omega}_-$) exists. Moreover, if all four operators $\Omega_\pm$, $W_\pm$ exist, then $\tilde{S}=S$ if and only if

$$W_\pm^*SW_-=S. \tag{2.13}$$

Proof: By virtue of (2.3), (2.7), (2.10), and the unitarity of U_t, one concludes that

$$\tilde{\Omega}_t=W^*\Omega_tW_t. \tag{2.14}$$

From this result, together with (2.2), (2.6), (2.9), and the familiar theorem that the strong limit of an operator product is the product of the strong limits of its factors when all the latter limits exist, we obtain

$$\tilde{\Omega}_+=W^*\Omega_+W_+ \quad (\tilde{\Omega}_-=W^*\Omega_-W_-) \tag{2.15}$$

if Ω_+ and W_+ (Ω_- and W_-) exist.

If all the operators $\Omega_\pm$ and $W_\pm$ exist, we infer that

$$\tilde{S} = W_+^* S W_- \tag{2.16}$$

by invoking (2.8), (2.15), and the unitarity of W. The last assertion of the theorem follows directly from (2.16).

This theorem makes it plain that $W_\pm = I$ is sufficient to guarantee that $\tilde{S} = S$.

It is natural to ask under what circumstances $W_\pm$ are unitary and when do W_+ and W_- coincide, since these two properties involve obvious and desirable simplifications of the necessary and sufficient condition (2.13).

As to the first question, $W_\pm$ are isometric when they exist, since they are then strong limits of unitary operators. However, it is not known whether the mere existence of $W_\pm$ guarantees their unitarity in general. On the other hand, the following example (due to Ekstein[1]) shows that it is possible to have $W_+ \neq W_-$. Suppose that $\Omega_\pm$ exist and are unitary. Then $W_+ = I$, $W_- = S^*$ for $W = \Omega_+$, and $W_+ = S$, $W_- = I$ for $W = \Omega_-$. Hence $W_+ \neq W_-$ for the nontrivial case $S \neq I$ in this example.

We mention the following desirable, easily derivable, consequences of the unitarity of $W_\pm$ and of the equality of W_+ and W_-, similar remarks applying in Sec. 3 and 4. If $W_\pm$ are unitary, then $\tilde{S}$ is unitary if and only if S is unitary. Furthermore, if $\Omega_\pm$ exist and $W_\pm$ are unitary, then a necessary and sufficient condition for $\tilde{\Omega}_\pm$ to have the property of strong asymptotic completeness[10] is that $\Omega_\pm$ have this property. The following simple result implicitly specifies a class of operators W which are such that the corresponding transformed systems yield no scattering: If $S = I$ and $W_\pm$ are unitary, then $\tilde{S} = I$ if and only if $W_+ = W_-$. A more general result of this last type is stated in Theorem 2.4.

A sufficient condition for the unitarity of $W_\pm$ is given by the following theorem, which is a special case of Theorem A.1 of Appendix A.

Theorem 2.2: Let the spectrum of W be a proper subset of the unit circle. Then $W_\pm$ are unitary whenever they exist.

We have already mentioned a class of unitary operators W which satisfy the spectral condition of Theorem 2.2 and are also such that $W_\pm$ exist, viz., those for which $W - I$ is compact. In Sec. 5 we will discuss examples of unitary operators W which do not satisfy the spectral requirement in question, but nevertheless have the property that the corresponding $W_\pm$ exist and are unitary.

The next theorem furnishes further sufficient conditions for the unitarity of $W_\pm$. It will be convenient to use the terminology of von Neumann algebras[11,12] in its proof. Thus, if $\mathcal{N}$ denotes a set of closed linear operators (not necessarily bounded) with domains dense in $\mathcal{H}$ and ranges in $\mathcal{H}$, $\mathcal{N}'$ will denote the collection of all bounded operators from $\mathcal{H}$ to $\mathcal{H}$ which commute with each operator in $\mathcal{N}$, $\mathcal{N}'' = (\mathcal{N}')'$, etc.

Theorem 2.3: When $W_+(W_-)$ exists, the following statements hold: (i) If W commutes with L^2 and L_3, then $W_+(W_-)$ is unitary. (ii) If W is invariant under spatial rotations and time reversal, then W_+ and W_- are unitary operators such that

$$W_+^* = W_-.$$

Remarks: Part (i) of this theorem can be immediately generalized as follows. Suppose that $\{T_i,\ i \in \mathcal{I}\}$, where $\mathcal{I}$ is an arbitrary index set, is a set of self-adjoint operators such that each T_i commutes with H_0 and W, and that $\{H_0, T_i, i \in \mathcal{I}\}'$ is Abelian. Then $W_+(W_-)$ is unitary whenever it exists. Theorem 2.4 and the pertinent portions of Theorems 3.1 and 4.1 can be analogously generalized.

Proof of Theorem 2.3: We first note that the existence of W_+ or W_- entails its commutativity with H_0. This follows from (2.9), (2.10), and the fact that $\{U_t, -\infty < t < \infty\}$ is a unitary group.

As before, we shall only prove the theorem for W_+, whose existence is henceforth assumed in this proof.

Proof of (i): It is well-known that $\mathcal{A} = \{H_0, L^2, L_3\}'$ is a maximal Abelian von Neumann algebra ($\mathcal{A}' = \mathcal{A}$). This is a consequence of the fact that the von Neumann algebra $\{(H_0)_{lm}\}''$ generated by the restriction $(H_0)_{lm}$ of H_0 to any given $\mathcal{H}_{lm}$ is maximal Abelian. Here $\mathcal{H}_{lm}$ is the subspace of $\mathcal{H}$ spanned by the simultaneous eigenfunctions of L^2 and L_3 of given l and m. Now, $W_+ \in \mathcal{A}$ if W commutes with L^2 and L_3. On the other hand, since $\mathcal{A}$ is an Abelian von Neumann algebra, all of its operators are normal. Because W_+ is isometric, we therefore infer that it is unitary when W commutes with L^2 and L_3.

Proof of (ii): Let us confine our attention to a fixed l and m. From the maximal Abelian property of $\{(H_0)_{lm}\}''$, in particular, one deduces that the restriction $(W_+)_{lm}$ of W_+ to $\mathcal{H}_{lm}$ is in $\{(H_0)_{lm}\}''$ and, therefore, is a function (in the sense of the usual functional calculus) $F_{lm}(H_0)_{lm}$ of $(H_0)_{lm}$ whenever W commutes with L^2 and L_3.

Let W, and hence W_+, possess the stronger property of rotational invariance. Denote by $\{E(\lambda),\ -\infty < \lambda < \infty\}$ the spectral family of H_0, take arbitrary $f,\ g \in \mathcal{H}_{lm}$, and let $L_\pm = L_1 \pm iL_2$. There exist $f_0,\ g_0 \in \mathcal{H}_{l\bar{l}}$ such that $f_0 = L_-^k f$ and $g = L_+^k g_0$, where $\bar{l} = -l$ and $k = l + m$, and we have

$$(f, W_+ g) = (f, W_+ L_+^k g_0) = (f_0, W_+ g_0)$$

$$= \int_{-\infty}^{\infty} F_{l\bar{l}}(\lambda) d(f_0, E(\lambda)_+ g_0) = \int_{-\infty}^{\infty} F_{l\bar{l}}(\lambda) d(f, E(\lambda) g), \tag{2.17}$$

where we have used the commutativity of W_+ and of each $E(\lambda)$ with $L_\pm$. In other words, the function $F_{lm}(\lambda)$ corresponding to $(W_+)_{lm}$ is independent of m in the present case.

If W is time-reversal invariant, then (2.9), (2.10), the existence of W_+, and the antiunitarity of θ jointly entail that W_- exists and is given by

$$W_- = \theta W_+ \theta. \tag{2.18}$$

When W is rotationally and time-reversal invariant, we have for arbitrary $f,\ g \in \mathcal{H}_{lm}$, making use of (2.17) and (2.18), and of the antiunitarity of θ, its commuta-

tivity with $E(\lambda)$, and the fact that it leaves $\oplus_{\mu=-l}^{l}\mathcal{H}_{l\mu}$ invariant,

$$(f, W_-g) = \overline{(\theta f, W_-\theta g)} = \int_{-\infty}^{\infty} \overline{F_{l\bar{l}}(\lambda)}\, d\overline{(\theta f, E(\lambda)\theta g)}$$

$$= \int_{-\infty}^{\infty} \overline{F_{l\bar{l}}(\lambda)}\, d(f, E(\lambda)g) = (f, W_+^* g). \tag{2.19}$$

The assumed rotational invariance of W_+ implies that of W_+^*, and hence W_+ and W_+^* leave each $\mathcal{H}_{lm}$ invariant when W_+ has this property, whence (2.19) and the rotational invariance of W_+ yield (2.18).

A necessary and sufficient condition for the equality of S and $\tilde{S}$ when S is unitary is a particular case of the following theorem.

Theorem 2.4: Let H and W commute with L^2 and L_3, and, in addition, let S, W_+, and W_- exist. Then we have:

(i) If

$$W_+ = W_-, \tag{2.20}$$

then $\tilde{S} = S$.

(ii) If S is unitary and $\tilde{S} = S$, then (2.20) holds.

Proof: When $\Omega_\pm$ exist, the commutativity of H with L^2 and L_3 implies that S also has these commutation properties. Since S commutes with H_0 when it exists, the assumptions in the first sentence of the theorem therefore imply that $S \in \mathcal{A}$. If these assumptions are made, Theorem 2.3 guarantees the unitarity of $W_\pm$ and an argument in the proof of part (ii) of that theorem entails that $W_\pm \in \mathcal{A}$. It follows that

$$[S, W_\pm] = 0 \tag{2.21}$$

under the latter hypotheses.

Suppose now that the conditions in (ii) are also satisfied. Then (2.13) and the unitarity of W_+ imply

$$W_+ S = S W_-. \tag{2.22}$$

From (2.21), (2.22), and the unitarity of S, (2.20) follows, and therefore (ii) has been established.

To prove that (i) obtains, we observe that (2.22) follows from (2.20) and (2.21) and that (2.16), (2.22), and the isometry of W_+ imply (2.12).

An immediate consequence of Theorems 2.3 and 2.4 is stated in the following corollary.

Corollary 2.1: Let H be rotationally invariant and W rotationally and time-reversal invariant, let $W_\pm$ and S exist, and let S be unitary. Then a necessary and sufficient condition for the equality of $\tilde{S}$ and S is that

$$W_+ = W_- = I - 2P,$$

where P is a projection.

Proof: Notice, particularly, that it follows from the hypotheses of the corollary that $W_\pm$ are both unitary and self-adjoint.

To arrive at the previous results of this section, we have assumed that W_+ or W_- exists. The next theorem entails that the existence of W_+ and W_- is *necessary* for $\tilde{S}$ to equal S for a very large class of scattering systems. In that theorem, $R_+(R_-)$ denotes the range of $\Omega_+(\Omega_-)$. It is obvious that a necessary condition for the validity of either of the inequalities $\dim(\mathcal{H} \ominus R_\pm) < \infty$ pertinent in Theorem 2.5 is that H should possess at most a finite number of linearly independent bound states and that this condition is also sufficient when strong asymptotic completeness obtains for $\Omega_\pm$.

Theorem 2.5: Let Ω_+ and $\tilde{\Omega}_+$ (Ω_- and $\tilde{\Omega}_-$) exist, and let $\dim(\mathcal{H} \ominus R_+) < \infty$ [$\dim(\mathcal{H} \ominus R_-) < \infty$]. Then $W_+(W_-)$ exists.

Proof: We limit ourselves to the case of W_+.

Using (2.7) and the unitarity of W and V_t, we find

$$W_t = \Omega_t^* W \tilde{\Omega}_t,$$

which we write in the form

$$W_t = \Omega_t^* P_+ W \tilde{\Omega}_t + \Omega_t^* Q_+ W \tilde{\Omega}_t. \tag{2.23}$$

Here $Q_+ = I - P_+$.

We proceed to prove that each term on the right-hand side of (2.23) approaches a strong limit as $t \to \infty$ if the conditions that Ω_+ and $\tilde{\Omega}_+$ exist and that $\dim(\mathcal{H} \ominus R_+) < \infty$ are all fulfilled.

That this statement is true for the first such term follows by combining the theorem concerning strong limits of operator products already invoked in the proof of Theorem 2.1, with the circumstance that the existence of Ω_+ implies that of $\text{s-lim}_{t \to \pm\infty} \Omega_t^* P_+$.

As for the second of these terms, the identity

$$\Omega_t^* Q_+ W \tilde{\Omega}_t = U_t^* Q_+ W U_t \tag{2.24}$$

holds when Ω_+ exists, as can be shown by invoking, in particular, the commutativity of Q_+ with V_t. Now, the requirement that $\dim(\mathcal{H} \ominus R_+) < \infty$ entails the compactness of Q_+ and therefore that of $Q_+ W$. Since U_t is unitary and converges weakly to zero as $t \to \infty$, we infer that $U_t^* Q_+ W U_t$ converges strongly to zero in this limit when this finite dimensionality requirement is satisfied. This completes the proof of the theorem.

3. GENERALIZATION TO INCLUDE LONG-RANGE POTENTIALS WITHOUT HARD CORES

It is well known that the Møller wave operators (2.2) fail to exist for local potentials which are $O(|\mathbf{x}|^{-\alpha})$ for some $0 < \alpha \leq 1$ as $|\mathbf{x}| \to \infty$. For appropriate Coulomb-like potentials[13] and for much wider classes of long-range potentials,[6,14] it has been shown that the modified wave operators

$$\Omega_\pm = \text{s-lim}_{t \to \pm\infty} \Omega_t' \tag{3.1}$$

exist if the operator G_t mentioned below is appropriately selected. Here,

$$\Omega_t' = V_t^* U_t'. \tag{3.2}$$

In (3.2), V_t is given by (2.1) in terms of a self-adjoint extension H of $-\Delta + V$ on $C_0^\infty(R^3)$, where V denotes a suitable long-range potential. For the class of such potentials considered by Alsholm,[14] such a self-adjoint extension always exists. The operator U_t' in (3.2) is defined by

$$U_t' = U_t \exp(-iG_t), \tag{3.3}$$

where G_t is a self-adjoint operator of multiplication by a real-valued Lebesgue-measurable function $G_t(\mathbf{k})$ on the momentum-space representation of R^3. A definition of $G_t(\mathbf{k})$ is given in Appendix B. If $G_t(\mathbf{k})$ is defined in this manner, and as discussed more precisely in the latter appendix, the pertinent results of Ref. 14 entail that G_t possesses the following two properties for a large class of long-range potentials and that, in addition, the $\Omega_\pm$ in (3.1) exist for such potentials:

$$\text{s-}\lim_{t\to\pm\infty} \exp[i(G_{t+s}-G_t)]=I, \quad -\infty<s<\infty, \tag{3.4}$$

$$\text{w-}\lim_{t\to\pm\infty} U_t'=0. \tag{3.5}$$

Whenever (3.4) or (3.5) is assumed to obtain in subsequent discussions, this hypothesis will be mentioned explicitly.

When the present $\Omega_\pm$ exist, (3.4) is a necessary and sufficient condition for them to have the intertwining property

$$V_t\Omega_\pm=\Omega_\pm U_t, \tag{3.6}$$

as follows by a trivial generalization of a theorem of Prugovečki.[15]

We again consider a unitary operator W and define $\tilde{H}$ and $\tilde{V}_t$ by (2.4) and (2.5), respectively, in terms of this W and the operator H mentioned in the sentence following (3.2). The wave operators pertaining to this $\tilde{H}$ are given by

$$\tilde{\Omega}_\pm=\text{s-}\lim_{t\to\pm\infty}\tilde{\Omega}_t', \tag{3.7}$$

where

$$\tilde{\Omega}_t'=\tilde{V}_t^*U_t'=W^*V_t^*WU_t'.$$

In this section, the notation $W_\pm$ will refer to the operators

$$W_\pm=\text{s-}\lim_{t\to\pm\infty} W_t' \tag{3.8}$$

when these limits exist, where

$$W_t'=U_t'^*WU_t'. \tag{3.9}$$

It is thus clear that the existence of any of the operators $\Omega_\pm$, $\tilde{\Omega}_\pm$, or $W_\pm$ introduced in this section implies its isometry.

In the next theorem, we shall understand $R_\pm$, S, and $\tilde{S}$ to be defined in the same way as in Sec. 2, but with $\Omega_\pm$ and $\tilde{\Omega}_\pm$ specified by (3.1) and (3.7), respectively.

Theorem 3.1: In Theorems 2.1—2.5 and Corollary 2.1, let H, $\Omega_\pm$, $R_\pm$, $W_\pm$, $\tilde{\Omega}_\pm$, S, and $\tilde{S}$ be understood in the generalized sense of the present section. Then these theorems and corollary are true, provided that in the cases of the generalized Theorems 2.3 and 2.4 the additional condition (3.4) holds and that in the case of Theorem 2.5 the conditions (3.4) and (3.5) obtain.[16]

Proof: The generalizations of Theorems 2.1—2.5 can be proved by steps analogous to those followed in establishing the original versions of those theorems. As far as the generalizations of Theorems 2.3 and 2.4 are concerned, these steps include the fact that $W_\pm$, when they exist, commute with H_0. This commutativity follows from (3.3), together with the definition of G_t, (3.4), (3.8), (3.9), and obvious arguments. Naturally, (3.5) plays a role in our proof of the generalization of Theorem 2.5 parallel to that played by the corresponding property of U_t in the earlier proof of that theorem.

Assuming that S exists in the sense of the present section, the generalized Theorem 2.1 referred to in Theorem 3.1 entails that $W_\pm=I$ is a sufficient condition for $\tilde{S}=S$ to be true in the present sense, a result precisely analogous to one in Sec. 2. This sufficient condition is fulfilled when $W-I$ is compact, if (3.5) obtains in addition.

4. GENERALIZATION TO INCLUDE SHORT-RANGE POTENTIALS WITH HARD CORES

In this section we shall deal only with short-range potentials whose hard cores, if any, are velocity independent. This restriction is made in the present treatment to avoid cumbersome complications which would obscure the ensuing theoretical developments. Certain short-range potentials with velocity-dependent, or more properly angular-momentum-dependent, hard cores[17] can be treated by a straightforward extension of the formalism of this section.

This formalism can also be generalized to apply to a wide class of long-range potentials with hard cores by defining $\Omega_\pm$ as in (4.1) and (4.2) below, but with U_t in the latter equation replaced by an operator analogous to the "renormalized" operator U_t' of Sec. 3. As we shall show in a separate paper, these $\Omega_\pm$ exist for a large family of local hard-core potentials of long range. With this result and the definitions just sketched, the theorems of the present section can be readily extended to potentials of this family satisfying suitable conditions.

Let Γ denote an open subset of R^3 whose complement γ is compact and may be empty. If γ is of positive measure, we interpret it as that part of R^3 occupied by the velocity-independent hard core of interest. In this case, we assume that the boundary $\partial\Gamma$ or Γ is a closed surface of class C^2. The Hilbert space of relative motion of the two interacting particles is $K=\mathcal{L}^2(\Gamma)$, where, given any measurable subset M of R^3, $\mathcal{L}^2(M)$ is defined as that subspace of $H=L^2(R^3)$ whose elements vanish a.e. on the complement of M. In the generalization of the theory of Sec. 2 developed in this section, it is convenient to employ suitable spaces $\mathcal{L}^2(M)$, rather than the customary spaces $L^2(M)$.

Ikebe[18] and Hunziker[19] have developed time-dependent scattering theories for local short-range hard-core potentials for single channel and multichannel scattering systems, respectively. We have found the former formulation to be particularly convenient as a base for our discussion. Thus, for the case when γ is nonempty and $\partial\Gamma$ is of the above mentioned type, Ikebe[20] has proved that there exists a self-adjoint extension h^0 (called H in Ref. 20) of $-\Delta$ appropriate to the exterior domain Γ when hard-core boundary conditions are imposed on $\partial\Gamma$. This extension has domain dense in $L^2(\Gamma)$ and range in $L^2(\Gamma)$ and is explicitly characterized in Ref. 20. We define the self-adjoint operator $H^0=U^*h^0U$ with domain dense in $\mathcal{L}^2(\Gamma)$ and range in $\mathcal{L}^2(\Gamma)$, where

U is a unitary operator mapping of $L^2(\Gamma)$ onto $L^2(\Gamma)$ such that $(Uf)(\mathbf{x}) = f(\mathbf{x})$ a.e. on Γ for each $f \in L^2(\Gamma)$. Similarly, in the present section, H will denote a self-adjoint operator $H^0 + V$ whose domain and range have the properties just mentioned, V being an appropriate potential.

To allow for the presence of hard cores, we define wave operators $\Omega_\pm$ for the system with Hamiltonian H in the same way as proposed in Ref. 18, except that we permit γ to be empty and for trivial modifications occasioned by our use of K. When they exist, these $\Omega_\pm$ lead to the same S operator as do the corresponding wave operators defined in the sense of the latter reference. We set

$$\Omega_\pm = \text{s-}\lim_{t \to \pm\infty} \Omega_t \tag{4.1}$$

whenever these limits exist, where

$$\Omega_t = V_t^* P U_t.$$

Here and for the remainder of this section, the operator $V_t : K \to K$ is defined by

$$V_t = \exp(-itH),$$

and P is a projection operator with domain H and range K.

Using, in particular, results and methods in Refs. 18 and 20, known results on self-adjointness and on the existence of wave operators for appropriate short-range potentials without hard cores, and straightforward arguments, one readily arrives at the following conclusions. Let V be a multiplication operator by a real-valued function in some class $L^2(\Gamma) + L^p(\Gamma) \cap L^\infty(R^3)$ [$L^p(\Gamma)$ is defined analogously to $L^2(\Gamma)$] or let V be a self-adjoint operator of finite rank. Define H as $H^0 + V$. Then H is self-adjoint and the wave operators (4.1) exist for this H. Local potentials similar to the first class mentioned above are considered in Ref. 19 for multichannel scattering.

Let $\tilde{\Gamma}$ denote a subset of R^3 whose complement $\tilde{\gamma}$ is compact and may be void. We set $\tilde{K} = L^2(\tilde{\Gamma})$ and denote by $\tilde{P}$ the projection with domain H and range $\tilde{K}$. We let W stand for a unitary operator from $\tilde{K}$ onto K and, of course, define the Hamiltonian $\tilde{H}$ of the transformed system by

$$\tilde{H} = W^* H W.$$

In analogy with the case of γ, $\tilde{\gamma}$ is to be interpreted as the region occupied by the hard core of $\tilde{H}$ if $\tilde{\gamma}$ has positive measure. Notice that the elements of the domain of $\tilde{H}$ do not generally obey the usual hard-core boundary conditions. Specifically, and contrary to the case of the elements of the domain of the operator H in the last paragraph, they need not be equivalent to continuous functions on R^3 tending to zero when $\mathbf{x} \in R^3$ tends to an arbitrary point of $\partial\Gamma$.

If they exist, the wave operators for the transformed system are

$$\tilde{\Omega}_\pm = \text{s-}\lim_{t \to \pm\infty} \tilde{\Omega}_t, \tag{4.2}$$

where

$$\tilde{\Omega}_t = \tilde{V}_t^* \tilde{P} U_t.$$

The operator $\tilde{V}_t : \tilde{K} \to \tilde{K}$ is defined by

$$\tilde{V}_t = \exp(-it\tilde{H}) = W^* V_t W.$$

Provided that the limits exist, we set

$$W_\pm = \text{s-}\lim_{t \to \pm\infty} W_t,$$

with

$$W_t = U_t^* W \tilde{P} U_t.$$

By means of an argument in Ref. 18 which exploits, in particular, the compactness of γ and $\tilde{\gamma}$, one easily shows that the operators $\Omega_\pm$, $\tilde{\Omega}_\pm$, and $W_\pm$ of this section are isometric when they exist.

We shall now generalize the results of Sec. 2, with the exception of Theorem 2.2, to apply to the type of hard core under discussion. Naturally, in the remainder of this section $R_\pm$ will stand for the ranges of the operators $\Omega_\pm$ in (4.1) and the scattering operators S and $\tilde{S}$ will be defined by (2.8) in terms of the $\Omega_\pm$ and $\tilde{\Omega}_\pm$ in (4.1) and (4.2), respectively.

Theorem 4.1: In Theorems 2.1, 2.3—2.5, and Corollary 2.1, let H, $\Omega_\pm$, $R_\pm$, $W_\pm$, $\tilde{\Omega}_\pm$, S, and $\tilde{S}$ be given by the generalized definitions of the present section. Then these theorems and corollary hold, if in the case of Theorems 2.3 and 2.4 and Corollary 2.1 the operators H and W are replaced by HP and $W\tilde{P}$, respectively, and in that of Theorem 2.5 the conditions $\dim(H \ominus R_\pm) < \infty$ are changed to $\dim(K \ominus R_\pm) < \infty$.

Proof: The generalizations of Theorems 2.1 and 2.3—2.5 can be proved similarly to the way in which these theorems were established earlier. In particular, comments analogous to those in the second and fourth sentences of the proof of Theorem 3.1 apply here.

The conditions that H commute with L^2 and L_3, imposed in Theorem 4.1, are both fulfilled if each of the following requirements is satsified: (i) γ is a finite closed sphere of positive radius centered at the origin; (ii) H is a self-adjoint operator $H^0 + V$, where V is a potential of one of the two types mentioned in the paragraph immediately after the one containing (4.1) and, moreover, commutes with L^2 and L^3. Examples of operators W having the properties assumed in Theorem 4.1 will be given in Sec. 5.

5. EXAMPLES

All of the examples of operators W considered in this section are such that the S operators corresponding to the Hamiltonians H and $\tilde{H}$ of the relevant original and transformed scattering systems are equal. In subsection A we discuss a W for which the operators $W_\pm$ defined in Sec. 2 are not equal to I. In subsection B we deal with two cases in which W is a Bohm—Gross—Baker transformation, a name chosen to honor three pioneers in the study of such transformations.[21] In these two cases, which include the possibility of velocity-independent hard cores, we show that the pertinent operators $W_\pm$, understood in the sense of Sec. 4, are equal to I. It is an easy matter to generalize the second of these transformations (W_2) so that it also applies in suitable cases involving angular-momentum-dependent hard cores, while preserving the last mentioned property of the pertinent $W_\pm$.

A. Example for which $W_\pm \neq I$

In the absence of an explicit indication to the contrary, all operators occurring in this subsection have domains dense in $\mathcal{H} = L^2(R^3)$ and ranges in $\mathcal{H}$.

The following statements hold:

1. Let H be a self-adjoint operator and let S, the scattering operator pertaining to H in the sense of Sec. 2, exist. Denote by W_0 a unitary operator commuting with H_0 and S and different from I. If $\tilde{S}$ is the scattering operator pertaining to the transformed Hamiltonian $W_0^* H W_0$ in the sense just specified, then $\tilde{S} = S$.

2. Suppose that S and $\tilde{S}$ are defined as indicated in Sec. 3, but that all of the remaining definitions and assumptions in Statement 1 are made and that the U_t' considered is a function of H_0. Then $\tilde{S} = S$.

Statement 1 follows immediately by using the assumed properties of W_0, the commutativity of S with H_0, and Theorem 2.1, and Statement 2 follows by similar arguments, including the use of Theorem 3.1. In connection with the latter statement, notice that U_t' can be chosen as a function of H_0 for local potentials V for which (B1)—(B3) of Appendix B obtain and which, moreover, are such that $V_L(\mathbf{x})$ in (B2b) is spherically symmetric.

Clearly, one has $W_\pm = W_0 \neq I$ for the operators $W_\pm$ corresponding to $W = W_0$ in the sense of Secs. 2 or 3, provided that the U_t' involved is a function of H_0 when these $W_\pm$ are understood in the sense of this last section.

Let S be as prescribed in Statements 1 or 2. Then an obvious example of such a W_0 is a unitary operator W_0' which is a function of H_0 and S, i.e., $W_0 \in \{H_0, S\}''$, and which differs from I. If, in addition, the S in question commutes with L^2 and L_3 and W_0' is also required to be rotationally invariant, then W_0' reduces to the form

$$W_0'' = \sum_{l=0}^{\infty} F_l P_l,$$

as can be seen by using arguments similar to ones invoked in the proof of Theorem 2.3. For each l, F_l is a unitary operator which is a function of H_l, and P_l is the projection onto $\mathcal{H}_l = \oplus_{m=-l}^{l} \mathcal{H}_{lm}$.

Notice that W_0'' is time-reversal invariant if and only if each F_l is of the form

$$F_l = I - 2G_l,$$

where, for every l, G_l is a projection operator which is a function of H_0, whence, by a theorem of Stone[22] and straightforward reasoning, we may write for each l without loss of generality

$$(G_l f)\hat{}(\mathbf{k}) = \chi_l(|\mathbf{k}|)\hat{f}(\mathbf{k}),$$

for $f \in \mathcal{H}_l$, where $\hat{g}$ is the Fourier transform of $g \in \mathcal{H}$ and χ_l the characteristic function of some measurable subset of $[0, \infty)$.

B. Bohm-Gross-Baker transformations

These are the operators W_1 and W_2 which we proceed to specify. The first applies to spherical and large classes of nonspherical hard cores of the original and transformed systems, while the second applies only when the hard cores, if any, of both of these systems are spheres centered at the origin. However, W_2 is a subcase of W_1 if and only if all the functions g_l defined below are the same.

To define W_1, we consider a pair of open sets Γ_1 and $\tilde{\Gamma}_1$ whose complements are compact and we introduce the following function:

I. $\mathbf{h}$ is a function from Γ_1 into $\tilde{\Gamma}_1$ which maps in a one—one manner an open subset $\mathcal{O} \subset \Gamma_1$ onto a subset $\tilde{\mathcal{O}} \subset \tilde{\Gamma}_1$, where the complement of $\mathcal{O}(\tilde{\mathcal{O}})$ in $\Gamma_1(\tilde{\Gamma}_1)$ has measure zero. Furthermore, for each $\mathbf{x} \in \mathcal{O}$, all the cartesian components of $\mathbf{h}(\mathbf{x})$ are continuously differentiable with respect to each cartesian component of $\mathbf{x}$ and the Jacobian $J(\mathbf{x})$ of the mapping $\mathbf{x} \mapsto h(\mathbf{x})$ is nonzero.

Let $K_1 = L^2(\Gamma_1)$, $\tilde{K}_1 = L^2(\tilde{\Gamma}_1)$, and define $W_1 : \tilde{K}_1 \to K_1$ by

$$(W_1 f)(\mathbf{x}) = \begin{cases} \sigma(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) & \text{a.e. on } \Gamma_1, \\ 0 & \text{a.e. on } \Gamma_1^c, \end{cases} \tag{5.1}$$

for each $f \in \tilde{K}_1$, where $\sigma(\mathbf{x}) = |J(\mathbf{x})|^{1/2}$ on $\mathcal{O}$ and, for any subset A of R^3, A^c signifies the complement of A.

To define W_2, let Γ_2 denote either the set $\{\mathbf{x} \in R^3 : |\mathbf{x}| > a\}$ for some $0 < a < \infty$ or R^3, and let $\tilde{\Gamma}_2$ denote either $\{\mathbf{x} \in R^3 : |\mathbf{x}| > b\}$ for some $0 < b < \infty$ or R^3. For every l, $\mathbf{h}_l$ is a radially symmetric transformation satisfying the conditions which we now state:

II. $\mathbf{h}_l$ is a one-one transformation from Γ_2 onto $\tilde{\Gamma}_2$ of the form

$$\mathbf{h}_l(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} = 0, \\ [g_l(|\mathbf{x}|)/|\mathbf{x}|]\mathbf{x}, & \mathbf{x} \in \Gamma_2 \setminus \{0\}, \end{cases} \tag{5.2}$$

where

$$g_l(r) = \beta + \int_\alpha^r \rho_l(\xi)\, d\xi, \quad r \geq \alpha.$$

Here $\alpha = a (\alpha = 0)$ when $\Gamma_2 \neq R^3$ ($\Gamma_2 = R^3$), $\beta = b$ ($\beta = 0$) when $\tilde{\Gamma}_2 \neq R^3$ ($\tilde{\Gamma}_2 = R^3$), $\rho_l \in L^1([\alpha, \infty))$, and $\rho_l(r) > 0$ a.e. on $[\alpha, \infty)$.

Let $K_2 = L^2(\Gamma_2)$, $\tilde{K}_2 = L^2(\tilde{\Gamma}_2)$, and define $K_{2l}(\tilde{K}_{2l})$ as the subset of $\mathcal{H}_l$ whose elements f have the property $f(\mathbf{x}) = 0$ a.e. on $\Gamma_2^c(\tilde{\Gamma}_2^c)$. It is clear that the subspaces $K_{2l}(\tilde{K}_{2l})$ are pairwise orthogonal and that

$$K_2 = \bigoplus_{l=0}^{\infty} K_{2l}, \qquad \tilde{K}_2 = \bigoplus_{l=0}^{\infty} \tilde{K}_{2l}.$$

For each $r = 1, 2$, we denote by $P_r(\tilde{P}_r)$ the projections from $\mathcal{H}$ onto $K_r(\tilde{K}_r)$ and, for every l, $P_{2l}(\tilde{P}_{2l})$ signify the projections from $\mathcal{H}$ onto $K_{2l}(\tilde{K}_{2l})$.

The operator $W_2 : \tilde{K} \to K_2$ is now defined by

$$W_2 = \bigoplus_{l=0}^{\infty} W_{2l},$$

where the $W_{2l} : \tilde{K}_{2l} \to K_{2l}$ are given by

$$(W_{2l} g)(\mathbf{x}) = \begin{cases} \sigma_l(\mathbf{x}) g(\mathbf{h}_l(\mathbf{x})) & \text{a.e. on } \Gamma_2, \\ 0 & \text{a.e. on } \Gamma_2^c, \end{cases} \tag{5.3}$$

for each $g \in \tilde{K}_{2l}$. Here $\sigma_l(\mathbf{x}) = |J_l(\mathbf{x})|^{1/2}$ a.e. on Γ_2, $J_l(\mathbf{x})$ being the Jacobian of the transformation $\mathbf{x} \mapsto \mathbf{h}_l(\mathbf{x})$. One has

$$\sigma_l(\mathbf{x}) = \rho_l^{1/2}(|\mathbf{x}|) g_l(|\mathbf{x}|)/|\mathbf{x}| \tag{5.4}$$

a.e. on Γ_2.

It follows directly from the definitions that

$$W_2 \tilde{\rho}_2 = \sum_{l=0}^{\infty} W_{2l} \tilde{\rho}_{2l}. \tag{5.5}$$

The operators W_1 and W_2 can be shown to be unitary by invoking standard theorems on changes of variables in Lebesgue integrals and other equally familiar theorems of integration theory. It is easily seen that $W_1\tilde{\rho}_1$ and $W_2\tilde{\rho}_2$ are time-reversal invariant and that $W_2\tilde{\rho}_2$ is rotationally invariant.

The principal objective of this subsection is to show that

$$\text{s-}\lim_{t\to\pm\infty} U_t^* W_r \tilde{\rho}_r U_t = I \tag{5.6}$$

for $r = 1, 2$, provided that certain conditions which are physically very natural are satisfied. In general, these conditions do not entail that $W_r\tilde{\rho}_r - I$ is compact for these r.

Theorem 5.1: Let h satisfy I and also each of the following requirements:

$$\sup_{\mathbf{x}\in\Gamma_1} |\mathbf{h}(\mathbf{x}) - \mathbf{x}| < \infty, \tag{5.7a}$$

$$\lim_{|\mathbf{x}|\to\infty} |\mathbf{h}(\mathbf{x}) - \mathbf{x}| = 0, \tag{5.7b}$$

$$\sup_{\mathbf{x}\in\Gamma_1} \sigma(\mathbf{x}) < \infty, \tag{5.7c}$$

$$\lim_{|\mathbf{x}|\to\infty} \sigma(\mathbf{x}) = 1. \tag{5.7d}$$

Then (5.6) holds for $r = 1$.

Moreover, if each of the $\mathbf{h}_l$ fulfills II and also each of the conditions (5.7) with $\mathbf{h}(\mathbf{x})$, $\sigma(\mathbf{x})$, and Γ_1 replaced by $\mathbf{h}_l(\mathbf{x})$, $\sigma_l(\mathbf{x})$, and Γ_2, respectively, then (5.6) obtains for $r = 2$.

Remarks: Suppose that h fulfills the conditions of the first paragraph of the theorem and that, in addition, $\mathbf{h}(\mathbf{x}) \neq \mathbf{x}$ only when $\mathbf{x}$ lies in a bounded subset of U. Then the first assertion of the theorem follows in a substantially simpler fashion than in the proof below, by arguments which include exploiting the unitarity of W_1 and the familiar fact that $\lim_{t\to\pm\infty}\|\Lambda U_t f\| = 0$ for each f in a suitable dense subset of $\mathcal{H}$ when Λ is a projection from $\mathcal{H}$ onto $L^2(N)$, N being a bounded and measurable subset of R^3.

It is easily shown from (5.2) and (5.4) that the conditions (5.7) imposed on the $\mathbf{h}_l$ in the second paragraph of Theorem 5.1 are obeyed if each of the requirements

$$\lim_{r\to\infty}\{g_l(r) - r\} = 0,$$

$$\lim_{r\to\infty} \rho_l(r) = 1,$$

$$\sup_{r\in\mathcal{Y}} \rho_l(r) < \infty,$$

is satisfied for every l, where $\mathcal{Y}$ denotes $[a, \infty)$ ($[0, \infty)$) when $\Gamma_2 \neq R^3$ ($\Gamma_2 = R^3$).

Proof of Theorem 5.1: To prove the assertion pertaining to $r = 1$ it is sufficient to show that

$$\text{s-}\lim_{t\to\pm\infty} \|(W_1\tilde{\rho}_1 - I)U_t\varphi_a\| = 0, \quad a \in R^3, \tag{5.8}$$

where

$$\varphi_a(\mathbf{x}) = \exp(-\tfrac{1}{2}|\mathbf{x} - \mathbf{a}|^2).$$

Indeed, (5.8) is equivalent to

$$\text{s-}\lim_{t\to\pm\infty} (W_1\tilde{\rho}_1 - I)U_t = 0, \tag{5.9}$$

by virtue of the uniform boundedness of the operators of the family $\{(W_1\tilde{\rho}_1 - I)U_t, \ -\infty < t < \infty\}$ and of the known fact that the set $\{\varphi_a, a \in R^3\}$ is dense in $L^2(R^3)$. In turn, (5.9) is clearly equivalent to (5.6) in the case when $r = 1$.

To derive (5.8), we first observe that a straightforward computation, using, in particular, (5.1) and a familiar formula for $(U_t\varphi_a)(\mathbf{x})$, yields

$$\begin{aligned}\|(W_1\tilde{\rho}_1 - I)U_t\varphi_a\|^2 &= \int_{R^3} \exp(-|\mathbf{y}|^2)\left\{\sigma_a^2(\mathbf{y}, t)\exp\left[-\frac{\lambda_a(\mathbf{y}, t)}{1+4t^2}\right]\right. \\ &\left. - 2\sigma_a(\mathbf{y}, t)\exp\left[-\frac{\lambda_a(\mathbf{y}, t)}{2(1+4t^2)}\right]\cos\left[\frac{t\lambda_a(\mathbf{y}, t)}{2(1+4t^2)}\right] + 1\right\},\end{aligned} \tag{5.10}$$

where we have introduced the change in variables

$$\mathbf{y} = (1 + 4t^2)^{-1/2}(\mathbf{x} - \mathbf{a}) \tag{5.11}$$

and have set

$$\begin{aligned}&\sigma_a(\mathbf{y}, t) = \sigma(\mathbf{x}), \\ &\lambda_a(\mathbf{y}, t) = |\mathbf{h}(\mathbf{x}) - \mathbf{x}|^2 + 2(\mathbf{x} - \mathbf{a})\cdot[\mathbf{h}(\mathbf{x}) - \mathbf{x}], \ \mathbf{x} \in \Gamma_1, \\ &\sigma_a(\mathbf{y}, t) = \lambda_a(\mathbf{y}, t) = 0, \quad \mathbf{x} \in \Gamma_1^c.\end{aligned} \tag{5.12}$$

Invoking (5.7b), (5.7d), (5.11), and (5.12), we conclude that $\sigma_a(\mathbf{y}, t) \to 1$ and $\lambda_a(\mathbf{y}, t)/(1 + 4t^2) \to 0$ for given $\mathbf{y}, \mathbf{a} \in R^3$ as $t \to \pm\infty$. Hence it only remains to show that it is permissible to interchange $\lim_{t\to\pm\infty}$ with the integration in (5.10) in order to complete the proof of (5.8).

Now, $M = \sup_{\mathbf{x}\in\Gamma_1}\sigma(\mathbf{x})$ and $N = \sup_{\mathbf{x}\in\Gamma_1}|\mathbf{h}(\mathbf{x}) - \mathbf{x}|$ are finite by hypothesis. From (5.11), (5.12), and obvious estimates, we therefore conclude that the integrand of (5.10) is bounded in absolute value on R^3 by the t-independent, R^3-summable function $\{M\exp[N(N + 2|\mathbf{y}|)] + 1\}^2 \times \exp(-|\mathbf{y}|^2)$ for $-\infty < t < \infty$. The legitimacy of the interchange of the limits in question now follows from the bounded convergence theorem.

For our proof of the assertion pertaining to $r = 2$ we consider, for each l, the operator $\tilde{W}_l : \mathcal{K}_2 \to \mathcal{K}_2$ defined by

$$(\tilde{W}_l g)(\mathbf{x}) = \begin{cases} \sigma_l(\mathbf{x}) g(\mathbf{h}_l(\mathbf{x})), & \text{a.e. on } \Gamma_2, \\ 0, & \text{a.e. on } \Gamma_2^c, \end{cases} \tag{5.13}$$

for each $g \in \mathcal{K}_2$. From (5.3), (5.5), and (5.13), it is clear that W_{2l} is the restriction of $\tilde{W}_l$ to $\mathcal{K}_{2l}$ for any given l.

From arguments essentially identical to previous ones in this proof, we infer that our hypotheses for $r = 2$ entail that

$$\text{s-}\lim_{t\to\pm\infty} (W_l\tilde{\rho}_2 - I)U_t = 0 \tag{5.14}$$

for every l. Using, among other facts, the ones that

the $\mathcal{H}_l$ reduce U_t and that $K_{2l} \subset \mathcal{H}_l$ for each l, the relation between W_l and W_{2l}, and (5.5), one concludes that (5.14) implies that

$$\text{s-}\lim_{t\to\pm\infty}(W_2\tilde{\rho}_2 - I)U_t = 0.$$

The truth of our assertion for $r=2$ immediately follows from this result.

ACKNOWLEDGMENTS

We wish to thank H. Ekstein, W.O. Amrein, W. Hunziker, J.-P. Marchand, P. Rejto, C. Chandler, A.G. Gibson, J.V. Corbett, I. Herbst, and B. Simon for discussions on foundational physical and mathematical topics concerning this work, and M.I. Haftel and M. Rosen for conversations on its nuclear physics applications. One of us (WWZ) would like to thank Professor A.S. Wightman for his kind hospitality at Princeton University.

APPENDIX A: THEOREM ON STRONG LIMITS OF SEQUENCES OF CERTAIN UNITARY OPERATORS

Let $W: \mathfrak{H} \to \mathfrak{H}$ and each $U_n: \mathfrak{H} \to \mathfrak{H}$ be unitary operators, where n runs over the positive integers and $\mathfrak{H}$ is an arbitrary separable Hilbert space. Define

$$\mathcal{W} = \text{s-}\lim_{n\to\infty} W_n \tag{A1}$$

whenever this limit exists, where

$$W_n = U_n^* W U_n \tag{A2}$$

for each n.

For an arbitrary operator $B: \mathfrak{H} \to \mathfrak{H}$ let $\sigma(B)$ be its spectrum and $N(B)$ the closure of its numerical range.[23] Then we have

Theorem A.1: Let $\sigma(W)$ be a proper subset of the unit circle. Then $\mathcal{W}$ is unitary whenever it exists.

Proof: Since $\sigma(W)$ is a closed subset of the unit circle C, we may suppose that $\sigma(W)$ lacks some closed arc of positive length. In particular, it will be sufficient to prove the theorem under the assumption that $\sigma(W)$ lacks an arc of the form

$$\{z \in C : |\arg z| \leq \theta_0 \text{ for some } 0 < \theta_0 < \pi\}, \tag{A3}$$

since all other cases in which a closed arc of positive length is missing from $\sigma(W)$ can be reduced to the case (A3) by multiplying W by an appropriate unimodular complex number.

If $\mathcal{W}$ exists, it is isometric. We shall prove that $\sigma(\mathcal{W})$ does not coincide with the closed unit disc $|z| \leq 1$. This will show that $\mathcal{W}$ is unitary by virtue of a theorem[24] on the spectra of isometric operators.

Take $z_0 \in N(\mathcal{W})$. The existence of $\mathcal{W}$ is easily seen to entail the existence of a positive integer $N_0(\epsilon)$ such that the distance between z_0 and $N(W_n)$ is less than ϵ for all $n \geq N_0(\epsilon)$. Now, it follows from the fact that the numerical range of an operator is a unitary invariant[25] that $N(W_n) = N(W)$ for each n. Hence, the distance between z_0 and $N(W)$ is zero. This and the fact that $N(W)$ is closed imply that $z_0 \in N(W)$ and therefore that

$$N(\mathcal{W}) \subset N(W). \tag{A4}$$

Consider the closed, convex, proper subset $\mathcal{J}$ of the disc $|z| \leq 1$ which is the intersection of this disc with the half-plane $\mathrm{Re}\, z \leq \cos\theta_0 < 1$. Since W is normal, $N(W)$ is the smallest closed convex set containing $\sigma(W)$,[26] so that

$$N(W) \subset \mathcal{J} \tag{A5}$$

because $\mathcal{J} \supset \sigma(W)$.

The isometry of $\mathcal{W}$ implies $\sigma(\mathcal{W}) \subset N(\mathcal{W})$[27] and we therefore infer that $\sigma(\mathcal{W}) \subset \mathcal{J}$ from (A4) and (A5). Hence, $\sigma(\mathcal{W})$ does not coincide with the disc $|z| \leq 1$, and the proof is complete.

APPENDIX B: WAVE OPERATORS AND RELATED OPERATORS FOR LOCAL LONG-RANGE POTENTIALS

In this appendix, we summarize for the reader's convenience some results concerning modified wave operators and other operators for a wide class of long-range potentials. Most of these results are simplifications of conclusions of Alsholm.[14]

We shall be concerned with a local potential V such that

$$V = V_S + V_L, \tag{B1}$$

V_S and V_L being multiplication operators in $L^2(R^3)$ by real-valued functions $V_S(\cdot)$ and $V_L(\cdot)$, respectively, where

$$(1 + |\mathbf{x}|)^{1+\epsilon} V_S(\mathbf{x}) \in L^2(R^3) + L^\infty(R^3), \tag{B2a}$$

$$|\nabla^p V_L(\mathbf{x})| \leq c(1 + |\mathbf{x}|)^{-p-\alpha}, \quad p = 0, \ldots, 2m, \tag{B2b}$$

for some ϵ, m, and α such that

$$\epsilon > 0, \quad m \geq 1, \quad (m+1)^{-1} < \alpha \leq 1. \tag{B3}$$

Here m is an integer and c a constant independent of the other constants. It is clear that if the decomposition (B1) of a given V into short-range and long-range parts V_S and V_L satisfying (B2a) and (B2b) for values of the constants mentioned above is possible it is not unique. This fact leads to nonunique wave operators and is a well-recognized property of long-range potentials.[28] Hereafter in this appendix, until further notice, V, V_S, and V_L will denote arbitrary fixed operators fulfilling all the requirements stated above.

Since $V_S(\cdot) + V_L(\cdot)$ is a real-valued function in $L^2(R^3) + L^\infty(R^3)$, it follows that the operator $H = H_0 + V$ is self-adjoint.[29] We define V_t just as before in terms of this H and also define self-adjoint operators G_t by

$$(G_t^{(j)} f)\hat{}\,(\mathbf{k}) = G_t^{(j)}(\mathbf{k})\hat{f}(\mathbf{k}),$$

at each $f \in \mathcal{H}$ in their respective domains. The functions $G_t^{(j)}(\mathbf{k})$ exist and are specified by the recursive formulas

$$G_t^{(0)}(\mathbf{k}) = 0$$

$$G_t^{(j)}(\mathbf{k}) = \int_0^t V_L(s\mathbf{k} + \nabla_k G_t^{(j-1)}(\mathbf{k}))\, ds,$$

for $j = 0, \ldots, 2m$, in particular, if m is a positive integer such that (B2b) is satisfied.

From the next two theorems, it follows that for the pertinent class of long-range potentials there exist op-

erators G_t satisfying (3.4) and (3.5) and that, in addition, the modified wave operators (3.1) pertaining to these G_t exist.

Theorem B.1.[30]: Let m and α be such that (B2b) and (B3) obtain. Then

$$\text{s-}\lim_{t\to\pm\infty} \exp\{i[G_{t+s}^{(m)} - G_t^{(m)}]\} = I, \quad -\infty < s < \infty,$$

and

$$\text{w-}\lim_{t\to\pm\infty} U_t \exp(-iG_t^{(m)}) = 0.$$

Theorem B.2 [31]: Let $\epsilon, m,$ and α be such that (B2a) – (B3) hold. Then the modified wave operators

$$\Omega_\pm^{(m)} = \text{s-}\lim_{t\to\pm\infty} V_t^\dagger U_t \exp(-iG_t^{(m)})$$

exist.

To apply these theorems in the important special case when V is an operator of multiplication in $L^2(R^3)$ by a real-valued function

$$V(\cdot) = \tilde{V}_S(\cdot) + \tilde{V}_L(\cdot)$$

where $V_S(\cdot)$ satisfies (B2a) for some $\epsilon > 0$ and

$$\tilde{V}_L(\mathbf{x}) = c|\mathbf{x}|^{-\alpha}$$

for some $0 < \alpha \leq 1$, we write

$$V(\cdot) = V_S(\cdot) + V_L(\cdot),$$

with

$$V_S(x) = \tilde{V}_S(x) + c[(\mathbf{x})^{-\alpha} - (1 + |\mathbf{x}|)^{-\alpha}],$$
$$V_L(x) = c(1 + |\mathbf{x}|)^{-\alpha}. \tag{B4}$$

Plainly, $V_S(\cdot)$ and $V_L(\cdot)$ in (B4) are such that V_S obeys (B2a) and $V_L(\cdot)$ satisfies the inequality (B2b) at all non-negative integers p for the present ϵ, c, and α.

[1]H. Ekstein, Phys. Rev. **117**, 1590 (1960).
[2]F. Coester, S. Cohen, B. Day, and C.M. Vincent, Phys. Rev. C 1, 769 (1970).
[3]We will cite only a few papers in this rapidly expanding field from which further references can be traced: P. Mittelstaedt, Acta Phys. Acad. Sci. Hung. **19**, 303 (1965); M. Ristig, Z. Phys. **199**, 325 (1967); M. Ristig and S. Kistler, Z. Phys. **215**, 419 (1968); M.I. Haftel and F. Tabakin, Phys. Rev. C 3, 921 (1971); F. Tabakin, in *The Two-Body Force in Nuclei*, edited by S.M. Austin and G.M. Crawley (Plenum, New York, 1972); C.W. Wong and T. Sawada, Ann. Phys. (N.Y.) **72**, 107 (1972); A. Mann, I. Unna, and R. Gross, Phys. Rev. C 6, 415 (1972); H.S. Picker and J.P. Lavine, Phys. Rev. C 6, 1542 (1972); E.P. Harper, Y.E. Kim, and A. Tubis, Phys. Rev. C 6, 1601 (1972); M.I. Haftel, Phys. Rev. C 7, 80 (1973); J.P. Vary, Phys. Rev. C 7, 521 (1973); J.P. Lavine and G.J. Stephenson, Jr., Phys. Rev. C 9, 2095 (1974); P.U. Sauer, Phys. Rev. C **11**, 1786 (1975).
[4]For the definition of H as a quadratic form, see B. Simon, *Quantum Mechanics for Hamiltonian defined as Quadratic Forms* (Princeton U.P., Princeton, N.J., 1971).
[5]If nothing to the contrary is mentioned, all subsequent statements and equations involving operators with the subscript t should be understood to hold for each $-\infty < t < \infty$.
[6]For a summary of these results and for further information concerning two-body time-dependent scattering we refer to the review by W.O. Amrein, in *Scattering Theory in Mathematical Physics*, edited by J. LaVita and J.-P. Marchand (Reidel, Dordrecht, Holland, 1974).
[7]A necessary and sufficient condition for $W_\pm = I$ is that $W - I$ be in the closure of compact operators with respect to H_0-convergence. See Definition 2 and Theorem 1 of H. Baumgärtel, Math. Nachr. 58, 279 (1973). We thank W.O. Amrein for this reference.
[8]In this paper, equations involving real or complex-valued functions on R^3 or $[0,\infty)$, considered at some value $\mathbf{x}$, k, r, of the pertinent independent variable, hold a.e., relative to Lebesgue measure of the appropriate dimensionality, on the respective domains, unless stated otherwise. The terms a.e. and a.a. will always refer to the latter measure.
[9]A rigorous treatment of time-reversal invariance in the context of time-dependent scattering theory has been given by W.O. Amrein, Ph. A. Martin, and B. Misra, Helv. Phys. Acta 43, 313 (1970).
[10]See Ref. 4, p. 99.
[11]J. Dixmier, *Les algèbres d'operateurs dans d'espace Hilbertien (algèbres de von Neumann)* (Gauthier-Villars, Paris, 1969), 2nd ed.
[12]J.M. Jauch, Helv. Phys. Acta 33, 711 (1960).
[13]J.D. Dollard, J. Math. Phys. 5, 729 (1964); Rocky Mount. J. Math. 1, 5 (1971); Erratum: 2, 317 (1972).
[14]P.K. Alsholm, "Wave Operators for Long-Range Scattering," dissertation, Univ. California, Berkeley, 1972.
[15]E. Prugovečki, Nuovo Cimento B 4, 105 (1971), Theorem 4.1.
[16]The conditions $\dim(\mathcal{H} \ominus R_\pm) < \infty$ imposed in the generalized Theorem 2.5 are obeyed in the most obvious application of the results of this section to nuclear physics—that of proton—proton scattering. However, these conditions are not particularly useful for long-range potentials which are purely attractive beyond a certain distance, since the corresponding bound-state subspaces are infinite-dimensional under the rather weak assumptions stated by B. Simon, Helv. Phys. Acta 43, 607 (1970).
[17]See, e.g., R.V. Reid, Jr., Ann. Phys. (N.Y.) 50, 411 (1968).
[18]T. Ikebe, J. Math. Kyoto Univ. 7, 93 (1967).
[19]W. Hunziker, Helv. Phys. Acta 40, 1052 (1967).
[20]T. Ikebe, Jpn. J. Math. 36, 33 (1967).
[21]Insofar as we are aware, the first physicists to define and use transformations of this type were D. Bohm and E.P. Gross. Their work is cited by J.S. Bell, in *The Many Body Problem*, edited by C. Fronsdal (Benjamin, New York, 1962), and is also cited by M. Eger and E.P. Gross, Ann. Phys. (N.Y.) 24, 63 (1963). See also G.A. Baker, Jr., Phys. Rev. 128, 1485 (1962).
[22]M.H. Stone, *Linear Transformations in Hilbert Space*, Am. Math. Soc. Colloq. Pub. 15 (Am. Math. Soc., New York, 1932), Theorem 6.6, p. 232.
[23]See, e.g., Stone,[22] Definition 4.3, p. 130.
[24]See, e.g., C.R. Putnam, *Commutation Properties of Hilbert Space Operators and Related Topics* (Springer, New York, 1967), Corollary 2, p. 44.
[25]See, e.g., Stone,[22] Theorem 4.5, p. 130.
[26]See, e.g., Stone,[22] Theorem 8.14, p. 327.
[27]See, e.g., Putnam,[24] Corollary 2, p. 47.
[28]See, e.g., Refs. 6 and 14.
[29]See, e.g., T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Theorem 5.4, p. 303.
[30]The first assertion of this theorem follows from Theorem 4, p. 21 of Ref. 14 and the second assertion thereof from Lemma 7, p. 6 of this reference.
[31]This theorem is a slight simplification of Corollary 4, p. 19, of Ref. 14.

Bounds on the admittance for KMS states

J. Naudts

University of Antwerpen, Belgium

A. Verbeure

*University of Leuven, Belgium**
(Received 18 August 1975)

Upper and lower bounds are proved for the static admittance of observables in a KMS state on a von Neumann algebra. As an application some exact results for the transverse Ising model are derived.

I. INTRODUCTION

Recently Roepstorff[1] derived a new upper and lower bound for the state admittance of observables in a Gibbs state. They are given in the case of a finite lattice system.

As is well-known the states of infinite systems (i.e., in the thermodynamic limit) are no longer of Gibbs' type, and it is now widely accepted that an equilibrium state of an infinite system should be described by a state satisfying the KMS condition.[2]

In this paper we give rigorous proofs of the upper and lower bounds given in Ref. 1, and derive new bounds for KMS states on a von Neumann algebra of observables. Hence the results are valid not only for infinite lattice systems but also for continuous systems.

In Sec. II we introduce the necessary mathematical material and prove some more properties of the Bogoliubov or Kubo–Mori scalar product, which was studied in Ref. 3. In Sec. III two upper bounds and essentially two lower bounds of the static admittance are derived. Finally in Sec. IV we apply the inequalities to prove some exact results for the transverse Ising model.

II. MATHEMATICAL FRAMEWORK

Let $\mathcal{M}$ be a von Neumann algebra on a Hilbert space $\mathcal{H}$. Let $t \to U_t$ be a strongly continuous map from the real numbers $\mathbb{R}$ into the group of unitaries on $\mathcal{H}$, then there exists a self-adjoint operator H on $\mathcal{H}$ such that $U_t = \exp itH$, and let $x_t = U_t x U_t^*$. Furthermore let ω be any vector state on $\mathcal{M}$, i.e., $\omega(x) = (\Omega, x\Omega)$ for all $x \in \mathcal{M}$, with with Ω a cyclic element of $\mathcal{H}$. The state ω is called an equilibrium state if it satisfies the following definition.

Definition II.1: The state ω on $\mathcal{M}$ satisfies the KMS condition at inverse temperature $\beta = 1/kT$, if for any pair x, y of observables in $\mathcal{M}$, there exists a complex function $F_{x,y}(z)$, defined, bounded and continuous on the strip $-\beta \leq \mathrm{Im} z \leq 0$, and analytic inside, with boundary values $F_{x,y}(t) = \omega(x_t y)$, $F_{xy}(t - i\beta) = \omega(y x_t)$.

Without restriction of generality, let $\beta = 1$ in the sequel. Any state ω satisfying the KMS condition has the following properties[4]:

(i) the vector Ω is separating;

(ii) for all t, $U_t\Omega = \Omega$;

(iii) there exists an operator $\Delta = \exp(-H)$ on $\mathcal{H}$, such that $\Delta = FS$ where S is the closure of the map $x\Omega \to x^*\Omega$, $x \in \mathcal{M}$, F is the adjoint of S; furthermore, $S = J\Delta^{1/2}$ is the polar decomposition of S and $J\Delta^{it} = \Delta^{it}J$, $\Delta^{it} = U_t$, $J\Omega = \Omega$, $\Delta\Omega = \Omega$;

(iv) there exists a subset $\mathcal{B}$ of elements of $\mathcal{M}$ such that $\mathcal{B}$ is invariant under left multiplication with Δ^α, $\alpha \in C$ (complex numbers), and such that $\mathcal{B}$ is dense in the Hilbert space $\mathcal{D}(\Delta^\alpha)$ (domain of Δ^α). $\mathcal{B}$ is generally called the set of analytic elements.

In Ref. 3 we defined the Hilbert space $\tilde{\mathcal{H}}$, as the closure of $\mathcal{M}$ with respect to the scalar product $(x,y)_\sim = (Tx\Omega, Ty\Omega)$; $x, y \in \mathcal{M}$, $T = [(\Delta - 1)/\ln\Delta]^{1/2}$.

Furthermore the following results were proved:

(i) there exists a unitary operator U from $\tilde{\mathcal{H}}$ onto $\mathcal{H}$, defined by $Ux = Tx\Omega$, $x \in \mathcal{M}$;

(ii) let

$$\chi_{x,y}(0) = \lim_{z \to 0\pm} i \int_0^\infty dt \exp(\mp izt)\, \omega([x_t, y])$$

be the static admittance of the pair of observables $x, y \in \mathcal{M}$, then

$$(x^*, y)_\sim = \chi_{x,y}(0) + (\Omega, xE_0 y\Omega), \tag{1}$$

where E_0 is the orthogonal projection on the set of U_t-invariant vectors of $\mathcal{H}$.

If E_0 is one-dimensional then

$$(x^*, y)_\sim = \chi_{x,y}(0) + \omega(x)\,\omega(y).$$

In the following we derive bounds for the scalar product $(x,y)_\sim$; the implications for the static admittance are given by formula (1).

(iii) for each pair x, y of elements in $\mathcal{M}$,

$$\begin{aligned}(x,y)_\sim &= \int_{-1}^{0} dt\, F_{x,y}(it), \\ &= \int_0^{1/2} dt \{(x\Omega, \Delta^t y\Omega) + (y^*\Omega, \Delta^t x^*\Omega)\}.\end{aligned} \tag{2}$$

Now we prove some more properties of this scalar product.

Proposition II.2: For all $x, y \in \mathcal{M}$, we have

$$(x,y)_\sim = \tfrac{1}{2} \int_{-1}^{1} dt\, (\Delta^{(1+t)/4} x\Omega, \Delta^{(1+t)/4} y\Omega). \tag{3}$$

Proof: Starting from formula (2), after a few substitutions for the integration variable one gets

$$(x,y)_\sim = \tfrac{1}{2}\int_{-1}^{0} dt\ (\Delta^{(1+t)/2}x\Omega,\ y\Omega)\,dt + \tfrac{1}{2}\int_{0}^{1} dt(\Delta^{(1-t)/2}y^*\Omega, x^*\Omega).$$

But

$$\begin{aligned}(\Delta^{(1/2)(1-t)}y^*\Omega,\ x^*\Omega) &= (\Delta^{(1-t)/4}J\Delta^{1/2}y\Omega,\ \Delta^{(1-t)/4}J\Delta^{1/2}x\Omega)\\ &= (J\Delta^{(t-1)/4}\Delta^{1/2}y\Omega,\ J\Delta^{(t-1)/4}\Delta^{1/2}x\Omega)\\ &= (J\Delta^{(1+t)/4}y\Omega,\ J\Delta^{(1+t)/4}x\Omega)\\ &= (\Delta^{(1+t)/4}x\Omega,\ \Delta^{(1+t)/4}y\Omega),\end{aligned}$$

and the result follows. QED

For any $x \in \mathcal{M}$, denote

$$f_x(t) = \log\|\Delta^{(1+t)/4}x\Omega\|^2. \tag{4}$$

As is easily checked:

$$f_x(-1) = \log\omega(x^*x), \tag{5}$$

$$f_x(1) = \log\omega(xx^*). \tag{6}$$

Furthermore

$$2\exp[f_x(t)]\frac{d}{dt}f_x(t) = -(\Delta^{(1+t)/4}x\Omega,\ H\Delta^{(1+t)/4}x\Omega) \quad \text{for } t \in (-1,1).$$

Hence, if $x^*\Omega \in \mathcal{D}(H)$

$$f_x'(1) = \omega(xHx^*)/2\omega(xx^*), \tag{7}$$

and if $x\Omega \in \mathcal{D}(H)$

$$f_x'(-1) = -\omega(x^*Hx)/2\omega(x^*x). \tag{8}$$

Proposition II. 3:

(i) for all $x \in \mathcal{M}$, the function $t \to f_x(t)$ is convex on the interval $[-1,+1]$;

(ii) for all $x = x^* \in \mathcal{M}$, the function satisfies $f_x(t) = f_x(-t)$.

Proof: For any $x \in \mathcal{B}$ (analytic elements), the function $t \to f_x(t)$ is analytic, hence it is sufficient to prove $f_x((t+s)/2) \leq \tfrac{1}{2}[f_x(t) + f_x(s)]$. But this follows from Schwartz's inequality,

$$\begin{aligned}\|\Delta^{[1+(t+s)/2]/4}x\Omega\|^2 &= (\Delta^{(1+t)/4}x\Omega,\ \Delta^{(1+s)/4}x\Omega)\\ &\leq \|\Delta^{(1/4)(1+t)}x\Omega\|\,\|\Delta^{(1/4)(1+s)}x\Omega\|.\end{aligned}$$

Let now x be any element of $\mathcal{M}$, from definition II. 1 (iv), there exists a sequence of elements $\{x_n\}$ in $\mathcal{B}$, such that $x_n\Omega$ tends to $x\Omega$ and $\Delta^{1/2}x_n\Omega$ tends to $\Delta^{1/2}x\Omega$.

Hence $f_{x_n}(t)$ tends to $f_x(t)$, and $f_x(t)$ is also a convex function as the limit of convex functions. This proves (i).

Now, if $x = x^*$, then

$$\begin{aligned}\|\Delta^{(1/4)(1+t)}x\Omega\|^2 &= \|J\Delta^{(1+t)/4}x\Omega\|^2\\ &= \|J\Delta^{(t-1)/4}\Delta^{1/2}x\Omega\|^2\\ &= \|\Delta^{(1/4)(1-t)}x^*\Omega\|^2 = \|\Delta^{(1/4)(1-t)}x\Omega\|^2,\end{aligned}$$

and (ii) follows. QED

III. BOUNDS

Theorem III. 1 (*Roepstorff*): For all $x \in \mathcal{M}$, we have

$$(x,x)_\sim \leq \omega([x,x^*])/\log\frac{\omega(xx^*)}{\omega(x^*x)}.$$

Proof: From formula (4) for all $x \in \mathcal{M}$,

$$(x,x)_\sim = \tfrac{1}{2}\int_{-1}^{1} dt\ \exp f_x(t).$$

From Proposition II. 3 (i), $t \in [-1,1]$,

$$\frac{f_x(1)-f_x(-1)}{2}t + \frac{f(1)+f(-1)}{2} \geq f(t) \quad \text{[see Fig. 1, curve (1)]},$$

hence

$$(x,x)_\sim \leq \frac{\exp f_x(1) - \exp f_x(-1)}{f_x(1)-f_x(-1)}$$

and by (5) and (6),

$$(x,x)_\sim \leq \omega([x,x^*])/\log[\omega(xx^*)/\omega(x^*x)]. \quad \text{QED}$$

Corollary III. 2. (*Bogoliubov—Roepstorff*): For all y of $\mathcal{M}$ and elements of x of $\mathcal{M}$ such that $x\Omega$ and $x^*\Omega$ belong to the domain $\mathcal{D}(H)$ of H we have,

$$|\omega([y^*,x])|^2 \leq \tfrac{1}{2}\omega(\{y,y^*\})\omega([[x.H],\ x^*])g(r),$$

where

$$g(r) = \frac{2r}{\log(1+r/1-r)} \leq 1,$$

$$r = \omega([y,y^*])/\omega(\{y,y^*\}).$$

Proof: Under the conditions of the corollary, as in Ref. 3, Theorem III. 8, one gets

$$|\omega([y^*,x])|^2 \leq (y,y)_\sim\,\omega([[x,H],x^*]).$$

Using Theorem III. 1 to majorize $(y,y)_\sim$, the corollary follows. QED

Remark: $g(r=0) = 1$, so that the inequality above is not stronger than the original Bogoliubov inequality in the case $\omega(yy^*) = \omega(y^*y)$, in particular if $y = y^*$ or if ω is

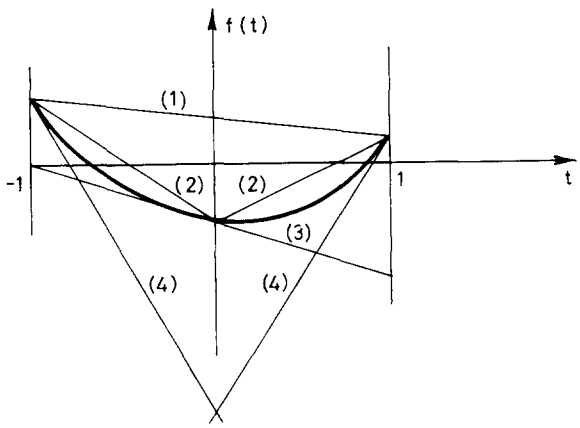


FIG. 1.

a central state. On the other hand $\lim_{r\to 1} g(r) = 0$. Hence the interest of the stronger version lies in the region $r \approx 1$. A stronger upper bound is obtained in the following theorem.

Theorem III. 3: For all $x \in \mathcal{M}$

$$(x,x)_\omega \leq \frac{1}{2}\left(\frac{\omega(x^*[\exp(-H/2)-1]x)}{\log[\omega(x^*\exp(-H/2)x)/\omega(x^*x)]} + \frac{\omega(x[\exp(-H/2)-1]x^*)}{\log[\omega(x\exp(-H/2)x^*)/\omega(xx^*)]}\right).$$

Proof: From Proposition II. 3 (i) [see Fig. 1, curve (2)],

$$(f_x(0)-f_x(-1))t + f_x(0) \geq f_x(t), \quad t\in[-1,0],$$
$$(f_x(1)-f_x(0))t + f_x(0) \geq f_x(t), \quad t\in[0,1].$$

Hence

$$(x,x)_\omega = \tfrac{1}{2}\int_{-1}^{1} dt\, \exp f_x(t) \leq \frac{1}{2}\left(\frac{\exp[f_x(0)]-\exp[f_x(-1)]}{f_x(0)-f_x(-1)} + \frac{\exp[f_x(1)]-\exp[f_x(0)]}{f_x(1)-f_x(0)}\right).$$

Using formulas (5) and (6), and the fact that

$$f_x(0) = \log\omega(x^*\exp(-H/2)x) = \log\omega(x\exp(-H/2)x^*),$$

one gets the result. QED

Remark: In any case, the inequality of Theorem III. 3 is a much stronger inequality than the one of Theorem III. 1. The associated Bogoliubov inequality will also be much better. We do not elaborate on this point here.

Now we turn to the lower bounds.

Theorem III. 4: For all $x \in \mathcal{M}$,

$$(x,x)_\omega \geq \sup_{-1\leq s\leq 1} \exp\{f_x(s) - sf_x'(s)\}\sinh[f_x'(s)]/f_x'(s).$$

In particular,

(i) if $x\Omega \in \mathcal{D}(H)$,

$$(x,x)_\omega \geq \frac{\omega(x^*x)^2}{2\omega(x^*Hx)}\left[1-\exp\left(-2\,\frac{\omega(x^*Hx)}{\omega(x^*x)}\right)\right];$$

(ii) if $x^*\Omega \in \mathcal{D}(H)$,

$$(x,x)_\omega \geq \frac{\omega(xx^*)^2}{2\omega(xHx^*)}\left[1-\exp\left(-2\,\frac{\omega(xHx^*)}{\omega(xx^*)}\right)\right];$$

(iii) if $x = x^*$,

$$(x,x)_\omega \geq \omega(x\exp(-H/2)x).$$

Proof: From Proposition II. 3 (i), for all $s \in (-1,1)$,

$$f_x(t) \geq f_x(s) + (t-s)f_x'(s) \quad \text{[see Fig. 1, curve (3)]},$$

hence

$$(x,x)_\omega \geq \tfrac{1}{2}\int_{-1}^{1} dt\, \exp[f(s) + (t-s)f'(s)].$$

After integration one gets

$$(x,x)_\omega \geq \exp[f(s) - sf'(s)]\sinh[f'(s)]/f'(s),$$

yielding the first part of the theorem.

If $x\Omega \in \mathcal{D}(H)$, then using (5) and (8) one gets (i) by taking $s = -1$.

If $x^*\Omega \in \mathcal{D}(H)$, put $s = 1$ and use (7) and (8) to get (ii).

If $x = x^*$, then by Proposition II. 3 (ii) the function $f_x(t)$ is symmetric around $t = 0$ and $f_x'(t=0) = 0$, hence

$$(x,x)_\omega \geq \exp f_x(0) = \omega(x\exp(-H/2)x). \quad \text{QED}$$

Finally we derive another lower bound, which was first derived in Ref. 1 for finite lattice systems. We prove it only in the case of elements $x = x^* \in \mathcal{M}$.

Using the same proof it may be derived in the general case.

Theorem III. 5: For all elements $x = x^*$ of $\mathcal{M}$ such that $x\Omega \in \mathcal{D}(H)$,

$$(x,x)_\omega \geq \omega(x^2)\left(\frac{1-e^{-C}}{C}\right)$$

where

$$C = \tfrac{1}{2}\omega(xHx)/\omega(x^2).$$

Proof: From Proposition II. 3 (i) and (ii) it follows that [see Fig. 1, curve (4)],

$$f_x(t) \geq f_x'(-1)(t+1) + f_x(-1) \quad \text{if} \quad -1 \leq t \leq 0,$$
$$f_x(t) \geq f_x'(1)(t-1) + f_x(1) \quad \text{if} \quad 0 \leq t \leq 1.$$

Hence

$$(x,x)_\omega \geq \tfrac{1}{2}\int_{-1}^{0} dt\, \exp[f_x'(-1)t + f_x(-1) + f_x'(-1)] + \tfrac{1}{2}\int_{0}^{1} dt\, \exp[f_x'(1)t + f_x(1) - f_x'(1)]$$

Using (5), (6), (7), and (8) and the fact that $f_x(1) = f_x(-1)$; $f_x'(-1) = -f_x'(1)$, and one gets the proof. QED

Remark: There is no strict relation between the two lower bounds which we proved. However the last inequality is stronger than the particular cases (i) or (ii) of Theorem III. 4. The essential difference consists in minorizing the convex function respectively by one or by two straight lines.

IV. APPLICATION

As an application we derive some exact results on the transverse Ising model,[5] sometimes called the Blinc model or Tunnel model,[6] described by the following Hamiltonian. Let Z^ν be a ν-dimensional lattice, Λ any finite subset of Z^ν, then the local Hamiltonian is given by

$$H_\Lambda = \Omega \sum_{k\in\Lambda} \sigma_k^x + \tfrac{1}{2}\sum_{k,l\in\Lambda} J(|k-l|)\sigma_k^z\sigma_l^z,$$

where Ω and $J(|k|)$ are real numbers and we take $J(0) = 0$. Furthermore, σ_p^α; $\alpha = x,y,z$; $p \in Z^\nu$ are the Pauli

matrices, satisfying $[\sigma_p^\alpha, \sigma_q^\beta] = 2i\epsilon_{\alpha\beta\gamma}\sigma_p^\gamma \delta_{p,q}$. The local algebra of observables for the volume Λ is the completed tensor product of the 2×2 matrices M_2,

$$\mathfrak{A}_\Lambda = \bigotimes_{k\in\Lambda} M_2.$$

The C^*-algebra of observables $\mathfrak{A}$ is then the norm closure of $U_{\Lambda\in Z^\nu}\mathfrak{A}_\Lambda$.

We suppose that $\sum_{k\in Z^\nu} |J(k)| < \infty$, such that the map

$$t\in \mathrm{R} \to \alpha_t(x) = \lim_{\Lambda\to\infty} \exp(itH_\Lambda)\, x \exp(-itH_\Lambda), \quad x\in\mathfrak{A}$$

exists and yields a strongly continuous group of $*$-automorphisms of $\mathfrak{A}$.[7]

As is easily checked, the following limits exist, and define a not necessarily bounded derivation H of the C^*-algebra,

$$H(\sigma_p^x) = \lim_{\Lambda\to\infty} [H_\Lambda, \sigma_p^x] = 2i\tau_p\sigma_p^y \tag{9}$$

where

$$\tau_p = \sum_{k\in Z^\nu} J(|k-p|)\sigma_k^z,$$

$$H(\sigma_p^y) = \lim_{\Lambda\to\infty} [H_\Lambda, \sigma_p^y] = 2i\Omega\sigma_p^z - 2i\tau_p\sigma_p^x, \tag{10}$$

$$H(\sigma_p^z) = \lim_{\Lambda\to\infty} [H_\Lambda, \sigma_p^z] = -2i\Omega\sigma_p^y. \tag{11}$$

Let ω be a time-invariant state on Ω, and let (π, Ω, U_t) be its GNS representation, i.e.,

$$\omega(x) = (\Omega, \pi(x)\Omega), \quad x\in\mathfrak{A},$$

where Ω is a cyclic vector of H for $\pi(\mathfrak{A})$,

$$\pi(\alpha_t(x)) = U_t\, \pi(x)\, U_t^*$$

$$M = \pi(\mathfrak{A})''.$$

Consider the extension of ω to M; we denote it by the same symbol ω. From now on we drop the notation π. Finally let us suppose that ω is a KMS state of M for the time evolution $x_t = U_t x U_t^*$. Now we are in a position to apply the inequalities.

Using (11) and Ref. 3, Theorem III. 2 we get

$$(\sigma_p^y, \sigma_p^y)_\sim = -(1/\Omega)\,\omega(\sigma_p^x) \geq 0. \tag{12}$$

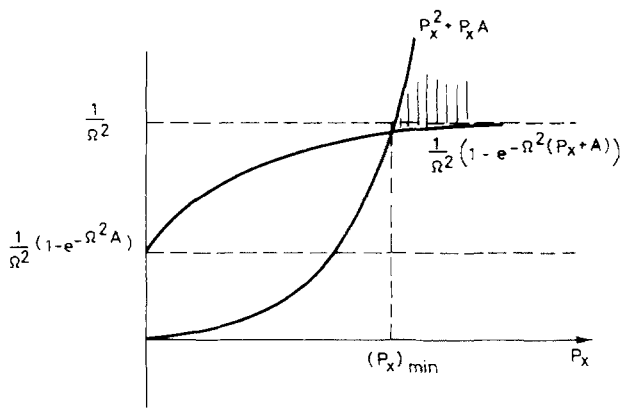


FIG. 2.

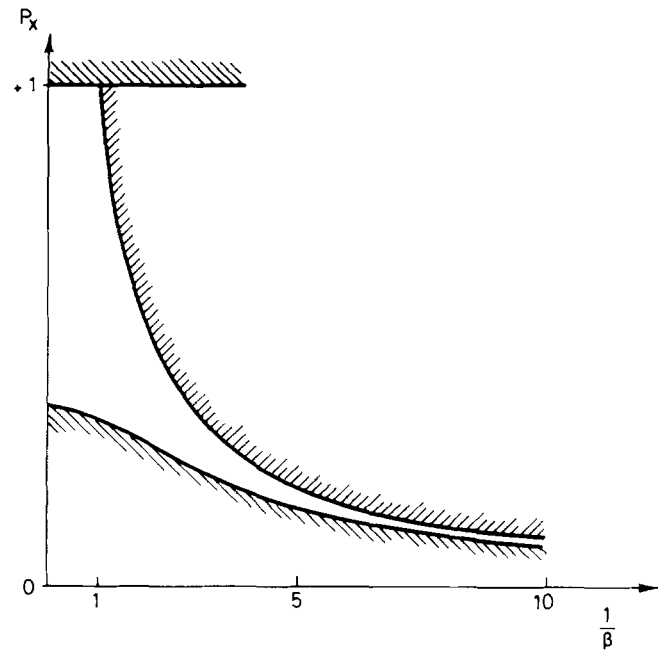


FIG. 3.

Let us first apply Corollary III. 2 with $y = \sigma_p^y$ and $x = \sigma_p^z$. Using Eqs. (9)—(12) one gets immediately

$$0 \leq -\omega(\sigma_p^x)/\Omega \leq 1. \tag{13}$$

This means that for any fixed temperature $\omega(\sigma_p^x)$ tends to zero, if the frequency Ω tends to zero. Or in other words the phase transition in the Ising models (i.e., $\Omega = 0$) is not due to a breaking of the symmetry along the x direction, which makes the model a quantum-mechanical one.

A lower bound for the spin polarization in the x direction at fixed temperature and fixed frequency Ω, can be found by applying the inequality of Theorem III. 5. Take $x = \sigma_p^y$, then again using (9)

$$-(1/\Omega)\omega(\sigma_p^x) \geq 1 - e^{-C}/C, \tag{14}$$

where $C = \Omega^2[-\omega(\sigma_p^x)/\Omega + A]$,

$$A = -\omega(\tau_p\sigma_p^z)/\Omega.$$

As now

$$(H\sigma_p^x, H\sigma_p^x) = -4\omega(\tau_p\sigma_p^z) \geq 0,$$

it follows that $A \geq 0$.

Denote

$$P_x = -\omega(\sigma_p^x)/\Omega,$$

then (14) becomes

$$P_x^2 + AP_x \geq \frac{1}{\Omega^2}\{1 - \exp[-\Omega^2(P_x + A)]\},$$

yielding a minimum value $(P_x)_{\min}$ for the polarization P_x as is shown in Fig. 2.

It is easy to reintroduce the inverse temperature β in the formulas and then Fig. 3 represents a numerical

calculation of the bounds (13) and (14) as a function of the temperature. The quantity A in (14) has been majorized as follows:

$$A \leq \sum_{k \in Z^\nu} |J(k)|.$$

*Postal address: Institute for Theoretical Physics, K. U. L., Celestijnenlaan 200 D, B-3030 Heverlee, Belgium.

[1]G. Roepstorff, "Correlation inequalities in Quantum Statistical Mechanics and their Application in the Kondo Problem," preprint.
[2]R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. 5, 215 (1967).
[3]J. Naudts, A. Verbeure, and R. Weder, Commun. Math. Phys. **44**, 87 (1975).
[4]M. Takesaki, "Tomita's Theory of Modular Hilbert Algebras and its applications," (Lecture Notes in Mathematics, Springer, 1970), near p. 128.
[5]J. Ginibre, Commun. Math. Phys. 14, 205 (1969).
[6]R. Blinc, J. Phys. Chem. Solids 13, 204 (1960); P. G. de Gennes, Solid State Commun. 1, 132 (1963).
[7]D. W. Robinson, Commun. Math. Phys. **7**, 337 (1968).

Conformal algebra in superspace and supergauge theory*

Peter G. O. Freund

Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637
(Received 6 October 1975)

The conformal algebra in a superspace with orthosymplectic metric is found to be an orthosymplectic algebra with two extra Bose dimensions. It is argued that all supergauge fields are also Nambu–Goldstone fields. In supergauge theory the nature of the gauged internal symmetry is found to be severely restricted: It must be O(2N), where N is the number of fundamental Dirac fermions or alternatively one-eighttth of the number of Fermi dimensions of superspace.

1. INTRODUCTION

Of the classical graded Lie algebras (GLA's)[1] the orthosymplectic algebras are characteristic of graded vector spaces endowed with a metric. They are the graded version of the ordinary (ungraded) orthogonal algebras.

In particle physics an important role is played by the conformal algebra[2] which supplements the orthogonal (or pseudo-orthgonal) algebras by the translations P_μ, dilatation D, and conformal boosts K_μ. If the metric of the underlying space is pseudo-Euclidean with m plus signs and n minus signs then the pseudo-orthogonal algebra is $O(m,n)$. The corresponding "Poincaré" algebra $IO(m,n)$ [$\equiv O(m,n)+$translations] is embedded in the conformal algebra $C(m,n)\equiv O(m+1,n+1)$. The *nonlinear* action of the group $C(m,n)$ on the $m+n$-dimensional pseudo-Euclidean space is that on the homogeneous space $O(m+1,n+1)/IO(m,n)\otimes D$ ($D\equiv$ Abelian group of dilatations). Here we generalize these concepts to the graded case. Specifically the inhomogeneous orthosymplectic algebra $IOSp(2r|s_1,s_2)$ (of a pseudo-Euclidean superspace[3] with $2r$ Fermi dimensions and s_1+s_2 Bose dimensions with Bose sector metric with s_1 plus signs and s_2 minus signs[1] can be again embedded into a conformal GLA $C(2r|s_1,s_2)\equiv OSp(2r|s_1+1,s_2+1)$ with a nonlinear realization over the $2r+s_1+s_2$-dimensional superspace. One can then generalize to the graded case the arguments[4] that show the gravitational field in Einstein's theory to be simultaneously a gauge field (of the Poincaré group) and a Nambu–Goldstone field associated with the spontaneous breaking of general covariance. All fields (Bose and Fermi) in supergauge theories[5,6] are thus simultaneously gauge and Nambu–Goldstone fields. They are gauge fields of the inhomogeneous orthosymplectic "group" and Nambu–Goldstone fields corresponding to the spontaneous breaking of the general covariance in superspace. In fact supergauge theories are the only theories involving Fermi fields where *every* field appearing in the Lagrangian is both a gauge and a Nambu–Goldstone field. The remarkable thing is that in supergauge theories even the nature of the admissible gauged internal symmetry (i.e., symmetry with Lorentz-scalar generators) is severely restricted. As we shall see, it must be of the form O(2N) with N the number of "fundamental" Dirac fermions (quarks + leptons).

2. A CANONICAL BASIS FOR AN ORTHOSYMPLECTIC ALGEBRA

Consider the operators x^a and ∂_a ($a=1,\ldots,N=2r+s_1+s_2$). Let the first $2r$ (last $s=s_1+s_2$) x^a's and ∂_a's be Fermi (Bose). The bracketing rules are

$$[x^a,x^b]=[\partial_a,\partial_b]=0,\quad [\partial_a,x^b]=\delta_a^b, \tag{2.1a}$$

where as in Ref. 1 the bracket [ ] stands for commutator except when both bracketed operators are fermionic in which case it is an anticommutator. The (pseudo-) orthosymplectic metric

$$g_{ab}=\begin{array}{c}\begin{matrix}2r & s_1 & s_2\end{matrix}\\ \begin{pmatrix} C & 0 & 0\\ 0 & +1 & 0\\ 0 & 0 & -1\end{pmatrix}\end{array}\begin{matrix}2r\\ s_1\\ s_2,\end{matrix}\qquad C=\begin{pmatrix} i\sigma_2 & & \\ & \ddots & \\ & & i\sigma_2\end{pmatrix}r, \tag{2.1b}$$

can be used to raise or lower indices, i.e.,

$$V_a=V^b g_{ba},\quad V^a=V_b g^{ba}, \tag{2.1c}$$

where we defined g^{ab} by

$$g_{ab}g^{bc}=\delta_a^c. \tag{2.1d}$$

The alternative definition $\tilde V_a=g_{ab}V^b$ only gives $\tilde V_a=(-1)^{\bar a}V_a$ where $\bar a$ is the grade of V_a (+1 for Fermi, 0 for Bose in a mod 2 grading).

Next we define the operators

$$M_b^a=x^a\partial_b-(-1)^{\bar a+\bar b+\bar a\bar b}x_b\partial^a \tag{2.2}$$

of grade $\bar a+\bar b$ (mod 2). They obey the structure relations

$$[M_b^a,M_d^c]=\delta_b^c M_d^a-(-1)^{\bar a+\bar d\bar b+\bar b\bar c+\bar c\bar a}\delta_d^a M_b^c + (-1)^{\bar d+\bar d\bar b+\bar b\bar a}g^{ac}g_{md}M_b^m-(-1)^{\bar b+\bar c+\bar b\bar c}g_{bd}g^{mc}M_m^a, \tag{2.3}$$

so that they span a GLA. This GLA is precisely $OSp(2r|s_1,s_2)$ as can be seen from

$$[M_b^a,x_c x^c]=0, \tag{2.4}$$

which means that the metric form $x^2\equiv x_c x^c\equiv g_{ab}x^a x^b$ is left invariant by the M_b^a. The canonical basis (2.2) will be extensively used below.

3. GRADED CONFORMAL LIE ALGEBRAS

Consider the canonical basis of Eq. (2.2) of the GLA $OSp(2r|s_1,s_2)$. To these $r(2r+1)+(s_1+s_2)(s_1+s_2-1)/2+2r(s_1+s_2)$ generators M_b^a add the $2(2r+s_1+s_2)+1$ generators

$$P_a=\partial_a,$$
$$D=x^a\partial_a, \tag{3.1}$$

$$K_a = x_c x^c \partial_a - (-1)^{\tilde{a}} x_a x^c \partial_c,$$

for a total of $r(2r+1) + (s_1 + s_2 + 2)(s_1 + s_2 + 1)/2 + 2r(s_1 + s_2 + 2)$ generators, i.e., the right number for $\mathrm{OSp}(2r|s_1+1, s_2+1)$. By direct computation it is readily checked that the generators (2.2) and (3.1) do indeed span $\mathrm{OSp}(2r|s_1+1,\ s_2+1)$. The P_a's are obviously translations, the K_a's conformal boosts, and D a dilatation. We thus see that the graded conformal algebra is $C(2r|s_1, s_2) \equiv \mathrm{OSp}(2r|s_1+1, s_2+1)$, and in Eqs. (2.2) and (3.1) we possess the nonlinear action of this algebra on $(2r + s_1 + s_2)$-dimensional superspace.

4. SUPERGAUGE FIELDS AS NAMBU-GOLDSTONE FIELDS

It has been shown by Cho and Freund[6] that non-Abelian gauge fields can simultaneously be Nambu–Goldstone fields corresponding to the spontaneous breaking of general covariance in a higher-dimensional space. As long as these higher dimensions are bosonic, just as in Kaluza–Klein theory, one has a hard time avoiding their observability. It was therefore suggested[6] that all dimensions in excess of the four basic dimensions of space–time are fermionic. They are then trivially unobservable, as fields depend only polinomially on them. Gauge invariance of the second kind for internal symmetries emerges just as in the case of extra Bose dimensions treated in Ref. 6. We are now dealing with a supergauge theory and are gauging the "graded Poincaré group" $\mathrm{IOSp}(8N|1,3)$ where N is the number of Dirac fermions among the superspace dimensions (four Fermi dimensions build a Majorana spinor, eight a Dirac spinor). Just as in the ungraded case this corresponds to the spontaneous breaking of general covariance in superspace. We do not wish to reproduce here the details of the generalizations of the arguments of Refs. 4 and 6 to this case. Rather we briefly note some of the novel features of such a generalization.

First of all the general coordinate transformations in superspace have a simple finite dimensional graded Lie subalgebra that is not classical (in the sense of Ref. 1) i.e., a hyperexceptional GL subalgebra. Indeed, the general coordinate transformations on the Fermi part of superspace have generators

$$(x^1)^{n_1} \circ\circ\circ (x^{8N})^{n_{8N}} \partial_a, \quad a = 1, \ldots, 8N,$$

and in virtue of the exclusion principle ($x^1, \ldots, x^{8N}$ are Fermi dimensions) the n_i here can only take the values 0 and 1 so that there are only a finite number, to wit $8N(2^{8N})$ such generators. This is precisely one of the hyperexceptional simple GLA's mentioned in Ref. 1. Its representations are not fully reducible and this algebra does not lead to a classification of fields into multiplets, a feature not unexpected for a nonlinearly realized symmetry.

Finally, just as in the purely bosonic case, in the course of the spontaneous breaking of general covariance one breaks the conformal invariance (of superspace). In the four-dimensional ungraded case this leads to the appearance of five Nambu–Goldstone fields, a scalar (dilaton) and a vector field corresponding to the breaking of dilatations and conformal transformations. The vector field can be set equal to the gradient of the dilaton field. Were one, however, to keep this vector field as an ind independent field, it would play the role of a gauge field of dilatations as in Weyl's unified field theory.[7] Similarly in the graded case one picks up one scalar and one "$(4+8N)$-vector" Nambu–Goldstone fields. The latter is the analog of the Weyl field but now along with the 4-vector field it also has N Dirac fermion supersymmetry partners. Again these correspond to the breaking of conformal transformations generated by the K_a's of Eq. (3.1).

These arguments will be more fully presented elsewhere.

5. PHYSICAL ASPECTS OF SUPERGAUGE THEORIES

In the absence of matter (or matter fields), Einstein's theory of gravitation has the feature that all fields (i.e., the gravitational field) appearing in the Lagrangian are simultaneously gauge fields and Nambu–Goldstone fields. Supergauge theories share this property with Einstein's theory with the additional virtue that Dirac fields are present and can provide a realistic description of matter. If the vector fields appearing in the supergauge Lagrangian are to have a Yang–Mills part, then the supergauge Lagrangian is fixed to be the scalar curvature density (in superspace) and not one of the scalars quadratic in the Riemann–Christoffel supertensor. Such terms are induced upon renormalization but at the classical level the theory is essentially unambiguous (modulo a cosmological term). It unifies fields of spins 0, $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2 all of which are gauge *and* Nambu–Goldstone fields. It is a very tight and well-defined structure. At this point we wish to note that this structure even restricts the nature of the gauged *internal* symmetry group. As we now show, not any compact Lie group but only certain orthogonal groups are candidates for internal symmetries compatible with supergauge theories. To see this recall that supergauge theories gauge $\mathrm{OSp}(8N|1,3)$, the Bose sector of which is $\mathrm{Sp}(8N) \otimes \mathrm{O}(3,1)$. $\mathrm{O}(3,1)$ is the Lorentz group and $\mathrm{Sp}(8N)$ the group that shuffles all components of $2N$ Majorana spinors. Therefore, not all generators of $\mathrm{Sp}(8N)$ are Lorentz scalars. The Lorentz scalars are only those that shuffle corresponding components (ith component with ith component, $i = 1, \ldots, 4$) of the $2N$ Majorana spinors. These span the subgroup $\mathrm{O}(2N)$ of $\mathrm{Sp}(8N)$. It is this $\mathrm{O}(2N)$ which is[8] the maximal internal symmetry contained in $\mathrm{OSp}(8N|1,3)$. It thus follows that only $\mathrm{O}(2N)$ internal symmetries are gauged in supergauge theories. This by no means implies that the $\mathrm{O}(2N)$ symmetry is to be exact. There are sufficient scalar fields around to spontaneously break $\mathrm{O}(2N)$. But for N Dirac-fermions it is $\mathrm{O}(2N)$ and not $\mathrm{SU}(N)$ which is the internal symmetry of the supergauge theories. Thus, e.g., for the antiquated but familiar case of three quarks one would expect O(6) and not SU(3). $\mathrm{O}(2N)$ contains $\mathrm{SU}(N)$ and as a result of the spontaneous breaking chain $\mathrm{SU}(N)$ is an obvious intermediate step. The important feature is that the *Fermi dimensionality of superspace* $(8N)$ *uniquely determines the nature of the internal symmetry group* $[\mathrm{O}(2N)]$. This makes it all the more interesting to work out such theories in full detail.

ACKNOWLEDGMENTS

Following the completion of this work I learned in a conversation with Dr. P. Srivastava that he has independently obtained the results of Secs. 2 and 3 of this paper. I wish to thank the Aspen Center of Physics for offering me the opportunity to complete and lecture on this work in a very pleasant natural (that includes human) environment.

*Work supported in part by the National Science Foundation Contract No. MPS74-08833.

[1]P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).

[2]See, e.g., G. Mack and A. Salam, Ann. Phys. (N.Y.) 53, 174 (1969).

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

[4]V. Ogievetsky, Nuovo Cimento Lett. 8, 988 (1973); V. Ogievetsky and A. B. Borisov, Dub a Preprint E2-7684; A. Salam, C. Isham, and J. Strathdee, Ann. of Phys. (N.Y.) 62, 98 (1971).

[5]P. Nath and R. Arnowitt, Phys. Lett. B 56, 177 (1975); R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B 56, 81 (1975).

[6]Y. M. Cho and P. G. O. Freund, Phys. Rev. (to be published).

[7]H. Weyl, *Raum*, *Zeit*, *Materie* (Springer-Verlag, Berlin, 1923), 5th ed.

[8]The same fermion number changing $O(2N)$ symmetry appears in a theory of N Dirac fields with S or P four Fermi interaction as noted by R. Marshak and S. Okubo, Nuovo Cimento 19, 1226 (1961).

A theorem on Ruch's principle of nonequilibrium statistics

Bernhard Lesche

Theoretische Physik III, Technische Universität Berlin, Berlin, Germany
(Received 30 September 1975)

It is proved that whenever two probability distributions, on a finite set, p and p' are given such that p' has a bigger mixing character than p, it is possible to find a unitary matrix U, such that $p'_i = \Sigma_j |U_{ij}|^2 p_j$. This theorem ensures that Ruch's principle of increasing mixing character provides the strongest assertion to be made upon the diagonal of the density matrix of some quantum mechanical system at time $t > 0$, without one's knowing the Hamiltonian operator but given that the density matrix was initially diagonal with a known but arbitrary diagonal.

INTRODUCTION

As the notions Ruch has introduced into statistical mechanics to formulate his principle of increasing mixing character are very new and thus not commonly known, some ideas are briefly recalled here.

Let $\mathcal{S}(N)$ be a finite set with $N \geq 2$ elements. If convenient, $\mathcal{S}(N)$ will be tacitly identified with the set $\{1, 2, \ldots, N\}$. To each element $i \in \mathcal{S}(N)$ let a number p_i be attached such that $p_i \geq 0$ for all $i \in \mathcal{S}(N)$ and $\sum_{i \in \mathcal{S}(N)} p_i = 1$. Such a function $p : \mathcal{S}(N) \to [0, 1]$ is called a probability distribution on $\mathcal{S}(N)$. Let the set of all probability distributions on $\mathcal{S}(N)$ be denoted by $V^{\mathcal{S}(N)}$. Now, one defines the following.

Definition 1: Two probability distributions $p, p' \in V^{\mathcal{S}(N)}$ are called mixing equivalent, if there exists a permutation $s \in S_N$ such that

$$p'_i = p_{s(i)}.$$

As S_N is a group this defines an equivalence relation.

Definition 2: The class of probability distributions that are mixing equivalent to $p \in V^{\mathcal{S}(N)}$ is called the mixing character of p and is written $[p]$.

Definition 3: $p' \in V^{\mathcal{S}(N)}$ is said to be more mixed than $p \in V^{\mathcal{S}(N)}$ if there exists a probability distribution c on S_N such that

$$p'_i = \sum_{s \in S_N} c_s p_{s(i)}.$$

Obviously, definition 3 provides a relation on the set of mixing characters. If p' is more mixed than p the mixing character $[p']$ is said to be bigger than $[p]$; $[p'] \succ [p]$.

One can prove that $[p'] \succ [p]$ is equivalent to the existence of a bistochastic matrix M such that $p'_i = \sum_j M_{ij} p_j$. Note that there may exist several ones. A matrix M is called bistochastic, if the following conditions hold

$$M_{ij} \geq 0,$$

$$\sum_{j=1}^{N} M_{ij} = 1,$$

$$\sum_{i=1}^{N} M_{ij} = 1.$$

Another equivalent way to define the relation " $\succ$ " is the following: Let $\bar{p} \in [p]$ and $\bar{p}' \in [p']$ be class-representatives in decreasing order, i.e., $i > j \Rightarrow (\bar{p}_i \leq \bar{p}_j \wedge \bar{p}'_i \leq \bar{p}'_j)$.

Then the assertion $[p'] \succ [p]$ is equivalent to

$$\forall \left(l \in \mathcal{S}(N) \right) \left(\sum_{i=1}^{l} \bar{p}'_i \leq \sum_{i=1}^{l} \bar{p}_i \right)$$

From this one concludes that " $\succ$ " is a partial order-relation on the set of mixing characters. Finally, we mention that $[p'] \succ [p]$ is equivalent to[2]

$$\forall (l \in \mathbb{R}) \left(\sum_{i \in \mathcal{S}(N)} |p'_i - l| \leq \sum_{i \in \mathcal{S}(N)} |p_i - l| \right)$$

Let $\mathcal{U}$ be the set of all unitary $N \times N$ matrices. The assertion $\exists (U \in \mathcal{U})\ (p'_i = \sum_j |U_{ij}|^2 p_j)$ will be abbreviated $p' \rhd p$. As the matrix $(|U_{ij}|^2)$ is bistochastic if $U \in \mathcal{U}$ we have

$$\forall (N \geq 2)\ \forall (p, p' \in V^{\mathcal{S}(N)})\ (p' \rhd p \Rightarrow [p'] \succ [p])$$

Our theorem is exactly the reverse implication.

Theorem 1:

$$\forall (N \geq 2)\ \forall (p, p' \in V^{\mathcal{S}(N)})([p'] \succ [p] \Rightarrow p' \rhd p).$$

First, we shall prove this theorem, and then illustrate its physical meaning.

THE PROOF OF THE THEOREM

Before proving the theorem we shall state some lemmas which provide us with knowledge necessary to advance a concise proof.

Lemma 1: Let P, P' be any permutation matrices. Then

$$\forall p\ \forall p'\ (p' \rhd p \Leftrightarrow P' p' \rhd P p).$$

Proof of Lemma 1: Lemma 1 can easily be inferred from the fact that permutation matrices in each row and column have the number 0 exactly $(N-1)$ times and the number 1 only once.

Lemma 2: Let M be a bistochastic 2×2 matrix. Then it is possible to find a unitary 2×2 matrix U such that $M_{ij} = |U_{ij}|^2$.

Proof of Lemma 2: M can be written as

$$\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix},$$

where $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$. A suitable U for instance is

$$\begin{pmatrix} \sqrt{\alpha} & \sqrt{\beta} \\ -\sqrt{\beta} & \sqrt{\alpha} \end{pmatrix}.$$

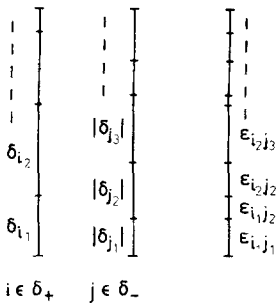


FIG. 1.

It is remarkable that bistochastic $N \times N$ matrices ($N \geq 3$) do not have this property. Note however that this does not contradict our theorem.

The next lemma tells us something about two probability distributions, p and p', which are in decreasing order. The differences $p_i' - p_i$ will be denoted by δ_i. Two subsets of $\mathcal{S}(N)$ are needed,

$$\delta_+ = \{i \mid i \in \mathcal{S}(N), \delta_i > 0\}, \quad \delta_- = \{i \mid i \in \mathcal{S}(N), \delta_i < 0\}.$$

We shall assume $p \neq p'$, so that δ_+ and δ_- are not empty. Supposing all this, we state Lemma 3.

Lemma 3: If $[p'] \succ [p]$, there exists a mapping $\epsilon : \delta_+ \times \delta_- \to [0, 1]$ such that

$$\sum_{i \in \delta_+} \epsilon_{ij} = |\delta_j|, \quad \text{and} \quad \sum_{j \in \delta_-} \epsilon_{ij} = \delta_i, \quad \text{and} \quad \epsilon_{ij} \neq 0 \Rightarrow i > j.$$

(Independent variables are written as indices.)

Proof of Lemma 3: As p and p' are both probability distributions, we have $\sum_{i \in \mathcal{S}(N)} \delta_i = 0$, or to put it differently, $\sum_{i \in \delta_+} \delta_i = \sum_{j \in \delta_-} |\delta_j|$. Thus, we have two partitions of the interval $[0, \sum_{i \in \delta_+} \delta_i]$ namely,

$$0 < \delta_{i_1} < \delta_{i_1} + \delta_{i_2} < \delta_{i_1} + \delta_{i_2} + \delta_{i_3} < \cdots \sum_{i \in \delta_+} \delta_i,$$

$$0 < |\delta_{j_1}| < |\delta_{j_1}| + |\delta_{j_2}| < \cdots \sum_{j \in \delta_-} |\delta_j|.$$

The sequences of indices $i_\nu \in \delta_+$ and $j_\nu \in \delta_-$ are nothing but the elements of δ_+ and δ_- in their natural order, i.e.,

$$\nu < \mu \Rightarrow i_\nu < i_\mu, \quad \nu < \mu \Rightarrow j_\nu < j_\mu.$$

Now, we construct the product (intersection of intervals) of these partitions, as shown in Fig. 1.

If the interval i_ν with the length δ_{i_ν} has a nonempty intersection with the interval j_μ with the length $|\delta_{j_\mu}|$ we determine $\epsilon_{i_\nu j_\mu}$ to be equal to the length of the interval of intersection. Otherwise, $\epsilon_{i_\nu j_\mu}$ is determined to be 0. Obviously, this construction ensures the conditions

$$\epsilon_{ij} \geq 0, \quad \sum_{i \in \delta_+} \epsilon_{ij} = |\delta_j|, \quad \sum_{j \in \delta_-} \epsilon_{ij} = \delta_i$$

to be valid. We still have to prove $\epsilon_{ij} \neq 0 \Rightarrow i > j$. As p' is more mixed than p and both are in decreasing order, one has $\forall l (\sum_{i=1}^{l} p_i' \leq \sum_{i=1}^{l} p_i)$. This can be expressed by using the quantities δ_i as

$$\forall [l \in \mathcal{S}(N)] \left(\sum_{\substack{i \in \delta_+ \\ i \leq l}} \delta_i \leq \sum_{\substack{j \in \delta_- \\ j \leq l}} |\delta_j| \right).$$

Suppose that $i \in \delta_+$ is given. Specializing the latter assertion, we get

$$\sum_{\substack{t \in \delta_+ \\ t \leq i}} \delta_t \leq \sum_{\substack{h \in \delta_- \\ h \leq i-1}} |\delta_h|.$$

To give an idea of what the meaning of this inequality is, we sketch an example given by Fig. 2.

On the right of the interval i only intervals with indices $j < i$ can be found. This concludes the proof of Lemma 3.

Let us now assume that p and p' are given as in Lemma 3 and the mapping ϵ whose existence has just been ensured, has already been chosen. Let $\{(i, j) \mid (i, j) \in \delta_+ \times \delta_- \wedge \epsilon_{ij} \neq 0\}$ be denoted by M. Then we have Lemma 4.

Lemma 4: Let $(n, m) \in M$ be given. The probability distribution

$$\bar{p}_i = \begin{cases} p_i' & \text{if } i \neq n \wedge i \neq m, \\ p_i' - \epsilon_{nm} & \text{if } i = n, \\ p_i' + \epsilon_{nm} & \text{if } i = m, \end{cases}$$

is more mixed than p, i.e., $[\bar{p}] \succ [p]$.

Proof of Lemma 4: We can assume that $M \setminus \{(n, m)\}$ is not empty, otherwise we would have $p = \bar{p}$ and thus $[\bar{p}] \succ [p]$. Now, choose a pair $(i, j) \in M \setminus \{(n, m)\}$ and denote the permutation matrix which permutes just i and j by Q_1. With the aid of Q_1 we define the probability distribution

$$p^1 = \alpha_1' p + \alpha_1 Q_1 p,$$

where

$$\alpha_1 = \epsilon_{ij}/(p_j - p_i), \quad \alpha_1' = 1 - \alpha_1.$$

On the other hand, p^1 can be written as

$$p_l^1 = \begin{cases} p_l & \text{if } l \neq i \wedge l \neq m, \\ p_l + \epsilon_{ij} & \text{if } l = i, \\ p_l - \epsilon_{ij} & \text{if } l = j. \end{cases}$$

As $(i, j) \in M$ we have $i > j$ and thus $p_i' \leq p_j'$. Using $p_i' = p_i + \sum_{l \in \delta_-} \epsilon_{il}$ and $p_j' = p_j - \sum_{l \in \delta_+} \epsilon_{lj}$ we infer

$$p_i + \epsilon_{ij} \leq p_j - \epsilon_{ij}.$$

This entails of course $\alpha_1 > 0$ and $\alpha_2 > 0$. Thus p^1 is a convex linear combination of p and $Q_1 p$. Therefore, p^1 is more mixed than p. If $p^1 = \bar{p}$ the lemma has already been proved. If $p^1 \neq \bar{p}$ we can choose another pair of indices $(i^*, j^*) \in M \setminus \{(n, m), (i, j)\}$ and define a new probability distribution with the aid of Q_2 permuting just the new pair and using p^1 instead of p, i.e.,

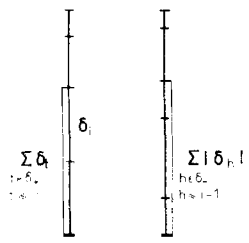


FIG. 2.

$$p^2=\alpha_2' p^1+\alpha_2 Q_2 p^1$$

where

$$\alpha_2=\frac{\epsilon_{i*j*}}{p^1_{j*}-p^1_{i*}},\quad \alpha_2'=1-\alpha_2.$$

Continuing in this way, one obtains $\bar{p}$ by a finite number of steps. With the same argument which has been used to show $[p^1]\succ[p]$ one can easily prove that $[p^{k+1}]\succ[p^k]$ holds for any step. As $\succ$ is transitive, this provides the desired assertion.

Proof of the theorem: Let the assertion

$$\forall(p,p'\in V^{S(N)})([p']\succ[p]\Rightarrow p'\rhd p)$$

be denoted by $G(N)$. The desired assertion $\forall(N\geq 2)\,G(N)$ will be proved by induction on N. Lemma 2 ensures that $G(2)$ is true. It remains to show that $\forall(N\geq 2)\,[G(N)\Rightarrow G(N+1)]$. Suppose $N\geq 2$ is given. Lemma 1 tells us that it is sufficient to deal with probability distributions in decreasing order. Let $p,p'\in V^{S(N+1)}$ be given such that both are in decreasing order and $[p']\succ[p]$. If $[p']=[p]$ we have $p'=p$ because they are both in decreasing order, and then $p'\rhd p$ is true. If $[p']\neq[p]$ we construct the map ϵ using the method which was successful in proving Lemma 3 (see Fig. 1). If $\delta_{i_1}\leq|\delta_{j_1}|$ we denote the index i_1 by n and the index j_1 by m, otherwise, we label i_1 as m and j_1 as n. After doing this, we construct the probability distribution

$$\bar{p}_l=\begin{cases} p_l+\sum\limits_{\substack{j\in\delta_-\\ j\neq n}}\epsilon_{lj} & \text{if } l\in\delta_+\setminus\{n\},\\ p_l-\sum\limits_{\substack{i\in\delta_+\\ i\neq n}}\epsilon_{il} & \text{if } l\in\delta_-\setminus\{n\},\\ p_l & \text{if } l\in S(N+1)\setminus(\delta_+\cup\delta_-),\\ p_n & \text{if } l=n.\end{cases}$$

$\bar{p}$ can be expressed by means of p', i.e.,

$$\bar{p}_l=\begin{cases} p'_l & \text{if } l\notin\{n,m\},\\ p'_n-\epsilon_{nm} & \text{if } l=n,\\ p'_m+\epsilon_{nm} & \text{if } l=m,\end{cases}\qquad \text{if } n\in\delta_+$$

$$\bar{p}_l=\begin{cases} p'_l & \text{if } l\notin\{n,m\},\\ p'_n+\epsilon_{mn} & \text{if } l=n,\\ p'_m-\epsilon_{mn} & \text{if } l=m.\end{cases}\qquad \text{if } n\in\delta_-$$

According to Lemma 4, this entails $[\bar{p}]\succ[p]$. Because $p_n=\bar{p}_n$, the distributions p and $\bar{p}$ can only differ in the remaining N elements. n is defined in such a way that $p_n\neq 1$. Let us denote the restriction of p and $\bar{p}$ to $S(N+1)\setminus\{n\}$ by q and $\bar{q}$. The probability distribution (on $S(N+1)\setminus\{n\}$) $[1/(1-p_n)]\bar{q}$ is more mixed than $[1/(1-p_n)]q$. This can easily be inferred from $[\bar{p}]\succ[p]$ by using $\forall p,p'\{[p']\succ[p]\leftrightarrow\forall(l\in\mathbb{R})(\sum_i|p'_i-l|\leq\sum_i|p_i-l|)\}$. Thus, according to $G(N)$, there exists a unitary $N\times N$ matrix $\tilde{U}$ such that

$$\bar{q}_i=\sum_j|\tilde{U}_{ij}|^2 q_j.$$

$\tilde{U}$ can be enlarged in such a way as to obtain a unitary $(N+1)\times(N+1)$ matrix U such that $\bar{p}_i=\sum_j|U_{ij}|^2 p_j$, i.e.,

$$U_{ij}=\begin{cases}\tilde{U}_{ij} & \text{if } i,j\neq n,\\ 0 & \text{if } (i=n\wedge j\neq n)\vee(i\neq n\wedge j=n),\\ 1 & \text{if } i=j=n.\end{cases}$$

To give an idea of what U looks like, we write it schematically in its explicit form, i.e.,

$$\begin{pmatrix} U_{11} & U_{12}\cdots & 0 & U_{1n+1}\cdots\\ U_{21} & U_{22}\cdots & 0 & \vdots\\ \vdots & \vdots & & \vdots\\ 0 & 0 & 1 & 0\ 0\cdots\\ U_{n+1,1} & \cdots & 0 & U_{n+1,n+1}\cdots\\ \vdots & & 0 & \vdots\\ & & \vdots & \end{pmatrix}.$$

If we denote the permutation matrix which permutes just n and m by P and the unit matrix by $\mathbf{1}$, we can write p' as

$$p'=(\alpha\mathbf{1}+\beta P)\bar{p},$$

where α and β are defined by

$\beta=\epsilon_{nm}/(\bar{p}_m-\bar{p}_n)$,	$\beta=\epsilon_{mn}/(\bar{p}_n-\bar{p}_m)$,
$\alpha=1-\beta$,	$\alpha=1-\beta$,
if $n\in\delta_+$,	if $n\in\delta_-$.

We have $\alpha\geq 0$ and $\beta\geq 0$, because

$p'_n\leq p'_m$,	$p'_m\leq p'_n$,
$\bar{p}_n+\epsilon_{nm}\leq\bar{p}_m-\epsilon_{nm}$,	$\bar{p}_m+\epsilon_{mn}\leq\bar{p}_n-\epsilon_{mn}$,
$0<2\epsilon_{nm}\leq\bar{p}_m-\bar{p}_n$,	$0<2\epsilon_{mn}\leq\bar{p}_n-\bar{p}_m$,
if $n\in\delta_+$,	if $n\in\delta_-$.

Let us now have a look at the unitary matrix V defined by

$$V_{ij}=\begin{cases} 1 & \text{if } i=j\wedge i\neq n\wedge i\neq m,\\ \sqrt{\alpha} & \text{if } i=j\wedge(i=n\vee i=m),\\ \sqrt{\beta} & \text{if } i=n\wedge j=m,\\ -\sqrt{\beta} & \text{if } i=m\wedge j=n,\\ 0 & \text{elsewhere.}\end{cases}$$

It looks schematically like

$$\begin{pmatrix} 1 & 0\cdots & & & & \\ 0 & 1\cdot & & & & \\ \vdots & \ddots & & & & \\ & & \sqrt{\alpha} & & \sqrt{\beta} & \\ & & & 1\cdot & & \\ & & -\sqrt{\beta} & & \sqrt{\alpha} & \\ & & & & & 1\cdot\ \ddots \end{pmatrix}.$$

Obviously we have $(\alpha\mathbf{1}+\beta P)_{ij}=|V_{ij}|^2$. Combining this with the previous result, we get

$$p'_i=\sum_j\Big(\sum_l|V_{il}|^2|U_{lj}|^2\Big)p_j.$$

As both U and V are unitary, $\sum_l V_{il}U_{lj}$ is unitary too. If we can show that $|\sum_l V_{il}U_{lj}|^2 = \sum_l |V_{il}U_{lj}|^2$, the proof will be complete. By using the definition of V and U it can easily be shown that the sum $\sum_l |V_{il}U_{lj}|^2$ consists of one term at the most. This ensures that $\sum_l |V_{il}U_{lj}|^2 = |\sum_l V_{il}U_{lj}|^2$ for any i and j and concludes the proof of the theorem.

THE PHYSICAL MEANING OF THE THEOREM

Let us consider a quantum-mechanical system whose density operator initially ($t=0$) has a finite (N) dimensional range H^N, where H^N is an eigenspace of some invariant of the motion. The density operator $\rho(t)$ will then have this range for all times $t \geq 0$. Thus H^N can be regarded as the space of states. Let $\{|1\rangle, \ldots, |N\rangle\}$ be an orthonormal basis of H^N and suppose that the density operator initially has the form

$$\rho(0) = \sum_{i=1}^{N} |i\rangle p_i \langle i|.$$

Let $t > 0$ be given. The density operator at time t is

$$\rho(t) = \exp[-(i/\hbar)Ht]\rho(0)\exp[(i/\hbar)Ht],$$

where H is the Hamiltonian operator of the system. We are interested in assertions concerning the diagonal of $\rho(t)$, i.e., $p_i(t) = \langle i|\rho(t)|i\rangle$. The number p_i and $p_i(t)$ can be lumped into probability distributions p and $p(t)$ so that that the notion of mixing character applies. We call an assertion concerning p and $p(t)$ general if it is independent of the special properties of the system (represented by the Hamiltonian operator) and if it does not refer to special initial distributions only. Or to put it differently, a general assertion can always be written in the form

$$\forall(H \in \mathcal{H}am)\ \forall(p)\, F(p, p(t)),$$

where $\mathcal{H}am$ is the set of all operators which can possibly be the Hamiltonian operator of the system. As H^N is finite dimensional, $\mathcal{H}am$ is nothing but the set of all Hermitian operators. $F(p,p')$ is an assertion that depends on the probability distributions p and p' but it must not explicitly depend on H. We call the general assertion $\forall(H \in \mathcal{H}am)\ \forall p\, S(p, p(t))$ better than $\forall(H \in \mathcal{H}am)\ \forall p\, F(p, p(t))$ if the condition

$$\forall(p)\ \forall(p')(S(p,p') \Longrightarrow F(p,p'))$$

holds. Now let us look at $p(t)$:

$$p_i(t) = \langle i|\exp[-(i/\hbar)Ht]\rho(0)\exp[(i/\hbar)Ht]|i\rangle$$
$$= \sum_j |\langle i|\exp[-(i/\hbar)Ht]|j\rangle|^2 p_j.$$

As $\exp[-(i/\hbar)Ht]$ is a unitary operator, this entails

$$\forall(H \in \mathcal{H}am)\ \forall(p)(p(t) \rhd p)$$

Thus $p' \rhd p$ provides a true general assertion. As every unitary operator can be written as $\exp(iA)$ where A is Hermitian, we have

$$\forall(p)\ \forall(p')\{p' \rhd p \Longrightarrow \exists(H \in \mathcal{H}am)[p' = p(t)]\}.$$

Let $F(\cdot,\cdot)$ be given such that $\forall(H \in \mathcal{H}am)\ \forall(p)\, F(p, p(t))$ is true. We then infer from the latter assertion that

$$\forall(p)\ \forall(p')[p' \rhd p \Rightarrow F(p,p')].$$

Thus $\forall(H \in \mathcal{H}am)\ \forall(p)(p(t) \rhd p)$ is better than an arbitrary true general assertion and therefore it is the best true general assertion. Our theorem tells us that this best assertion is equivalent to $\forall(H \in \mathcal{H}am)\ \forall(p)([p(t)] \succ [p])$. Ruch's principle of increasing mixing character[1,2] is therefore distinguished.

The assertion $[p(t)] \succ [p]$ can be interpreted very intuitively. Let us assume that, referring to a certain set of instruments, the diagonal of $\rho(t)$ is the only part of $\rho(t)$ that is experimentally relevant at time t. The knowledge which is represented by the off-diagonal elements has then lost its relevance (later it may of course become relevant again). In the course of time a part of the knowledge which was initially located in the diagonal of $\rho(0)$ flows into the off-diagonal elements and loses its relevance thereby. This phenomenon of relevancy-loss is described by the assertion $[p(t)] \succ [p]$ and our theorem ensures that this is the best and hence a complete description. For many reasons it will be sensible to require that any principle of statistical physics which describes the knowledge-decay that occurs if an isolated system approaches to equilibrium provides a description of the phenomenon of relevancy-loss as well. (The entropy principle, for example, provides a description of the phenomenon of relevancy loss.) As according to our theorem the assertion $[p(t)] \succ [p]$ is the best description of this phenomenon, we are justified in calling Ruch's principle of increasing mixing character the strongest principle of statistical physics, that means, any other principle can be inferred from this one.

ACKNOWLEDGMENT

I wish to thank Professor Dr. A. Schönhofer for very valuable discussions.

[1] E. Ruch, Theor. Chim. Acta 38, 167 (1975).
[2] E. Ruch and A. Mead, Theor. Chim. Acta (in print).

Matrix elements of multiple electron exchange between spin functions of maximum multiplicity*

M. D. Girardeau

Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403
(Received 8 September 1975)

A general formula for such matrix elements is obtained, in terms of a sum of terms each proportional to the square root of a product of eight binomial coefficients.

1. INTRODUCTION

Consider a wavefunction $\varphi_{ns}(x_1 \cdots x_N)$ describing the N electrons in an unfilled atomic shell. Here $x_j = (\mathbf{r}_j, \sigma_j)$ with $\mathbf{r}_j$ the position vector and $\sigma_j = \pm\frac{1}{2}$ the spin Z component variable of the jth electron. We assume that φ_{ns} is an eigenstate of the N-electron spin with eigenvalue S, and of the N-electron spin z component with eigenvalue $s = -S, -S+1, \ldots, S$; the label S is suppressed since it will be assumed to have the same fixed value for all wavefunctions φ_{ns} involved. The label n stands for all other quantum numbers necessary to specify the state. If direct spin—spin (dipole—dipole) and spin—orbit interactions are negligible and the shell is not more than half-filled, then those φ_{ns} describing the $(2S+1)$-fold degenerate ground state are expressible according to Hund's rule in the form

$$\varphi_{ns}(x_1 \cdots x_N) = u_n(\mathbf{r}_1 \cdots \mathbf{r}_N)\chi_s(\sigma_1 \cdots \sigma_N), \tag{1}$$

where the spatial function u_n is completely antisymmetric, the spin function χ_s is completely symmetric, and the total spin $S = \frac{1}{2}N$. In a case in which the shell is more than half-filled, φ_{ns} must be expressed not as a single product (1), but as a sum of such products in which the spatial factor is not completely antisymmetric and the spin factor not completely symmetric. Such more general cases will not be considered in this paper, although the spin parts of exchange matrix elements between such wavefunctions might be derivable by a generalization of the method used here.

A product wavefunction $\varphi_{ns}(x_1 \circ\circ\circ x_N)\varphi_{ms'}(x_1' \circ\circ\circ x_N')$ is not completely antisymmetric under all permutations of the $2N$ electrons $(x_1 \circ\circ\circ x_N x_1' \circ\circ\circ x_N')$ but can be made antisymmetric by premultiplication by the $2N$-electron antisymmetrizer $\mathcal{A}_{2N}$. Let A be a permutation-invariant operator representing any physical observable independent of the spin variables $(\sigma_1 \circ\circ\circ \sigma_N \sigma_1' \circ\circ\circ \sigma_N')$. The matrix element of A between two wavefunctions of the form $\mathcal{A}_{2N}\varphi_{ns}\varphi_{ms'}$ is

$$\begin{aligned}&(n_1 s_1, n_2 s_2 \,|A|\, n_3 s_3, n_4 s_4)\\ &\quad = \int \varphi_{n_1 s_1}^*(x_1 \cdots x_N)\varphi^*_{n_2 s_2}(x_1' \cdots x_N')\\ &\times A\mathcal{A}_{2N}\varphi_{n_3 s_3}(x_1 \cdots x_N)\varphi_{n_4 s_4}(x_1' \cdots x_N')\\ &\times dx_1 \cdots dx_N\, dx_1' \cdots dx_N',\end{aligned} \tag{2}$$

where each $\int dx$ stands for an integration over $\mathbf{r}$ and a summation over $\sigma = \pm\frac{1}{2}$; the projection property $\mathcal{A}_{2N}^2 = \mathcal{A}_{2N}$ of the antisymmetrizer has already been used in eliminating one factor of $\mathcal{A}_{2N}$ from the matrix element. Upon writing $\mathcal{A}_{2N}$ as a sum of terms involving all possible permutations of the $2N$ variables $(x_1 \circ\circ\circ x_N x_1' \circ\circ\circ x_N')$ and noting that the product $\varphi_{ns}(x_1 \circ\circ\circ x_N)\varphi_{ms'}(x_1' \circ\circ\circ x_N')$ is already antisymmetric under *intra*-atomic permutations, i.e., those involving only permutations *within* the set $(x_1 \circ\circ\circ x_N)$ and/or within $(x_1' \circ\circ\circ x_N')$, one can write (2) as a sum of terms involving j-fold *inter*atomic exchange, with j running from 0 (no exchange) to N. The term involving j-fold interatomic exchange is found from (1) to be proportional[1] to

$$M_j(n_1 \circ\circ\circ n_4)I_j(s_1 \cdots s_4) \tag{3}$$

where M_j is the matrix element representing the coupling between the observable A and j-fold spatial interatomic electron exchange,

$$\begin{aligned}&M_j(n_1 \cdots n_4)\\ &\quad = \int u^*_{n_1}(\mathbf{r}_1 \cdots \mathbf{r}_N)u^*_{n_2}(\mathbf{r}_1' \cdots \mathbf{r}_N')\\ &\times AU_{n_3}(\mathbf{r}_1' \cdots \mathbf{r}_j'\mathbf{r}_{j+1} \cdots \mathbf{r}_N)u_{n_4}(\mathbf{r}_1 \cdots \mathbf{r}_j\mathbf{r}_{j+1}' \cdots \mathbf{r}_N')\\ &\times d^3r_1 \cdots d^3r_N\, d^3r_1' \cdots d^3r_N',\end{aligned} \tag{4}$$

and I_j is the matrix element of j-fold interatomic electron spin exchange,

$$\begin{aligned}&I_j(s_1 \circ\circ\circ s_4)\\ &\quad = \sum_{\sigma_1 \cdots \sigma_N}\ \sum_{\sigma_1' \cdots \sigma_N'} \chi_{s_1}(\sigma_1 \cdots \sigma_N)\chi_{s_2}(\sigma_1' \cdots \sigma_N')\\ &\times \chi_{s_3}(\sigma_1' \cdots \sigma_j'\sigma_{j+1} \cdots \sigma_N)\chi_{s_4}(\sigma_1 \cdots \sigma_j\sigma_{j+1}' \cdots \sigma_N').\end{aligned} \tag{5}$$

Note that M_j carries all the dependence on the observable A and on the spatial quantum numbers $(n_1 \circ\circ\circ n_4)$ but is independent of the spin z-component quantum numbers $(s_1 \cdots s_4)$, whereas I_j carries all the dependence on $(s_1 \circ\circ\circ s_4)$ but in independent of the observable A and of the spatial quantum numbers $(n_1 \cdots n_4)$. The phases of the χ_s can be chosen so that they are all real and positive; reality has already been assumed in (5). Since the χ_s are also assumed to be normalized, one has

$$0 \le I_j(s_1 \cdots s_4) \le 1 \tag{6}$$

by the Schwartz inequality.

In the investigations of the magnetic effects of many-electron exchange[2–4] which served as motivation[5] for the calculation reported here, the relevant observables A were the interatomic Coulomb interaction and the total intra-atomic Hamiltonians. However, since the I_j are independent of A, they are of more general significance. In Sec. 3 we shall sketch the derivation of the general formula for the I_j as a function of $(s_1 \cdots s_4)$, j, and N. Before doing so, in Sec. 2 we shall point out some rather

obvious selection rules which severely restrict the combinations of spin quantum numbers for which the matrix elements I_j are nonvanishing.

2. SELECTION RULES

In the first place, since electron exchange between two atoms does not change the sum of the spin z components of the two atoms, we have the selection rule

$$s_1 + s_2 = s_3 + s_4 \tag{7}$$

or

$$\Delta s = -\Delta s' \tag{8}$$

where Δs and $\Delta s'$ are the changes in spin z components of the two atoms whose spin wavefunctions are involved in (5),

$$\Delta s = s_1 - s_3, \quad \Delta s' = s_2 - s_4. \tag{9}$$

In addition, since each electron has total spin $\frac{1}{2}$, the maximal change of spin z component of each atom as a result of j-fold electron exchange is equal to j,

$$|\Delta s| \leq j. \tag{10}$$

The I_j necessarily vanish for any combination of $(s_1 \cdots s_4)$ violating (8) and/or (10).

3. EVALUATION OF I_j

Expressions for I_1 and I_2 for certain values of N have been worked out by Popović-Božić[2]; the calculation described here leads to a general expression for all values of N and all j, $1 \leq j \leq N$, by a different method based on combinatorial analysis.

According to Hund's rule, the χ_s in (5) are totally symmetric and belong to total spin $S = \frac{1}{2}N$. Recalling that we have assumed a choice of phase such that they are real and positive, the normalized χ_s of maximal spin z component $s = S = \frac{1}{2}N$ is

$$\chi_{(1/2)N}(\sigma_1 \cdots \sigma_N) = \delta_{\sigma_1(1/2)} \cdots \delta_{\sigma_N(1/2)} = \prod_{i=1}^{N} \delta_{\sigma_i(1/2)}, \tag{11}$$

where $\delta_{\sigma(1/2)}$ is the usual Kronecker delta, equal to unity if $\sigma = \frac{1}{2}$ and zero if $\sigma = -\frac{1}{2}$. The other χ_s can be generated by repeated application of the spin lowering operator S^-, defined implicitly by

$$(s' | S^- | s) = [(\tfrac{1}{2}N + s)(\tfrac{1}{2}N - s + 1)]^{1/2} \sigma_{s',s-1} \tag{12}$$

and explicitly by

$$S^- \chi(\sigma_1 \cdots \sigma_N) = \sum_{i=1}^{N} \chi(\sigma_1 \cdots \sigma_{i-1}, \tfrac{1}{2}, \sigma_{i+1} \cdots \sigma_N) \delta_{\sigma_i, -1/2}. \tag{13}$$

This iterates to

$$(S^-)^p \chi(\sigma_1 \cdots \sigma_N) = \frac{N!}{(N-p)!} (S^-)^p \chi_{(1/2)N}(\sigma_1 \cdots \sigma_N) \times \delta_{\sigma_{N-p+1}, -1/2} \cdots \delta_{\sigma_N, -1/2}], \tag{14}$$

where $\mathcal{S}_N$ is the N-spin symmetrizer.[6] For the special case (11) this reduces to

$$(S^-)^p \chi_{1/2N}(\sigma_1 \cdots \sigma_N) = \frac{N!}{(N-p)!} \mathcal{S}_N (\delta_{\sigma_1(1/2)} \cdots \delta_{\sigma_{N-p}(1/2)} \delta_{\sigma_{N-p+1}, -1/2} \cdots \delta_{\sigma_N, -1/2}). \tag{15}$$

By iteration of the standard expression[7] relating χ_s and χ_{s-1} [cf. (12)] one has

$$\chi_s(\sigma_1 \cdots \sigma_N) = \left[\frac{(\frac{1}{2}N + s)!}{N!(\frac{1}{2}N - s)!}\right]^{1/2} (S^-)^{(1/2)N - s} \times \chi_{(1/2)N}(\sigma_1 \cdots \sigma_N), \qquad -\tfrac{1}{2}N \leq s \leq \tfrac{1}{2}N. \tag{16}$$

Then by (15)

$$\chi_s(\sigma_1 \cdots \sigma_N) = \left[\frac{N!}{(\frac{1}{2}N + s)!(\frac{1}{2}N - s)!}\right]^{1/2} \mathcal{S}_N (\delta_{\sigma_1(1/2)} \cdots \delta_{\sigma_{1/2N+s}(1/2)} \times \delta_{\sigma_{(1/2)N+s+1}, -1/2} \cdots \delta_{\sigma_N, -1/2}). \tag{17}$$

It follows that χ_s is only nonzero if $(\frac{1}{2}N + s)$ of the σ_i are equal to $\frac{1}{2}$ and the remaining $(\frac{1}{2}N - s)$ are equal to $-\frac{1}{2}$, and that this nonzero value is the same for every set of σ_i satisfying this criterion. The constant value is easily found by normalization, noting that there are $[N!/(\frac{1}{2}N + s)! \times (\frac{1}{2}N - s)!]$ terms in the sum over $\sigma_1 \cdots \sigma_N$ satisfying the aforementioned criterion. Thus

$$\chi_s(\sigma_1 \cdots \sigma_N) = [(\tfrac{1}{2}N + s)!(\tfrac{1}{2}N - s)!/N!]^{1/2} \delta_s(\sigma_1 \cdots \sigma_N) \tag{18}$$

with

$$\delta_s(\sigma_1 \cdots \sigma_N) = \begin{cases} 1, & \text{if } (\frac{1}{2}N + s)\ \sigma_i \text{ are } \frac{1}{2} \text{ and } (\frac{1}{2}N - s) \text{ are } -\frac{1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The same result can be found by using the definition[6] of $\mathcal{S}_N$ in (17) and counting the number of permutations giving rise to the same combination of values of $\sigma_1 \cdots \sigma_N$.

Upon substituting (18) into (5) and noting the selection rules (7)—(10), one can write the matrix element I_j in the form

$$I_j(s_1 \cdots s_4) = (N!)^{-2} \sum_{k=-j}^{j} [(\tfrac{1}{2}N + s_1)!(\tfrac{1}{2}N + s_1 - k)!(\tfrac{1}{2}N - s_1)!(\tfrac{1}{2}N - s_1 + k)! \times (\tfrac{1}{2}N + s_2)!(\tfrac{1}{2}N + s_2 + k)!(\tfrac{1}{2}N - s_2)!(\tfrac{1}{2}N - s_2 - k)!]^{1/2} \times \delta_{s_3, s_1 - k} \delta_{s_4, s_2 + k} c(j,k;s_1 s_2), \tag{20}$$

with

$$c(j,k;s_1 s_2) = \sum_{\sigma_1 \cdots \sigma_N} \sum_{\sigma_1' \cdots \sigma_N'} \delta_{s_1}(\sigma_1 \cdots \sigma_N) \delta_{s_2}(\sigma_1' \cdots \sigma_N') \times \delta_{s_1 - k}(\sigma_1' \cdots \sigma_j' \sigma_{j+1} \cdots \sigma_N) \delta_{s_2 + k}(\sigma_1 \cdots \sigma_j \sigma_{j+1}' \cdots \sigma_N') = \sum_{\substack{\sigma_1 \cdots \sigma_N \\ (\sigma_1' + \cdots + \sigma_j' = \sigma_1 + \cdots + \sigma_j - k)}} \sum_{\sigma_1' \cdots \sigma_N'} \delta_{s_1}(\sigma_1 \cdots \sigma_N) \delta_{s_2}(\sigma_1' \cdots \sigma_N'), \tag{21}$$

where use has been made of the fact that with the indicated restrictions on the summations, $\delta_{s_1 - k}$ is nonzero if and only if δ_{s_1} is, and similarly for $\delta_{s_2 + k}$ and δ_{s_2}.

To evaluate this combinatorial factor, note first that $\delta_s(\sigma_1\cdots\sigma_N)$ is nonzero if and only if $(\frac{1}{2}N+s_1)$ of the σ_i are $+\frac{1}{2}$, and $(\frac{1}{2}N-s_1)$ are $-\frac{1}{2}$. Let n be the number of the set $(\sigma_1\cdots\sigma_j)$ which are $+\frac{1}{2}$; then $(j-n)$ of the same set are $-\frac{1}{2}$, $(\frac{1}{2}N+s_1-n)$ of $(\sigma_{j+1}\cdots\sigma_N)$ are $+\frac{1}{2}$, and $(\frac{1}{2}N-s_1-j+n)$ of $(\sigma_{j+1}\cdots\sigma_N)$ are $-\frac{1}{2}$. Furthermore, because of the constraint on the summation in (21), $(n-k)$ of the set $(\sigma_1'\cdots\sigma_j')$ are $+\frac{1}{2}$, $(j-n+k)$ of the same set are $-\frac{1}{2}$, $(\frac{1}{2}N+s_2-n+k)$ of the set $(\sigma_{j+1}'\cdots\sigma_N')$ are $+\frac{1}{2}$, and $(\frac{1}{2}N-s_2-j+n-k)$ of $(\sigma_{j+1}'\cdots\sigma_N')$ are $-\frac{1}{2}$. For each possible value of n [the number of the set $(\sigma_1\cdots\sigma_j)$ equal to $+\frac{1}{2}$], the contribution to c is the number of choices of the set $(\sigma_1\cdots\sigma_N\sigma_1'\cdots\sigma_N')$ consistent with these constraints. Consider any choice of $(\sigma_1\cdots\sigma_j)$ such that n are $+\frac{1}{2}$ and hence $(N-j)$ are $-\frac{1}{2}$. There are $j!$ possible permutations of these j quantities, but the corresponding combinatorial factor is only $[j!/n!(j-n)!]$, since the permutations of the n $+\frac{1}{2}$'s among themselves or of the $(j-n)$ $-\frac{1}{2}$'s among themselves do not correspond to distinct choices of the summation indices $(\sigma_1\cdots\sigma_j)$. Determining the combinatorial factors for $(\sigma_{j+1}\cdots\sigma_N)$, $(\sigma_1'\cdots\sigma_j')$, and $(\sigma_{j+1}'\cdots\sigma_N')$ similarly, one finds

$$c(j,k;s_1s_2)=\sum_n\left[\frac{j!}{n!(j-n)!}\right]\left[\frac{(N-j)!}{(\frac{1}{2}N+s_1-n)!(\frac{1}{2}N-s_1-j+n)!}\right]\times\left[\frac{j!}{(n-k)!(j-n+k)!}\right]\times\left[\frac{(N-j)!}{(\frac{1}{2}N+s_2-n+k)!(\frac{1}{2}N-s_2-j+n-k)!}\right]. \tag{22}$$

To make this expression more explicit it is necessary to determine the possible values of the summation index n. Note that j can take on all integral values from 1 (exchange of a single electron) to N (exchange of all N electrons), whereas k can take on all integral values from $-j$ to j. Both s_1 and s_2 can take on the $(N+1)$ values $-\frac{1}{2}N, -\frac{1}{2}N+1,\cdots,\frac{1}{2}N$, where N is an integer ≥ 2. The summation index n then ranges over all integer values such that the quantities n, $(j-n)$, $(\frac{1}{2}N+s_1-n)$, $(\frac{1}{2}N-s_1-j+n)$, $(n-k)$, $(j-n+k)$, $(\frac{1}{2}N+s_2-n+k)$, and $(\frac{1}{2}N-s_2-j+n-k)$ are all nonnegative. These restrictions are all incorporated if one rewrites (22) in the form

$$c(j,k;s_1s_2)=\sum_{n=k}^{j}\binom{j}{n}\binom{N-j}{\frac{1}{2}N+s_1-n}\binom{j}{n-k}\binom{N-j}{\frac{1}{2}N+s_2-n+k} \tag{23}$$

and notes that the binomial coefficient $\binom{n_1}{n_2}$ vanishes if either n_2 or (n_1-n_2) is a negative integer.[8]

The desired expression for I_j, found by substitution of (23) into (20) and use of appropriate identities for binomial coefficients (or their definition), is

$$I_j(s_1\cdots s_4)=\left(\frac{N}{j}\right)^{-2}\sum_{k=-j}^{j}\sum_{n=k}^{j}\left[\binom{\frac{1}{2}N+s_1}{n}\times\binom{\frac{1}{2}N+s_1-k}{n-k}\binom{\frac{1}{2}N-s_1}{j-n}\binom{\frac{1}{2}N-s_1+k}{j-n+k}\times\binom{\frac{1}{2}N+s_2}{n-k}\binom{\frac{1}{2}N+s_2+k}{n}\times\binom{\frac{1}{2}N-s_2}{j-n+k}\binom{\frac{1}{2}N-s_2-k}{j-n}\right]^{1/2}\delta_{s_3,s_1-k}\delta_{s_4,s_2+k}. \tag{24}$$

The desired constraints on the range of the summation over n again follow from the vanishing of $\binom{n_1}{n_2}$ when either n_2 or (n_1-n_2) is negative.[8]

ACKNOWLEDGMENT

I wish to thank Mirjana Popović-Božić for prepublication copies of some of her work and for correspondence which served as motivation for this work.

*Work supported in part by the National Science Foundation, Grant DMR72-03211.

[1]The proportionality constant is merely $(-1)^j$ times a numerical combinatorial factor.

[2]M. Popović-Božić, Phys. Status Solidi B 65, 305 (1974); see also M. Popović-Božić, in *Magnetism and Magnetic Materials*-1974, AIP Conference Proceedings No. 24 (American Institute of Physics, New York, 1975), pp. 331, 332.

[3]Z. Marić and M. Popović-Božić, Ann. Phys. (NY) (to be published).

[4]M.D. Girardeau and M. Popović-Božić, J. Phys. C (to be published).

[5]The calculations in Refs. 2–4 were all carried out in second quantization representations, rather than in the Schrödinger representation implied by (2). However, the same matrix elements (3) have to be evaluated in the end, although use of second quantization simplifies the combinatorics, as usual.

[6]Explicitly one has $\zeta_N=(N!)^{-1}\Sigma_P P$ where P is one of the $N!$ permutations of $(\sigma_1'\cdots\sigma_N)$.

[7]See, e.g., A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, New Jersey, 1957), p. 17.

[8]Note that s_1 and s_2 are half-integral if and only if N is odd, since $S=\frac{1}{2}N$. Hence, in this case n_1 and n_2 in $\binom{n_1}{n_2}$ in (23) are always integers. In this case the statement that

$$\binom{n_1}{n_2}=\Gamma(n_1+1)/\Gamma(n_2+1)\Gamma(n_1-n_2+1)$$

vanishes if n_2 or (n_1-n_2) is a negative integer can be replaced by the statement that $\binom{n_1}{n_2}$ vanishes if n_2 or (n_1-n_2) is negative, which implies the desired summation constraint.

Multipole expansion of the density of states about a crystal cell*

G. Strinati and U. Fano

Department of Physics, University of Chicago, Chicago, Illinois 60637
(Received 24 September 1975)

We construct the expansion of a Bloch wave with energy E into a complete set of multipole waves around a "center" of a crystal as an analog of the expansion in spherical waves in free space. Crystal point group symmetry is used to classify the set. The density of states in a cell is then analyzed into multipole components whose magnitude depends on the cell's distance from the "center."

1. INTRODUCTION

The band theory of crystalline solids considers a single expression for the electronic density of states of each band, states which are regarded as fully delocalized with periodic probability over the whole crystal. Emphasis on delocalized states becomes, however, inappropriate when treating phenomena which select a particular lattice site, as, for example, in the presence of an isolated impurity or in photoabsorption from inner shells which leaves a localized hole. Experimental evidence of the nonuniqueness of the relevant density of final states can be seen, for example, in the difference between K and L spectra of solid Al,[1] which has been accounted for recently in a qualitative way by Hayes and Sen.[2] In effect, the introduction of a "center" spoils the translational invariance of the lattice and shifts the analysis toward a local point of view.

Central symmetry is, of course, essential to the states of isolated atoms. Here, orbital momentum eigenstates represent a very natural basis for analysis. Indeed one can resolve the density of states into contributions from separate orbital momenta. In crystals, however, anisotropy spoils the conservation of angular momentum and a corresponding analysis of the density of states requires the construction of a new suitable basis. Orbital momentum eigenstates are replaced in a crystalline medium by states that transform according to the irreducible representations of the appropriate crystal group. However, the number of representations of finite groups and their dimensionality are finite in contrast to the infinity of angular momentum eigenvalues. In a crystal, therefore, a complete basis must include an infinity of states which transform according to the same row of the same irreducible representation. What we need, then, is a systematic classification of such states. The "appropriate" crystal groups we consider in this paper for the classification of the new set of states are the isogonal point groups of the space groups, whose elements, together with the inversion, transform the constant energy surface onto itself.

A solution to this problem has been sketched in a brief communication.[3] In this paper we develop the solution in some detail, with explicit application of the point group symmetry, and with the specific aim of resolving the density of states in a cell into contributions from different multipole waves. Further applications remain to be developed.

2. MULTIPOLE WAVE EXPANSION OF THE DENSITY OF STATES

In accordance with the local character of impurity effects and related phenomena, we consider a set of localized Wannier wavefunctions $\langle \mathbf{r}|t\mathbf{n}\rangle$ of an electron, where t is a band index and $\mathbf{n}$ is a lattice vector. We shall drop the band index throughout since all our considerations will refer to a simple band. A Bloch wave, eigenfunction of the perfect lattice Hamiltonian corresponding to the energy E, can be expressed as a superposition of Wannier functions:

$$\langle \mathbf{r}|\mathbf{k}\rangle_E = \sum_{\mathbf{n}} \langle \mathbf{r}|\mathbf{n}\rangle \langle \mathbf{n}|\mathbf{k}\rangle_E,$$
$$\langle \mathbf{n}|\mathbf{k}\rangle_E = w_E^{1/2}(\mathbf{k}) \exp(i\mathbf{n}\cdot\mathbf{k}). \tag{2.1}$$

The wave vector $\mathbf{k}$ is restricted to the first Brillouin zone and ranges over the constant energy surface defined by the dispersion relation

$$D(E;\mathbf{k}) = 0, \tag{2.2}$$

which we assume to be known and which incorporates the crystal field properties relevant to our problem. The amplitude $\langle \mathbf{n}|\mathbf{k}\rangle_E$ includes both the phase factor appropriate to the nth lattice point and a normalization coefficient. The definition (2.1) sets the phase of the Bloch wave at zero in the central cell, $\mathbf{n}=0$, in accordance with the recent work of Kohn.[4] Since we will work at a fixed energy E, the Bloch wave (2.1) will be normalized per unit range of energy and of the solid angle centered around the wave vector's direction $\mathbf{k}$. The density of states is thus incorporated in the normalization of the wavefunctions, by identifying the coefficient of (2.1) as:

$$w_E(\mathbf{k}) = \frac{\Omega_c}{(2\pi)^3} \left(\frac{\partial(E,\hat{k})}{\partial(k_x,k_y,k_z)} \right)^{-1}$$
$$= \frac{\Omega_c}{(2\pi)^3} \frac{k^2}{|\hat{k}\cdot\nabla_{\mathbf{k}} E(\mathbf{k})|}, \tag{2.3}$$

where Ω_c is the volume of the unit cell (which we take as the symmetrical Wigner–Seitz cell) and the last factor takes into account the obliquity of the constant energy surface. The Jacobian of the transformation is evaluated from the dispersion relation (2.2).

The points of the constant energy surface are in one to one correspondence with two continuous parameters (the polar angles of $\mathbf{k}$). Our goal is to replace the two-

parameter set of Bloch waves (2.1) of given energy by a new set, whose elements will be called multipole waves because of their analogy to the orbital momentum eigenstates of free space. These waves will be labeled by indices which are discrete but must run over an infinite range of values. Their construction by a unitary transformation of the set (2.1) will preserve the normalization per unit energy.

The space variables of the new set of multipole waves are the coordinates n of the Bravais lattice nodes, whose point-symmetry elements form the holosymmetric point group of the crystal system. This group always includes the space inversion. However, since we are constructing the multipole waves at fixed energy E, we will classify them according to the irreducible representations of the symmetry group of the constant energy surface $E(\mathbf{k}) = E$ which is a subgroup of the holosymmetric group. We call Γ an irreducible representation and i one of its rows. As previously noted, the set of Γ and i is finite and, therefore, insufficient to classify a complete set. Thus, *for each* Γ *and* i, we require a further set of *two* discrete indices L and q which can run over an infinite range of values and whose meaning remains to be determined.

The matrix elements $\langle \Gamma i L q \,|\, \mathbf{k}\rangle_E$ of the unitary transformation, which we seek to construct, constitute the coefficients of the expansion of the Bloch wave amplitudes,

$$\langle \mathbf{n} \,|\, \mathbf{k}\rangle_E = \sum_{\Gamma i} \sum_{Lq} \langle \mathbf{n} \,|\, \Gamma i L q\rangle_E \langle \Gamma i L q \,|\, \mathbf{k}\rangle_E, \tag{2.4}$$

into multipole waves $\langle \mathbf{n} \,|\, \Gamma i L q\rangle_E$; each of these waves will in fact be constructed by working out the inverse expansion:

$$\langle \mathbf{n} \,|\, \Gamma i L q\rangle_E = \int d\hat{k}\, \langle \mathbf{n} \,|\, \mathbf{k}\rangle_E \langle \mathbf{k} \,|\, \Gamma i L q\rangle_E. \tag{2.5}$$

The integration extends over the solid angle subtended by the constant energy surface. Both the varying radius of this surface and its obliquity are taken into account in the integrand of Eq. (2.5), in particular through the factor $w_E(\mathbf{k})$ which appears in the expression (2.1) of $\langle \mathbf{n} \,|\, \mathbf{k}\rangle_E$ and will also appear in $\langle \mathbf{k} \,|\, \Gamma i L q\rangle_E$.

The role of these multipole waves in the analysis of the density of states can be described even before their actual construction. The translational invariance of the crystal ensures that the total density of states $N(E)$ is a sum of equal contribution $N(E; \mathbf{n})$ from the various cells of the crystal, where

$$N(E; \mathbf{n}) = \frac{\Omega_c}{(2\pi)^3} \int \frac{dS_E}{|\nabla_{\mathbf{k}} E(\mathbf{k})|} = \int d\hat{k}\, w_E(\mathbf{k}) = \int d\hat{k} \langle \mathbf{n} \,|\, \mathbf{k}\rangle_E \langle \mathbf{k} \,|\, \mathbf{n}\rangle_E. \tag{2.6}$$

If we substitute here the expansion (2.4) of $\langle \mathbf{n} \,|\, \mathbf{k}\rangle_E$ into multipole waves,

$$N(E; \mathbf{n}) = \sum_{\Gamma i} \sum_{Lq} \sum_{\Gamma' i'} \sum_{L' q'} \int d\hat{k} \langle \mathbf{n} \,|\, \Gamma i L q\rangle_E \langle \Gamma i L q \,|\, \mathbf{k}\rangle_E \times \langle \mathbf{k} \,|\, \Gamma' i' L' q'\rangle_E \langle \Gamma' i' L' q' \,|\, \mathbf{n}\rangle_E, \tag{2.7}$$

the unitarity of the transformation

$$\int d\hat{k} \langle \Gamma i L q \,|\, \mathbf{k}\rangle_E \langle \mathbf{k} \,|\, \Gamma' i' L' q'\rangle_E = \delta_{\Gamma\Gamma'} \delta_{ii'} \delta_{LL'} \delta_{qq'} \tag{2.8}$$

reduces $N(E; \mathbf{n})$ to the form:

$$N(E; \mathbf{n}) = \sum_{\Gamma i} \sum_{Lq} |\langle \mathbf{n} \,|\, \Gamma i L q\rangle_E|^2. \tag{2.9}$$

This expression subdivides the density of states per cell into contributions from the various multipoles. The squared magnitude of the multipole wave $\langle \mathbf{n} \,|\, \Gamma i L q\rangle_E$ thus represents the density of states belonging to the $\Gamma i L q$-multipole at the n cell. The expression (2.6) of $N(E; \mathbf{n})$ can also be interpreted as a "sum rule" for the set of multipoles waves.

The "local" density of states—as defined, e.g., by Heine and Weaire[5]—is similarly subdivided by transforming the multipole waves and (2.9) itself to the position representation by means of the Wannier functions,

$$n(E; \mathbf{r}) = \sum_{\Gamma i} \sum_{Lq} |\langle \mathbf{r} \,|\, \Gamma i L q\rangle_E|^2, \tag{2.10}$$

where

$$\langle \mathbf{r} \,|\, \Gamma i L q\rangle_E = \sum_{\mathbf{n}} \langle \mathbf{r} \,|\, \mathbf{n}\rangle \langle \mathbf{n} \,|\, \Gamma i L q\rangle_E. \tag{2.11}$$

3. CONSTRUCTION OF THE TRANSFORMATION $\langle k \,|\, \Gamma i L q\rangle_E$

The set of Bloch waves with energy E reduces to plane waves with the same energy when the crystal becomes isotropic (empty lattice). In this limit the constant energy surface is a sphere of radius $|\mathbf{k}| = \sqrt{2mE}/\hbar$ whose symmetry group is the full rotation group so that the indices Γ and i coincide with the indices L and m of the angular momentum theory. The transformation $\langle \mathbf{k} \,|\, \Gamma i L q\rangle_E$ consists then simply of the spherical harmonics $Y_{Lm}(\hat{k})$ which indeed form a complete set of orthonormal functions with weight factor 1 over a sphere of arbitrary radius. What we want now is to construct a generalization of the set of spherical harmonics for a nonspherical surface with the symmetry discussed above. The functions of the new set must, however, depend both on the direction and the magnitude of k because this magnitude varies over the surface.

(a) To allow for the variation of $|\mathbf{k}|$ explicitly, we rewrite the condition of unitarity, Eq. (2.8), by extending the integration formally over the whole Brillouin zone and then restricting it to the constant energy surface by insertion of a factor $\delta[E - E(\mathbf{k})]$ which represents the surface equation $E(\mathbf{k}) = E$. Indicating the analogs of the spherical harmonics by $P_{Lq}^{(\Gamma i)}(\mathbf{k})$, we write then the orthonormality condition in the form

$$[\Omega_c/(2\pi)^3] \int d\mathbf{k}\, P_{Lq}^{(\Gamma i)}(\mathbf{k})\, P_{L'q'}^{(\Gamma' i')}(\mathbf{k})\, \delta[E - E(\mathbf{k})] = \int d\hat{k}\, P_{Lq}^{(\Gamma i)}(\mathbf{k})\, w_E(\mathbf{k})\, P_{L'q'}^{(\Gamma' i')}(\mathbf{k}) = \delta_{\Gamma\Gamma'} \delta_{ii'} \delta_{LL'} \delta_{qq'}. \tag{3.1}$$

Comparison with Eq. (2.8) shows that the nonspherical shape of the surface requires the unitary transformation to include a weight factor $w_E^{1/2}(\mathbf{k})$:

$$\langle \mathbf{k} \,|\Gamma i L q\rangle_E = w_E^{1/2}(\mathbf{k})\, P_{Lq}^{(\Gamma i)}(\mathbf{k}), \tag{3.2}$$

while the remaining factor $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ may reduce to a polynomial as it does for a spherical surface.

(b) Following a general procedure of mathematical physics, we determine the $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ as *orthogonal polynomials belonging to the weight function* $w_E(\mathbf{k})$ which is positive over the constant energy surface. The procedure starts from *any* convenient set of linearly independent homogeneous polynomials in the components of $\mathbf{k}$. In our case we want the polynomials to be symmetry-adapted, that is, each of them should transform according to a row i of an irreducible representation Γ of the point group. We make explicit the symmetry of the polynomial by factoring out its angular part as a symmetry-adapted spherical harmonic $X_{\Gamma i\lambda}^{(l)}(\hat{k})$, which is the λth linear combination of spherical harmonics of degree l that transforms according to the Γi irreducible representation of the point group (see, e.g., Bradley and Cracknell 1972 [6]). Symmetry-adapted homogeneous polynomials in $\mathbf{k}$ are then obtained by multiplying the $X_{\Gamma i\lambda}^{(l)}(\hat{k})$ with the invariant quantity $|\mathbf{k}|^l$ and any additional power of $|\mathbf{k}|^2$, which is itself an invariant polynomial in k_x, k_y, k_z.

Proceeding now to regroup *all* the polynomials which are homogeneous of degree l in $\mathbf{k}$, we identify $\frac{1}{2}(l+1)\,(l+2)$ symmetry-adapted polynomials for each l which we write as:

$$\begin{aligned} &v_{ls\lambda}^{(\Gamma i)}(\mathbf{k}) = A(l,s)\,|\mathbf{k}|^l\, X_{\Gamma i\lambda}^{(l-2s)}(\hat{k}), \\ &A(l,s) = \{4\pi/[2(l-s)+1]!!\,(s!)2^s\}^{1/2}, \end{aligned} \tag{3.3}$$

where $s = 0, 1, \ldots, [\frac{1}{2}l]$ (integer part of $\frac{1}{2}l$). The coefficients $A(l,s)$ are so chosen that the expansion of the plane wave $\exp(i\mathbf{r}\cdot\mathbf{k})$ has the form

$$\exp(i\mathbf{r}\cdot\mathbf{k}) = \sum_l i^l l!^{-1}(\mathbf{r}\cdot\mathbf{k})^l = \sum_{\Gamma i}\sum_{ls\lambda} i^l v_{ls\lambda}^{(\Gamma i)}(\mathbf{r}) v_{ls\lambda}^{(\Gamma i)}(\mathbf{k}), \tag{3.4}$$

with each term factored into identical polynomials in $\mathbf{r}$ and $\mathbf{k}$. The series (3.4) represents a rearrangement of the usual expansion of a plane wave into spherical waves, designed for easy adaptation to nonspherical symmetries. Successive terms of the power expansion of each spherical Bessel function $j_{l'}$ have been incorporated, in Eq. (3.4), into the various polynomials $v_{ls\lambda}^{(\Gamma i)}$ with equal $l' = l - 2s$ and different l. This parcelling out of the Bessel series was made necessary by the fact that each factor $|\mathbf{k}|^2$ is no longer independent of $\hat{k}$ in a crystal.

It is emphasized that for nonspherical surfaces over which $|\mathbf{k}|$ is not constant there are

$$\sum_{s=0}^{[l/2]} [2(l-2s)+1] = \tfrac{1}{2}(l+1)\,(l+2) \tag{3.5}$$

linearly independent polynomials $v_{ls\lambda}^{(\Gamma i)}$ homogeneous in $\mathbf{k}$ with degree l, in contrast to the familiar number $2l+1$ of harmonic polynomials $|\mathbf{k}|^l Y_{lm}(\hat{k})$ for the case of a spherical surface. In the spherical case the orthogonalization problem requires no special attention because different l values correspond to different group representations.

(c) The determination of the orthogonal polynomials $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ is worked out separately for each Γ and for each i because both the constant energy surface and the weight function $w_E(\mathbf{k})$ are group-invariant. Moreover, each of the irreducible representations Γ is even (+) or odd (−) under inversion since the groups we deal with include this operation. The degree l of the symmetry-adapted polynomials $v_{ls\lambda}^{(\Gamma i)}(\mathbf{k})$ is similarly even or odd for each Γ.

The orthogonalization of the $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ will be carried out, as usual, by recursion, starting with the lowest degree, l_0, that occurs for the given Γ in the process of symmetry-adapting the spherical harmonics. This value is given in the compatibility tables for the representations of the full orthogonal group and the representations of the point group (e.g., Bradley and Cracknell [6]). Generally, there is a single symmetry-adapted polynomial of degree l_0, $v_{l_0 0}^{(\Gamma i)}(\mathbf{k})$, for each Γ and i, and we can set

$$P_{l_0}^{(\Gamma i)}(\mathbf{k}) = d^{(l_0)} v_{l_0 0}^{(\Gamma i)}(\mathbf{k}), \tag{3.6}$$

where the indices q and λ, which distinguish polynomials of the same degree, have been dropped as superfluous in this particular case. (For the cases when the irreducible representation Γ occurs more than once at l_0, orthogonalization of the corresponding $P_{l_0 q}^{(\Gamma i)}(\mathbf{k})$ should present no difficulty.) The normalization coefficient $d^{(l_0)}$ is determined by (3.1) as

$$d^{(l_0)} = \langle v_{l_0 0} \,|W_E(\Gamma i)\,|\, v_{l_0 0}\rangle^{-1/2}; \tag{3.7}$$

the matrix elements

$$\begin{aligned} &\langle v_{ls\lambda}\,|W_E(\Gamma i)\,|\,v_{l's'\lambda'}\rangle \\ &\quad = \int d\hat{k}\; v_{ls\lambda}^{(\Gamma i)}(\mathbf{k}) w_E(\mathbf{k}) v_{l's'\lambda'}^{(\Gamma i)}(\mathbf{k}) \end{aligned} \tag{3.8}$$

constitute the essential structural parameters for our problem.

For each $L > l_0$, we set up the orthogonalization procedure by representing each polynomial $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ as the sum of one group of terms homogeneous of degree L and of a second group of terms of lower degree designed to insure the orthogonalization to all $P_{L'q'}^{(\Gamma i)}(\mathbf{k})$ with $L' < L$. Thus we set:

$$P_{Lq}^{(\Gamma i)}(\mathbf{k}) = \sum_{s\lambda} v_{Ls\lambda}^{(\Gamma i)}(\mathbf{k})\, d_{s\lambda}^{(Lq)} + \sum_{L'q'}^{L'<L} P_{L'q'}^{(\Gamma i)}(\mathbf{k})\, g_{L'q'}^{(Lq)}, \tag{3.9}$$

where the coefficients $g_{L'q'}^{(Lq)}$ will be determined by Schmidt orthogonalization and the $d_{s\lambda}^{(Lq)}$ by a separate procedure. The orthogonality condition for $L' < L$ reads:

$$\begin{aligned} &\langle P_{L'q'}\,|W_E(\Gamma i)\,|\,P_{Lq}\rangle \\ &\quad = \sum_{s\lambda} \langle P_{L'q'}\,|W_E(\Gamma i)\,|\,v_{Ls\lambda}\rangle d_{s\lambda}^{(Lq)} \\ &\qquad + \sum_{L''q''}^{L''<L} \langle P_{L'q'}\,|W_E(\Gamma i)\,|\,P_{L''q''}\rangle g_{L''q''}^{(Lq)} = 0. \end{aligned} \tag{3.10}$$

Owing to the previous orthonormalization of the $P_{L'q'}^{(\Gamma i)}(\mathbf{k})$ with $L' < L$, this condition gives simply

$$g_{L'q'}^{(Lq)} = -\sum_{s\lambda} \langle P_{L'q'}\,|W_E(\Gamma i)\,|\,v_{Ls\lambda}\rangle d_{s\lambda}^{(Lq)}. \tag{3.11}$$

Substitution of this result in (3.9) reduces the orthonormalization condition on the $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ with equal degree L to the form:

$$\langle P_{Lq_1}|W_E(\Gamma i)|P_{Lq_2}\rangle = \sum_{s\lambda}\sum_{s'\lambda'} d_{s\lambda}^{(Lq_1)}\Big\{\langle v_{Ls\lambda}|W_E(\Gamma i)|v_{Ls'\lambda'}\rangle - \sum_{L'q'}^{L'<L}\langle v_{Ls\lambda}|W_E(\Gamma i)|P_{L'q'}\rangle \times\langle P_{L'q'}|W_E(\Gamma i)|v_{Ls'\lambda'}\rangle\Big\} d_{s'\lambda'}^{(Lq_2)} = \delta_{q_1q_2}. \tag{3.12}$$

The construction of the polynomials of each degree L thus reduces to the determination of the eigenvectors $\mathbf{d}^{(Lq)}$ of the symmetric and real matrix

$$\langle v_{Ls\lambda}|\overline{W}_E(\Gamma i)|v_{Ls'\lambda'}\rangle = \langle v_{Ls\lambda}|W_E(\Gamma i)|v_{Ls'\lambda'}\rangle - \sum_{L'q'}^{L'<L}\langle v_{Ls\lambda}|W_E(\Gamma i)|P_{L'q'}\rangle\langle P_{L'q'}|W_E(\Gamma i)|v_{Ls'\lambda'}\rangle. \tag{3.13}$$

The order of this matrix equals the number of symmetry-adapted polynomials $v_{Ls\lambda}^{(\Gamma i)}(\mathbf{k})$ of degree L. The normalization of the eigenvectors $\mathbf{d}^{(Lq)}$, implied by (3.12), coincides with that given by (3.7) for the special case $L=l_0$. Examples of the construction of the $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ are shown in the Appendix.

4. PROPERTIES OF THE MULTIPOLE WAVES

Entering the symmetry-adapted expansion of the plane wave (3.4) into the expression (2.5) of the multipole wave $\langle \mathbf{n}|\Gamma i L q\rangle_E$, yields

$$\langle \mathbf{n}|\Gamma i L q\rangle_E = \int d\hat{k}\,\exp(i\mathbf{n}\cdot\mathbf{k})w_E(\mathbf{k})P_{Lq}^{(\Gamma i)}(\mathbf{k}) = \sum_{ls\lambda} i^l v_{ls\lambda}^{(\Gamma i)}(\mathbf{n})\langle v_{ls\lambda}|W_E(\Gamma i)|P_{Lq}\rangle. \tag{4.1}$$

Because of the orthogonality of $P_{Lq}^{(\Gamma i)}(\mathbf{k})$ to the entire space of polynomials of degree lower than L, the expansion (4.1) starts with terms of degree $l=L$. This property was introduced by Fano[3] as the characteristic of the dependence of each multipole wave on the distance from the "center".

This result permits us now to specify that the first term of the series expansion of the multipole wave $\langle \mathbf{n}|\Gamma i L q\rangle_E$ is

$$i^L\sum_{s\lambda}\sum_{s'\lambda'} v_{Ls\lambda}^{(\Gamma i)}(\mathbf{n})\,\overline{W}_E(\Gamma i)|v_{Ls'\lambda'}\rangle d_{s'\lambda'}^{(Lq)}. \tag{4.2}$$

Recalling that $\mathbf{d}^{(Lq)}$ is eigenvector of the matrix $\langle v_{Ls\lambda}|\overline{W}_E(\Gamma i)|v_{Ls'\lambda'}\rangle$ corresponding to the eigenvalue $\overline{W}_{\Gamma Lq}$, we obtain that the first term of the series expansion of the density of states $|\langle \mathbf{n}|\Gamma i L q\rangle_E|^2$ belonging to the $\Gamma i L q$-multipole is of degree $2L$ and is given by

$$\Big\{\overline{W}_{\Gamma Lq}\sum_{s\lambda} v_{Ls\lambda}^{(\Gamma i)}(\mathbf{n})\,d_{s\lambda}^{(Lq)}\Big\}^2. \tag{4.3}$$

Since the eigenvectors $\mathbf{d}^{(Lq)}$ are normalized in accordance with (3.12), we conclude that (4.2) is linear, rather than quadratic, in the eigenvalue $\overline{W}_{\Gamma Lq}$ of the matrix (3.13).

This conclusion represents the central result of the paper; we have shown how the density of states around a center cell in a crystalline medium can be characterized in practice by a few parameters only of the constant energy surface, namely the eigenvalues $\overline{W}_{\Gamma Lq}$ for rather low L. More specifically, the structural similarity of the polynomials $v_{Ls\lambda}^{(\Gamma i)}(\mathbf{n})$ and of the Bessel functions $j_L(|\mathbf{n}|)$ shows that $v_{Ls\lambda}^{(\Gamma i)}(\mathbf{n})\ll 1$ whenever $|\mathbf{n}|\ll L$. However, we regard this result as the first step of a broader investigation of physical parameters appropriate to the study of phenomena with local character.

This point of view is close to that which motivated Kohn[4] and collaborators to express energy band properties of solids directly in terms of Wannier functions and of such local quantities as the matrix elements $\langle O|H_c|\mathbf{n}\rangle$, where H_c is the 1-electron Hamiltonian of the crystal. As an example of the connection of the two approaches, the equation of the constant energy surface, which is given by Kohn as a Fourier series with coefficients $\langle O|H_c|\mathbf{n}\rangle$, can be expressed in terms of the symmetry-adapted polynomials belonging to Γ_1^+:

$$E(\mathbf{k}) = \sum_{ls\lambda} E_{ls\lambda} v_{ls\lambda}^{(\Gamma_1^+)}(\mathbf{k}), \tag{4.3}$$

where

$$E_{ls\lambda} = i^l\sum_{\mathbf{n}}\langle O|H_c|\mathbf{n}\rangle v_{ls\lambda}^{(\Gamma_1^+)}(\mathbf{n}). \tag{4.4}$$

Owing to the localization of the Wannier functions, the sum over $\mathbf{n}$ extends in effect only to a limited number of cells around the center. Owing, once again, to the properties of the polynomials $v_{ls\lambda}$, the sum in (4.3) also converges rapidly with increasing l.

APPENDIX: EXAMPLES OF ORTHOGONAL POLYNOMIALS FOR THE GROUP O_h

We use the notation of Bradley and Cracknell[6]:

$$Y_l^{m,c} = (2)^{-1/2}(Y_l^m + Y_l^{-m}),\quad Y_l^{m,s} = -i(2)^{-1/2}(Y_l^m - Y_l^{-m}).$$

Case 1: Γ_1^+

(a) $l=0$, $v_{00}(\mathbf{k})=1$, $P_0(\mathbf{k}) = \langle v_{00}|W_E|v_{00}\rangle^{-1/2}$.

(b) $l=2$, $v_{21}(\mathbf{k}) = (6)^{-1/2}|\mathbf{k}|^2$,

$$P_2(\mathbf{k}) = \langle v_{21}|\overline{W}_E|v_{21}\rangle^{-1/2}[v_{21}(\mathbf{k}) - P_0(\mathbf{k})\langle P_0|W_E|v_{21}\rangle].$$

(c) $l=4$

$$v_{40}(\mathbf{k}) = (4\pi/945)^{1/2}|\mathbf{k}|^4[(\tfrac{7}{12})^{1/2} Y_4^0(\hat{k}) + (\tfrac{5}{12})^{1/2} Y_4^{4,c}(\hat{k})],$$

$$v_{42}(\mathbf{k}) = (120)^{-1/2}|\mathbf{k}|^4,$$

$$P_{4\pm}(\mathbf{k}) = \overline{W}_{4\pm}^{-1/2}[(\overline{W}_{4\pm} - \langle v_{40}|\overline{W}_E|v_{40}\rangle)^2 + \langle v_{40}|\overline{W}_E|v_{42}\rangle^2]^{-1/2}[\langle v_{40}|\overline{W}_E|v_{42}\rangle u_{40}(\mathbf{k}) + (\overline{W}_{4\pm} - \langle v_{40}|\overline{W}_E|v_{40}\rangle)u_{42}(\mathbf{k})],$$

where

$$u_{4s}(\mathbf{k}) = v_{4s}(\mathbf{k}) - P_2(\mathbf{k})\langle P_2|W_E|v_{4s}\rangle$$

$$- P_0(\mathbf{k})\langle P_0|W_E|v_{4s}\rangle \quad (s=0,2),$$

and $\overline{W}_{4\pm}$ are the roots of the quadratic equation

$$(\langle v_{40}|\overline{W}_E|v_{40}\rangle - \overline{W}_{4q})(\langle v_{42}|\overline{W}_E|v_{42}\rangle - \overline{W}_{4q}) = \langle v_{40}|\overline{W}_E|v_{42}\rangle^2.$$

Case 2: Γ_4^-

(a) $l=1$,
$$\left.\begin{matrix} v_{10}^{(1)}(\mathbf{k}) \\ v_{10}^{(2)}(\mathbf{k}) \\ v_{10}^{(3)}(\mathbf{k}) \end{matrix}\right\} = \left(\frac{4\pi}{3}\right)^{1/2} |\mathbf{k}| \left\{\begin{matrix} Y_1^{1,c}(\hat{k}) \\ Y_1^{1,s}(\hat{k}) \\ Y_1^{0}(\hat{k}) \end{matrix}\right. = \left\{\begin{matrix} k_x, \\ k_y, \\ k_z, \end{matrix}\right.$$

$$P_1^{(i)}(\mathbf{k}) = \langle v_{10}|W_E(i)|v_{10}\rangle^{-1/2} v_{10}^{(i)}(\mathbf{k}) \quad (i=1,2,3).$$

(b) $l=3$,
$$\left.\begin{matrix} v_{30}^{(1)}(\mathbf{k}) \\ v_{30}^{(2)}(\mathbf{k}) \\ v_{30}^{(3)}(\mathbf{k}) \end{matrix}\right\}$$

$$= \left(\frac{4\pi}{105}\right)^{1/2} |\mathbf{k}|^3 \left\{\begin{matrix} (\tfrac{3}{8})^{1/2} Y_3^{1,c}(\hat{k}) - (\tfrac{5}{8})^{1/2} Y_3^{3,c}(\hat{k}), \\ (\tfrac{3}{8})^{1/2} Y_3^{1,s}(\hat{k}) + (\tfrac{5}{8})^{1/2} Y_3^{3,s}(\hat{k}), \\ - Y_3^{0}(\hat{k}), \end{matrix}\right.$$

$$\left.\begin{matrix} v_{31}^{(1)}(\mathbf{k}) \\ v_{31}^{(2)}(\mathbf{k}) \\ v_{31}^{(3)}(\mathbf{k}) \end{matrix}\right\} = \left(\frac{4\pi}{30}\right)^{1/2} |\mathbf{k}|^3 \left\{\begin{matrix} Y_1^{1,c}(\hat{k}), \\ Y_1^{1,s}(\hat{k}), \\ Y_1^{0}(\hat{k}), \end{matrix}\right.$$

$$P_{3\pm}^{(i)}(\mathbf{k}) = \overline{W}_{3\pm}^{-1/2}[(\overline{W}_{3\pm} - \langle v_{30}|\overline{W}_E(i)|v_{30}\rangle)^2 + \langle v_{30}|\overline{W}_E(i)|v_{31}\rangle^2]^{-1/2} \times[\langle v_{30}|\overline{W}_E(i)|v_{31}\rangle u_{30}^{(i)}(\mathbf{k}) + (\overline{W}_{3\pm} - \langle v_{30}|\overline{W}_E(i)|v_{30}\rangle) u_{31}^{(i)}(\mathbf{k})] \quad (i=1,2,3),$$

where

$$u_{3s}^{(i)}(\mathbf{k}) = v_{3s}^{(i)}(\mathbf{k}) - P_1^{(i)}(\mathbf{k})\langle P_1|W_E(i)|v_{3s}\rangle \quad (s=0,1),$$

and $\overline{W}_{3\pm}$ are the roots of the quadratic equation

$$(\langle v_{30}|\overline{W}_E(i)|v_{30}\rangle - \overline{W}_{3q})(\langle v_{31}|\overline{W}_E(i)|v_{31}\rangle - \overline{W}_{3q}) = \langle v_{30}|W_E(i)|v_{31}\rangle^2.$$

Case 3: Γ_3^+

(a) $l=2$,
$$\left.\begin{matrix} v_{20}^{(1)}(\mathbf{k}) \\ v_{20}^{(2)}(\mathbf{k}) \end{matrix}\right\} = \left(\frac{4\pi}{15}\right)^{1/2} |\mathbf{k}|^2 \left\{\begin{matrix} Y_2^{0}(\hat{k}), \\ Y_2^{2,c}(\hat{k}), \end{matrix}\right.$$

$$P_2^{(i)}(\mathbf{k}) = \langle v_{20}|W_E(i)|v_{20}\rangle^{-1/2} v_{20}^{(i)}(\mathbf{k}) \quad (i=1,2).$$

(b) $l=4$
$$\left.\begin{matrix} v_{40}^{(1)}(\mathbf{k}) \\ v_{40}^{(2)}(\mathbf{k}) \end{matrix}\right\}$$

$$= \left(\frac{4\pi}{945}\right)^{1/2} |\mathbf{k}|^4 \left\{\begin{matrix} (\tfrac{5}{12})^{1/2} Y_4^{0}(\hat{k}) - (\tfrac{7}{12})^{1/2} Y_4^{4,c}(\hat{k}), \\ - Y_4^{2,c}(\hat{k}), \end{matrix}\right.$$

$$\left.\begin{matrix} v_{41}^{(1)}(\mathbf{k}) \\ v_{41}^{(2)}(\mathbf{k}) \end{matrix}\right\} = \left(\frac{4\pi}{210}\right)^{1/2} |\mathbf{k}|^4 \left\{\begin{matrix} Y_2^{0}(\hat{k}), \\ Y_2^{2,c}(\hat{k}), \end{matrix}\right.$$

$$P_{4\pm}^{(i)}(\mathbf{k}) = \overline{W}_{4\pm}^{-1/2}[(\overline{W}_{4\pm} - \langle v_{40}|\overline{W}_E(i)|v_{40}\rangle)^2 + \langle v_{40}|\overline{W}_E(i)|v_{41}\rangle^2]^{-1/2} \times[\langle v_{40}|\overline{W}_E(i)|v_{41}\rangle u_{40}^{(i)}(\mathbf{k}) + (\overline{W}_{4\pm} - \langle v_{40}|\overline{W}_E(i)|v_{40}\rangle) u_{41}^{(i)}(\mathbf{k})] \quad (i=1,2),$$

where

$$u_{4s}^{(i)}(\mathbf{k}) = v_{4s}^{(i)}(\mathbf{k}) - P_2^{(i)}(\mathbf{k})\langle P_2|W_E(i)|v_{4s}\rangle \quad (s=0,1),$$

and $\overline{W}_{4\pm}$ are the roots of the quadratic equation

$$(\langle v_{40}|\overline{W}_E(i)|v_{40}\rangle - \overline{W}_{4q})(\langle v_{41}|W_E(i)|v_{41}\rangle - \overline{W}_{4q}) = \langle v_{40}|\overline{W}_E(i)|v_{41}\rangle^2.$$

*Work supported by the U.S. Energy Research and Development Administration, Contract No. C00-1674-108.

[1]U. Fano in *Vacuum Ultraviolet Radiation Physics*, edited by E.E. Koch, R. Haensel, and C. Kunz (Vieweg, Braunschweig, 1974), p. 84.
[2]T.M. Hayes and P.N. Sen, Phys. Rev. Lett. 34, 956 (1975).
[3]U. Fano, Phys. Rev. Lett. 31, 234 (1975).
[4]W. Kohn, Phys. Rev. B 7, 4388 (1973).
[5]V. Heine and D. Weaire, *Solid State Phys.* 24 (Academic, New York, 1970), pp. 249–463.
[6]C.J. Bradley and A.P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).

Duality transformation in a many-component spin model

F. Y. Wu* and Y. K. Wang

Department of Physics, Northeastern University, Boston, Massachusetts 02115
(Received 6 October 1975)

It is shown that the duality transformation relates a spin model to its dual whose Boltzmann factors are the eigenvalues of the matrix formed by the Boltzmann factors of the original spin model. The duality relation valid for finite lattices is obtained, and applications are given.

The duality relation for two-dimensional spin models can be considered both from a topological and an algebraic point of view. A comprehensive discussion of these aspects for the Potts and the Ashkin-Teller (AT) models has been given by Mittag and Stephen.[1] More recently, Wegner[2] has reformulated the duality relation as an instance of a more general transformation. In this note we point out one further aspect of the duality transformations. Our result helps to clarify the reasoning in Wegner's formulation and also provides straightforward extensions of duality to other spin models.

Consider a q component spin model on a two-dimensional lattice L which has N sites. Let $\xi_i = 1, 2, \ldots, q$ denote the spin state of the ith site. The Hamiltonian can be generally written as

$$\mathcal{H} = -\sum_{\langle ij \rangle} J(\xi_i, \xi_j), \tag{1}$$

where $-J(\xi, \xi')$ is the interaction between the spin states ξ and ξ'. The summation in (1) is over all interacting pairs $\langle ij \rangle$ which we assume to be noncrossing. The partition function is

$$Z = \sum_{\xi_i=1}^{q} \prod_{\langle ij \rangle} u(\xi_i, \xi_j), \tag{2}$$

with

$$u(\xi, \xi') = \exp[J(\xi, \xi')/kT]. \tag{3}$$

We shall restrict our attention to the case that

$$u(\xi_i, \xi_j) = u(\xi_i - \xi_j), \quad (\mathrm{mod}\ q). \tag{4}$$

Thus the matrix U whose elements are $u(\xi, \xi')$ is cyclic. It is not necessary for our discussion to further assume that U is symmetric, although in most applications this will be the case. In order to distinguish ξ_i from ξ_j for a given edge connecting sites i and j, we place an arrow on it pointing from i to j. Thus the lattice is directed. We shall also have occasion to consider the situation, such as for the AT model, that U is block-cyclic. These cases will be explored in later discussions.

We can rewrite the partition function in two different ways. First, instead of specifying the spin states by ξ_i, we may label the edge in (4) by the difference $\xi_{ij} \equiv \xi_i - \xi_j$. However, to ensure that each set of ξ_{ij} will correspond to some spin states, it is necessary (and sufficient) to require

$$\sum_{\mathrm{cw}} \xi_{ij} = \sum_{\mathrm{ccw}} \xi_{ij} \tag{5}$$

around each face of L. Here the summation cw (ccw) is over the edges carrying clockwise (counterclockwise) arrows around the face. Clearly, the $\xi_i \Leftrightarrow \xi_{ij}$ mapping is q to 1. Denoting the restriction (5) by a prime over the summation sign, we can now rewrite the partition sum (2) as

$$Z(u) = q \sum_{\xi_{ij}=1}^{q}{}' \prod_{\langle ij \rangle} u(\xi_{ij}). \tag{6}$$

To make connection with the partition function on the dual of L, or L^D, we now cast Z into another form. Direct the edges of L^D such that the arrows on L^D coincide with those on L if each edge of L^D is rotated 90° clockwise. The situation around a site on L is shown in Fig. 1. Now the eigenvalues of the $q \times q$ cyclic matrix U are

$$\lambda(\eta) = \sum_{\xi=1}^{q} \exp(2\pi i \xi \eta/q)\, u(\xi), \quad \eta = 1, \ldots, q, \tag{7}$$

or, conversely,

$$u(\xi_{ij}) = \sum_{\eta=1}^{q} T(\xi_i, \eta)\, \lambda(\eta)\, T^*(\xi_j, \eta), \tag{8}$$

where

$$T(\xi, \eta) = q^{-1/2} \exp(2\pi i \xi \eta/q). \tag{9}$$

We substitute (8) into (2) and carry out the sums over ξ_i. At each site of L, we have for each outgoing (incoming) arrow a factor $T(\xi, \eta)$ $[T^*(\xi, \eta)]$. Denote the spin states of the spin model on L^D by η_α and identify the η in (8) as $\eta_{\alpha\beta} \equiv \eta_\alpha - \eta_\beta$, where the arrow runs from site α to site β on the corresponding edge of L^D. Then the summation over ξ_i (cf. Fig. 1) leads to a factor

$$\sum_{\xi_i=1}^{q} T(\xi_i, \eta_{21})\, T^*(\xi_i, \eta_{23}) \cdots T(\xi_i, \eta_{1n_i})$$
$$= q^{1-n_i/2}\, \delta_{\mathrm{Kr}} \left(\sum_{\mathrm{cw}} \eta_{\alpha\beta} - \sum_{\mathrm{ccw}} \eta_{\alpha\beta} \right), \tag{10}$$

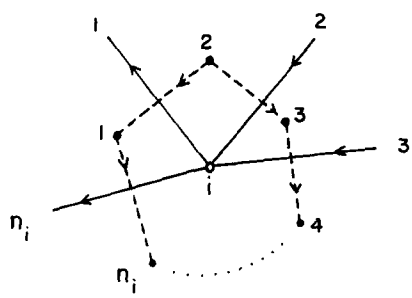


FIG. 1. The directed edges around the ith site on L. The solid (broken) lines are the edges of L (L^D).

where n_i is the number of neighbors of the ith site. The restriction imposed by the Kronecker delta on the rhs of (10) is exactly the same as in (5) for a face of L^D. Thus, after combining (8) with (2) and using (10), the partition function takes the form

$$Z = q^{N-E} \sum_{\eta_{\alpha\beta}=1}^{q}{}' \prod_{(\alpha\beta)} \lambda(\eta_{\alpha\beta}), \tag{11}$$

where E is the number of edges of L (or L^D). Finally, by comparing (11) with (6) and using the Euler's relation for a connected planar graph,[3]

$$N + N_D = E + 2, \tag{12}$$

we obtain the identity

$$Z(u) = q^{1-N_D} Z^{(D)}(\lambda). \tag{13}$$

This is our main result and it is valid for any *finite* lattice. Here $Z^{(D)}(\lambda)$ is the partition function of the spin model on L^D whose Boltzmann factors are given by (7). While this result is implicit in Ref. 2, our discussion does bring out in a natural way the role played by the U matrix, thus clarifying the reasoning behind Wegner's formulation.

An example is the Potts model[4] with

$$U = \begin{pmatrix} e^K & 1 & \cdots & 1 \\ 1 & e^K & \cdots & 1 \\ \cdots & & \cdots & \\ 1 & 1 & \cdots & e^K \end{pmatrix}. \tag{14}$$

The eigenvalues of U are

$$\lambda_1 = e^K + q - 1, \quad \lambda_2 = \cdots = \lambda_q = e^K - 1, \tag{15}$$

so that the equivalence (13) reads

$$Z(e^K) = q^{1-N_D}(e^K - 1)^E Z^{(D)}(e^{K*}), \tag{16}$$

where

$$e^{K*} = \lambda_1/\lambda_2 = (e^K + q - 1)/(e^K - 1). \tag{17}$$

The above result is readily extended to the case where U is block-cyclic. An example is the AT model for which

$$U = \begin{pmatrix} U_1 & U_2 \\ U_2 & U_1 \end{pmatrix}, \tag{18}$$

where U_1 and U_2 are themselves 2×2 cyclic matrices. Generally we consider a matrix U which is m-fold cyclic. That is to say, U is composed of q_1 cyclic matrices, each of which in turn contains q_2 cyclic matrices, etc., the dimension of U being $q = q_1 q_2 \cdots q_m$. Thus, an element of U, which specifies the spin states of the model, is described by an m component vector $\xi \equiv (\xi_1, \ldots, \xi_m)$ whose components can take on, respectively, $q_1, q_2, \ldots, q_m$ different values. Treating the previous ξ and η as vectors, we can carry through all the steps and again arrive at the equivalence (13), provided that in place of (7) we have

$$\lambda(\eta) = \sum_{\xi} \exp[2\pi i(\xi_1\eta_1/q_1 + \cdots + \xi_m\eta_m/q_m)]u(\xi). \tag{19}$$

For the AT model we have $q_1 = q_2 = 2$, $\xi_i, \eta_i = 1, 2$. Equation (19) then leads to the duality relations derived by Ashkin and Teller.[5] As a further illustration consider the six-component spin model whose U matrix is

$$U = \begin{pmatrix} U_1 & U_2 & U_2 \\ U_2 & U_1 & U_2 \\ U_2 & U_2 & U_1 \end{pmatrix}, \tag{20}$$

where $U_1 = \left(\begin{smallmatrix} a & b \\ b & a \end{smallmatrix}\right)$ and $U_2 = \left(\begin{smallmatrix} \alpha & \beta \\ \beta & \alpha \end{smallmatrix}\right)$ are 2×2 matrices. It is easily seen that the eigenvalues of U form a similar cyclic matrix whose elements are

$$\begin{aligned} a^* &= \lambda_1 = a + b + 2(\alpha + \beta), \\ b^* &= \lambda_2 = a - b + 2(\alpha - \beta), \\ \alpha^* &= \lambda_3 = \lambda_4 = a + b - (\alpha + \beta), \\ \beta^* &= \lambda_5 = \lambda_6 = a - b - (\alpha - \beta). \end{aligned} \tag{21}$$

This is the duality transformation.

Note added in proof: Finally we remark that our result (13) is valid even if the Boltzmann factor (3) is edge-dependent. In this case the eigenvalues (7) or (19) are introduced for each edge ij and in (13) we have $u = \{u_{ij}\}$, $\lambda = \{\lambda_{ij}\}$.

*Supported in part by National Science Foundation Grant No. DMR72-03213A01.

[1]L. Mittag and J. Stephen, J. Math. Phys. **12**, 441 (1971).
[2]F.J. Wegner, Physica **68**, 570 (1973).
[3]We have used here $N_D = S + 1$, where S is the number of independent circuits in the graph.
[4]R. B. Rotts, Proc. Cambridge Philos. Soc. **48**, 106 (1952).
[5]J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).

Erratum: Formula for the computation of the representation matrix elements of the group $SO(n)$ [J. Math. Phys. 16, 334 (1975)]

Takayoshi Maekawa

Department of Physics, Kumamoto University, Kumamoto, Japan
(Received 7 October 1975)

(1) On p. 335, (2.7) should be changed as follows:

$$D^{(n)}(\theta_{21}, \cdots, \theta_{n\,n-1}) = D^{(n-1)}\left[\prod_{k=n}^{4} R_{k\,k-1}(\theta_{n\,n-k+1})\right] R_{31}(\theta_{n\,n-2}) R_{12}(\theta_{n\,n-1}) = \prod_{j=n-2}^{0}\left[\left(\prod_{k=n-j}^{4} R_{k\,k-1}(\theta_{n-j\,n-j-k+1})\right) R_{31}(\theta_{n-j\,n-j-2}) \times R_{12}(\theta_{n-j\,n-j-1})\right].$$

(2) On p. 337, the last condition of (4.2) should be changed to inequality:

$$m_{2j+1,j} \geq |m_{2j,j}|.$$

(3) On p. 338, the (+) sign of the second terms on the right-hand side of (4.6) and (4.7) should be changed to a (−) sign.

(4) On p. 338, (4.9) should be changed as follows

$$D^{(n)} = R_{12}(\theta_{21}) R_{31}(\theta_{31})\left(\prod_{k=4}^{n} R_{k\,k-1}(\theta_{k1})\right) D^{(n-1)'}.$$

(5) On p. 339, the (+) sign of the second term on the right-hand side of (4.16) should be changed to a (−) sign.

Erratum: On the stationary gravitational fields [J. Math. Phys. 15, 1096 (1974)]

S. Kloster, M. M. Som, and A. Das

Department of Mathematics, Simon Fraser University, Burnaby 2, B.C., Canada
(Received 10 October 1975)

(1) The last sentence of the abstract [and the one before Eq. (7.14)] should say that the class of metrics found is mostly outside the P.E. class.

(2) In the fifth paragraph of the introduction, C_ω should be replaced by C^ω.

(3) In the sixth line of Eq. (2.6), a^β should be replaced by a_β, and in the seventh line, 2 should be replaced by 1.

(4) The first equation in (F_1) should assert that $\sigma_{\mu\nu} = 0$. The zero was omitted.

(5) The third term of the integrand in (3.2) should be $-\frac{1}{4}\exp(2\omega) f^{\alpha\beta} f_{\alpha\beta}$.

(6) In the second line of Eq. (6.2), the factor $(d\theta^2 + r^2 \sin^2\theta\, d\phi^2)$ should be replaced by $r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$.

(7) In Eq. (6.4), the signs should be as follows:

$$\Phi = +(2k)^{-1}[(1 + m/2R)^4 + (1 - m/2R)^4][dR^2 - R^2(d\psi^2 + \sinh^2\psi\, d\phi^2)] - 2k(1 - m^2/4R^2)^2[(1 + m/2R)^4 + (1 - m/2R)^4]^{-1}[(2m/k)\cosh\psi\, d\phi + dt]^2.$$

(8) The existence of conformastationary solutions outside of the P.E. class is still an open question. The coordinate conditions used form an overdetermined system.

(9) In Eq. (7.1) the second dz^2 should be $d\theta^2$.

(10) To be more precise, Eq. (7.13) should be written

$$g(\zeta) = f(\zeta, \bar{\zeta}) + \frac{1}{\pi} \times \iint \frac{(\omega_{,\bar{\zeta}'} - 1/4\rho') f(\rho', z') + (\omega_{,\zeta'} - 1/4\rho') \bar{f}(\rho', z')}{\zeta' - \zeta}\, d\rho'\, dz'.$$

(11) M.M. Som's address is: Centro Brasileiro de Pesquisas Fisicas, A.V. Wenceslau, Braz, 71-Zc-82, Rio de Janeiro-GB, Brasil.

