JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 2 February 1964

Group-Theoretical Models of Local-Field Theories*

HANS JOOS†
Institute for Advanced Study, Princeton, New Jersey
(Received 7 August 1963)

Local fields on a finite, discrete, space–time model are introduced as a guide for axiomatic discussion of quantum field theory.

1. INTRODUCTION

THE axiomatic discussion of relativistic quantum field theory[1] has not yet been very successful in clarifying the content and the interdependence of the main principles of this theory. Therefore we follow a mathematical tradition[2] when we study in this paper mathematical structures which are similar to the scheme of formal quantum field theory.

Let us first generalize the notion of a relativistic local quantum field. Consider a transitive transformation group **C** of a set **M** with a subgroup $\mathbf{R} \subset \mathbf{C}$, and a subset **S** of the set of all irreducible, unitary representations of **C**; then we call $A(x)$ a covariant local operator field in a Hilbert space **H**, if it satisfies the following conditions:

(i) $x \to A(x)$ is a mapping of **M** into the linear operators of **H**.[3]

(ii) There is an unitary representation $U(\mathbf{C})$ of **C** by operators $U(c)$ of **H** which satisfies

$$U(c)A(x)U^{-1}(c) = A(cx),$$

for all $x \in \mathbf{M}$ and $c \in \mathbf{C}$.

(iii) (a) There exists one and only one independent vector $\Omega \in \mathbf{H}$ which is invariant under all $U(c)$: $U(c)\Omega = \Omega$ for all $c \in \mathbf{C}$.

(b) The multiplicity in $U(\mathbf{C})$ of irreducible representations which are not elements of **S** is zero.

(iv) $[A(x), A(x')]_- = 0$ if $x = rx'$, $r \in \bigcup_c c\mathbf{R}c^{-1}$, $c \in \mathbf{C}$. [or (iv′) $[A(x), A(x')]_+ = 0$ if $x = rx'$, $r \neq 1$, and $r \in \bigcup_c c\mathbf{R}c^{-1}$].

These are the usual "axioms" of a scalar field, if **M** is the set of the 4-dimensional space–time points x, if the group **C** consists of the inhomogeneous orthochronous Lorentz transformations of x with the subgroup **R** of 3-dimensional space translations, and if the elements of **S** are the irreducible unitary representations of **C** with 4-momenta P_μ in the forward cone: $P_\mu P^\mu \geq 0$, $P_0 > 0$.

We consider in the following covariant, local operator fields of certain finite groups **C** which have a structure similar to the 2-dimensional inhomogeneous Lorentz group. These groups are

* This research was supported wholly or in part by the United States Air Force under grant No. AF-AFOSR-61-19, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

† Present address: Deutsches Elektronen Synchrotron Desy, Hamburg, Germany.

[1] A. S. Wightman, "Problèmes mathématiques de la théorie quantique relativiste," in Colloque sur les problèmes mathématiques des champs, Lille, 1957; K. Symanzik, "Grundlagen und gegenwärtiger Stand der feldgleichungsfreien Feldtheorie" in *Werner Heisenberg und die Physik unserer Zeit* (Friedr. Vieweg und Sohn, Braunschweig, Germany, 1961).

[2] The following shows that we are influenced by the axiomatic discussion of finite geometries, as described by E. Artin in *Geometric Algebra* (Interscience Publishers, Inc., New York, 1957). M. Hall, *The Theory of Groups* (The Macmillan Company, New York, 1959), p. 346. G. Pickert, "Projektive Ebenen" in *Die Grundlehren der mathematischen Wissenschaften*, (Springer-Verlag, Berlin, 1955).

[3] $A(x)$ are, in general, operator distributions. This fact is not important for our discussion.

the inhomogeneous Lorentz groups over certain prime fields of a characteristic $q \neq 0$.[4] In choosing finite groups, we avoid topological difficulties and gain advantages in the calculation of examples. Further, we believe that we can get an insight in the algebraic structure of such fields, which is valuable for the discussion of physical quantum fields. In the following section, we describe the "geometry" connected with Lorentz groups over finite prime fields. Section 3 contains the discussion of the unitary representations of these groups. In Sec. 4 we formulate field theory for such space–time models. In Sec. 5 we relate the construction of local operator fields to the general theory of group representations. Finally we discuss examples in Sec. 6.

2. THE MODEL OF SPACE AND TIME

Let $\mathbf{M} = \{x\}$ be the 2-dimensional vector space over the prime field $GF(q)$ of characteristic $q \neq 0, 2$, i.e., $x = (x_0, x_1)$, with x_0, x_1 residue classes mod q. Then the group ${}^q\mathbf{L}$ of nonsingular linear transformations Λ of $\mathbf{M}$, which leave the form

$$xy = x_0y_0 - x_1y_1 = \tfrac{1}{2}(x_+y_- + x_-y_+)$$
$$x, y \in \mathbf{M}, \quad x_+ = x_0 + x_1, \quad x_- = x_0 - x_1, \text{ etc.} \tag{2.1}$$

invariant, is called the 2-dimensional homogeneous Lorentz group over $GF(q)$. From the invariance of x_+x_-, we see that $\Lambda \in {}^q\mathbf{L}$ has the form $\Lambda = \Lambda(\lambda)$ or $\Lambda = \Lambda(\lambda)\Pi$, with

$x' = \Lambda(\lambda)x$:

$$x'_+ = \lambda x_+, \quad x'_- = \lambda^{-1}x_-, \quad \lambda \in GF(q), \quad \lambda \neq 0;$$
$$x' = \Pi x: \qquad x'_+ = x_-, \qquad x'_- = x_+. \tag{2.2}$$

As the multiplicative group of the $\lambda \in GF(q)$, $\lambda \neq 0$, is cyclic of order $q - 1$, the formulas

$$\Lambda(\lambda)\Lambda(\lambda') = \Lambda(\lambda\lambda'),$$
$$\Pi\Lambda(\lambda) = \Lambda(\lambda^{-1})\Pi, \qquad \Pi^2 = 1 \tag{2.3}$$

demonstrate that ${}^q\mathbf{L}$ is isomorphic with the dihedral group of order $2(q - 1)$.

We consider some subgroups of ${}^q\mathbf{L}$. The *proper homogeneous Lorentz group* ${}^q\mathbf{L}_+$ consists of the λ with $\det\Lambda = 1$. It follows from (2.2) that the elements of ${}^q\mathbf{L}_+$ are the $\Lambda(\lambda)$. The group ${}^q\mathbf{L}_+$ is isomorphic to the multiplicative group of $GF(q)$. Now we determine the *commutator subgroup* ${}^q\mathbf{L}'$ of ${}^q\mathbf{L}$. With the help of (2.3) we calculate the commutators $\Lambda_1\Lambda_2\Lambda_1^{-1}\Lambda_2^{-1}$:

$$\Lambda_1\Lambda_2\Lambda_1^{-1}\Lambda_2^{-1} = (1; \Lambda^2(\lambda_1); \Lambda^2(\lambda_2^{-1}); \Lambda^2(\lambda_1\lambda_2^{-1})),$$

for

$$\Lambda_1 = (\Lambda(\lambda_1); \Lambda(\lambda_1); \Lambda(\lambda_1)\Pi; \Lambda(\lambda_1)\Pi),$$
$$\Lambda_2 = (\Lambda(\lambda_2); \Lambda(\lambda_2)\Pi; \Lambda(\lambda_2); \Lambda(\lambda_2)\Pi). \tag{2.4}$$

Therefore the elements of ${}^q\mathbf{L}'$ are the squares of ${}^q\mathbf{L}$, i.e., ${}^q\mathbf{L}' = \{\Lambda(\lambda^2)\}$. We call $\lambda \in GF(q)$ a square (or nonsquare) if there exists (or does not exist) a $\rho \neq 0$, $\rho \in GF(q)$, and $\lambda = \rho^2$. For notation we introduce the symbol[5] $\epsilon(\lambda, q)$:

$$\epsilon(\lambda, q) = 1; 0; -1$$

for

$$\lambda = \rho^2; \qquad \lambda = 0; \qquad \lambda \neq \rho^2. \tag{2.5}$$

These equations follow immediately:

$$\epsilon(\lambda_1\lambda_2, q) = \epsilon(\lambda_1, q)\epsilon(\lambda_2, q),$$
$$\epsilon(\lambda^{-1}, q) = \epsilon(\lambda, q). \tag{2.6}$$

The group ${}^q\mathbf{L}'$ is isomorphic to the multiplicative group of the squares in $GF(q)$ and has the order $[{}^q\mathbf{L}' : 1] = \frac{1}{2}(q - 1)$. Equations (2.2), (2.3), and (2.4) are also valid for the 2-dimensional Lorentz group over the field of the real numbers ${}^0\mathbf{L}$. The commutator subgroup of ${}^0\mathbf{L}$ is the group of the proper, orthochronous Lorentz transformations. The *orthochronous homogeneous Lorentz group* ${}^q\mathbf{L}^\uparrow$ is the subgroup generated by ${}^q\mathbf{L}'$ and Π. The *center* $\mathbf{Z}$ of ${}^q\mathbf{L}$ consists of the Lorentz transformations which commute with all $\Lambda \in {}^q\mathbf{L}$. It follows from (2.3) that for $q \neq 3$ the centrale lements are $\Lambda(1)$ and $\Lambda(-1) = \mathrm{I}$.[6]

The relations between the subgroups ${}^q\mathbf{L}_+$, ${}^q\mathbf{L}^\uparrow$, ${}^q\mathbf{L}'$, and $\mathbf{Z}$ are somewhat different for $\frac{1}{2}(q - 1)$ even or $\frac{1}{2}(q - 1)$ odd :

(a) $\frac{1}{2}(q - 1)$ odd: In this case ${}^q\mathbf{L}'$ has no subgroup of order 2, therefore $\mathbf{Z} \not\subset {}^q\mathbf{L}'$ and $\epsilon(-1, q) = -1$. From the description of the subgroups of ${}^q\mathbf{L}$ above, we see that ${}^q\mathbf{L}'$, ${}^q\mathbf{L}'$, ${}^q\mathbf{L}_+$, ${}^q\mathbf{L}^\uparrow$ are normal subgroups of index 2 of ${}^q\mathbf{L}_+$, ${}^q\mathbf{L}^\uparrow$, ${}^q\mathbf{L}$, ${}^q\mathbf{L}$, respectively. The groups ${}^q\mathbf{L}_+$, ${}^q\mathbf{L}^\uparrow$, ${}^q\mathbf{L}$, ${}^q\mathbf{L}$ are generated by the elements of ${}^q\mathbf{L}'$, ${}^q\mathbf{L}'$, ${}^q\mathbf{L}_+$, ${}^q\mathbf{L}^\uparrow$, respectively, together with the corresponding reflections I, Π, Π, I. We have, therefore, in this case the same situation, with respect to the subgroups $\mathbf{L}'$, L_+, $\mathbf{L}^\uparrow$ and the reflections Π, I and $\mathrm{T} = \mathrm{I}\Pi$, as in the physical case $q = 0$.

(b) $\frac{1}{2}(q - 1)$ even: There are two essential differences compared with case (a). Now we have $\epsilon(-1, q) = 1$, and therefore $\mathbf{Z} \subset {}^q\mathbf{L}'$. While, as in (a), ${}^q\mathbf{L}'$ and ${}^q\mathbf{L}^\uparrow$ are normal subgroups of index 2 of ${}^q\mathbf{L}_+$ (or ${}^q\mathbf{L}$), there are no reflections, i.e., no

[4] For this notion, compare J. Dieudonné, "Sur les groupes classiques" in *Actualités scientifique et industrielles* (Hermann, et Cie., Paris, 1958) The author is indebted to Dr. A. Grossmann for bringing this paper to his attention.

[5] $\epsilon(\lambda, q)$ is essentially the Legendre symbol (λ/q). According to Euler's criterion we have $(\lambda/q) = \lambda^{\frac{1}{2}(q-1)} \bmod q$. See H. Hasse, "Vorlesungen über Zahlentheorie" in *Die Grundlagen der mathematischen Wissenschaften* (Springer-Verlag, Berlin, 1950).

[6] ${}^3\mathbf{L}$ is Abelian.

$\Lambda \in {}^qL$ with $\Lambda^2 = 1$, which generate, together with ${}^q\mathbf{L}'$ (or ${}^q\mathbf{L}^\uparrow$), these groups. Because of these differences we restrict the following discussion to the case $\frac{1}{2}(q-1)$ odd.

The "space–time geometry" induced in $\mathbf{M}$ by the transformations of ${}^q\mathbf{L}$ shows many similarities to the continious case, which may be seen by the discussion of the orbits of ${}^q\mathbf{L}$ in $\mathbf{M}$. Let us first introduce some definitions. A vector $x \in \mathbf{M}$ is called timelike, spacelike, or lightlike, according as $\epsilon(x^2, q) = 1$, $\epsilon(x^2, q) = -1$, or $\epsilon(x^2, q) = 0$. As a consequence of (2.6), all points y on the line $y = \alpha x$, $\alpha \in GF(q)$, and $x \in \mathbf{M}$, fixed, are spacelike, timelike, or lightlike, according to the character of x. Therefore we may speak about spacelike, etc., directions. The "cones" $\mathbf{V}^+$, $\mathbf{V}^-$, $\mathbf{V}^r$, $\mathbf{V}^l$ and $\bar{\mathbf{V}}$ are subsets of $\mathbf{M}$ defined by

$$\begin{aligned} x \in \mathbf{V}^+&: \epsilon(x_+, q) \geq 0, \quad \epsilon(x_-, q) \geq 0, \\ x \in \mathbf{V}^-&: \epsilon(x_+, q) \leq 0, \quad \epsilon(x_-, q) \leq 0, \\ x \in \mathbf{V}^r&: \epsilon(x_+, q) = +1, \quad \epsilon(x_-, q) = -1, \\ x \in \mathbf{V}^l&: \epsilon(x_+, q) = -1, \quad \epsilon(x_-, q) = +1, \\ \bar{\mathbf{V}} &= \mathbf{V}^r \cup \mathbf{V}^l. \end{aligned} \tag{2.7}$$

The different cones are invariant under the transformations of ${}^q\mathbf{L}' = \{\Lambda(\lambda^2)\}$. They are also invariant under "dilations", i.e., with $x \in \mathbf{V}^n$ there are all $\alpha x \in \mathbf{V}^n$ with $\alpha \in GF(q)$, $\epsilon(\alpha, q) = 1$. With respect to the reflections, we notice $\Pi\mathbf{V}^* = \mathbf{V}^*$, $\Pi\mathbf{V}^l = \mathbf{V}^r$, $\Pi\mathbf{V}^r = \mathbf{V}^l$, $\Pi\bar{\mathbf{V}} = \bar{\mathbf{V}}$, and $I\mathbf{V}^* = \mathbf{V}^*$, $I\mathbf{V}^l = \mathbf{V}^r$, $I\mathbf{V}^r = \mathbf{V}^l$, $I\bar{\mathbf{V}} = \bar{\mathbf{V}}$.

The orbit $\mathbf{Q}_y^i$ of a point y under a subgroup ${}^q\mathbf{L}^i$ in $\mathbf{M}$ is $\mathbf{Q}_y^i = \bigcup_\Lambda \Lambda y$, $\Lambda \in {}^q\mathbf{L}^i$. We shall discuss first the orbits of ${}^q\mathbf{L}'$, $\frac{1}{2}(q-1)$ odd:

(a) $y^2 \neq 0$, $y \in \mathbf{V}^n$, $n = +, -, r, l$. We state $x \in \mathbf{Q}_y'$ if and only if $x^2 = y^2$ and $x \in \mathbf{V}^n$. In order to prove this, set $\lambda = x_+/y_+$; then it follows from (2.2) that $x = \Lambda(\lambda)y$. As x and y are assumed to be in the same cone, we have $\epsilon(\lambda, q) = 1$ and therefore $\Lambda(\lambda) \in {}^qL'$. On the other hand, as x^2 and $\mathbf{V}^n$ are invariant under ${}^q\mathbf{L}'$, $x^2 = y^2$, $y, x \in \mathbf{V}^n$ is necessary for $x \in \mathbf{Q}_y'$. Let us denote these orbits by $Q'(y^2, \sigma) \equiv Q_y'$,

$$\begin{aligned} \sigma = 1 \quad &\text{if} \quad \epsilon(y^2, q) = 1, \quad y \in \mathbf{V}^+, \\ &\text{or} \quad \epsilon(y^2, q) = -1, \quad y \in \mathbf{V}^r, \\ \sigma = -1 \quad &\text{if} \quad \epsilon(y^2, q) = 1, \quad y \in \mathbf{V}^-, \\ &\text{or} \quad \epsilon(y^2, q) = -1, \quad y \in \mathbf{V}^l. \end{aligned} \tag{2.8a}$$

(b) $y^2 = 0$, $y \neq 0$. There are four orbits of this type, which is proved as in (a),

$$\begin{aligned} &y_- = 0, \quad \epsilon(y_+, q) = \pm 1: x \in \mathbf{Q}_y' \equiv \mathbf{Q}'(0, \pm, +) \\ &\text{if and only if } x_- = 0 \text{ and } \epsilon(x_+, q) = \pm 1, \\ &y_+ = 0, \quad \epsilon(y_-, q) = +1: x \in \mathbf{Q}_y' \equiv \mathbf{Q}'(0, \pm, -) \\ &\text{if and only if } x_+ = 0 \text{ and } \epsilon(x_-, q) = \pm 1. \end{aligned} \tag{2.8b}$$

(c) $y = 0$ is a one-point orbit $\mathbf{Q}_0'$.

As space reflection Π transforms $\mathbf{V}^r$ in $\mathbf{V}^l$ and leaves $\mathbf{V}^*$ invariant, we get, as orbits of ${}^q\mathbf{L}^\uparrow$,

$$\mathbf{Q}^\uparrow(y^2, \pm) = \mathbf{Q}'(y^2, \pm) \text{ for } \epsilon(y^2, q) = +1, \tag{2.9a}$$

$$\mathbf{Q}^\uparrow(y^2) = \mathbf{Q}'(y^2, +) \cup \mathbf{Q}'(y^2, -) \quad \text{for } \epsilon(y^2, q) = -1, \tag{2.9b}$$

$$\mathbf{Q}^\uparrow(0, \pm) = \mathbf{Q}'(0, \pm, +) \cup \mathbf{Q}'(0, \pm, -) \quad \text{for } y^2 = 0, \quad y \neq 0, \tag{2.9c}$$

$$\mathbf{Q}_0^\uparrow = \mathbf{Q}_0'. \tag{2.9d}$$

For any two $y, y' \neq 0$ with the same length, there exists $\Lambda \in {}^qL$ so that $y' = \Lambda y$.

This classification of the points in spacelike, timelike and lightlike points, as well as their grouping in orbits, is very similar to the continuous case. There are, of course, some important differences (See Fig. 1). As the total number of points is q^2, the number of all geometric objects is also finite. We have $(q+1)$ lines through the origin, each consisting of q points. $\frac{1}{2}(q-1)$ lines have spacelike, $\frac{1}{2}(q-1)$ lines have timelike, and 2 have lightlike directions. There are $\frac{1}{2}(q-1)$ timelike "hyperbolas" $\mathbf{Q}^\uparrow(y^2, +) \cup \mathbf{Q}^\uparrow(y^2, -)$, $\epsilon(y^2, q) = 1$, and $\frac{1}{2}(q-1)$

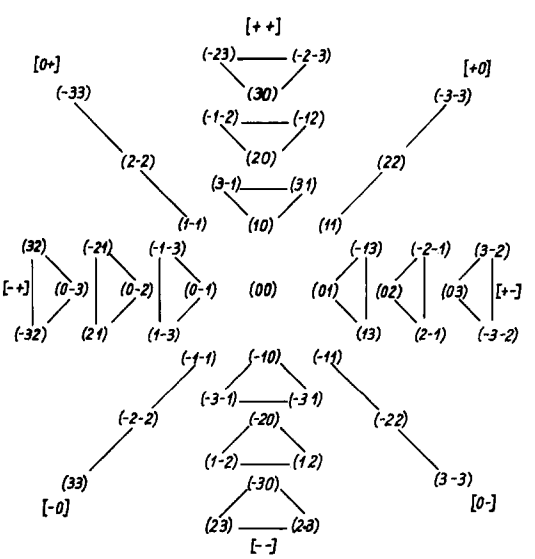


FIG. 1. The orbits of the proper, orthochronous Lorentz group ${}^7L'$ in the 2-dimensional vector space ${}^7M$ over $GF(7)$.

spacelike "hyperbolas" $\mathbf{Q}^{\uparrow}(y^2)$, $\epsilon(y^2, q) = -1$ with $\frac{1}{2}(q-1)$ points on each branch $\mathbf{Q}'(y^2, \sigma)$. The cyclic character of the additive group of $GF(q)$ is the origin of other differences. For example, the sum of two vectors of $\mathbf{V}^+$ sometimes is not an element of $\mathbf{V}^+$ and this affect seriously our further discussion.

Finally, we want to introduce the groups generated by inhomogeneous Lorentz transformations of $\mathbf{M}: x' = \Lambda x + a = (\Lambda, a)x$. These transformations (Λ, a) form a group with the rules

$$(\Lambda, a)(\Lambda', a') = (\Lambda\Lambda', \Lambda a' + a), \qquad (\Lambda, a)^{-1} = (\Lambda^{-1}, -\Lambda^{-1}a). \tag{2.10}$$

We call $\{(\Lambda, a)\}$ the inhomogeneous Lorentz group ${}^{q}\mathbf{C}$ over $GF(q)$, the proper, inhomogeneous Lorentz group ${}^{q}\mathbf{C}_{+}$ over $GF(q)$, the orthochronous inhomogeneous Lorentz group ${}^{q}\mathbf{C}^{\uparrow}$ over $GF(q)$, the proper orthochronous inhomogeneous Lorentz group ${}^{q}\mathbf{C}^{\uparrow}$ over $GF(q)$ if Λ is restricted to ${}^{q}\mathbf{L}$, ${}^{q}\mathbf{L}_{+}$, ${}^{q}\mathbf{L}^{\uparrow}$, ${}^{q}\mathbf{L}'$, respectively. The translation group generated by $\{(1, a)\}$ is a normal subgroup of the different inhomogeneous Lorentz groups with the corresponding homogeneous groups as factor groups.

The group ${}^{q}\mathbf{C}^{\uparrow}$ will play the role of the symmetry group $\mathbf{C}$ in our models of local fields.

3. THE UNITARY REPRESENTATIONS OF THE SPACE-TIME TRANSFORMATION GROUPS

In this section we determine the representations of the inhomogeneous Lorentz group ${}^{q}\mathbf{C}^{\uparrow}$ over $GF(q)$ by linear transformations of a finite or infinite Hilbert space $\mathbf{H}$ over the ordinary complex numbers. As ${}^{q}\mathbf{C}$ is of finite order, every representation is equivalent to a direct sum of irreducible, unitary representations.[7] Therefore it is sufficient to look for these representations.

The group ${}^{q}\mathbf{C}^{\uparrow}$ is the semidirect product of ${}^{q}\mathbf{L}^{\uparrow}$, with the Abelian translation group $\mathbf{T} = \{(1, a)\}$. We may apply the general method of Wigner–Mackey[8] for the construction of the representations of semidirect products. As we are interested in various groups of this structure, we shall state next the results of Wigner and Mackey in general terms.

Let $\mathbf{A}$ and $\mathbf{B}$ be two groups, and assume that for every $\beta \in \mathbf{B}$ there is an automorphism of $\mathbf{A}: a \rightleftarrows a^{\beta}$, $(a_1 a_2)^{\beta} = (a_1)^{\beta}(a_2)^{\beta}$, such that

$$a^1 = a, \qquad (a^{\beta_2})^{\beta_1} = a^{\beta_1\beta_2} \qquad \beta_1, \beta_2 \in \mathbf{B}. \tag{3.1}$$

Then we define a group structure for the pairs (β, a), $\beta \in \mathbf{B}$, $a \in \mathbf{A}$ by the following rules:

$$(\beta, a)(\beta', a') = (\beta\beta', a(a')^{\beta}), \qquad (\beta^{-1}, a)^{-1} = (\beta, (a^{-1})^{\beta}), \qquad (1_B, 1_A) = 1, \tag{3.2}$$

as we can verify by straightforward calculations. This group $\mathbf{G} = \{(\beta, a)\}$ is called the semidirect product of $\mathbf{A}$ and $\mathbf{B}$: $\mathbf{G} = \mathbf{BA}$. The rules (3.2) are identical with the rules (2.10) for the ${}^{q}\mathbf{C}$, if we identify $\mathbf{B}$ with ${}^{q}\mathbf{L}^{\uparrow} = \{\Lambda\}$, $\mathbf{A}$ with the additively written Abelian group $\mathbf{T} = \{a\}$, and therefore, a^{β} with Λa.

In order to construct the unitary representations of $\mathbf{BA}$, we consider first the irreducible unitary representations $\hat{\mathbf{A}}$ of $\mathbf{A}$. We assume now that $\mathbf{A}$ is Abelian. Then $\hat{\mathbf{A}}$ is the set $\{\chi(a, p)\}$ of complex valued functions on $\mathbf{A}$, which satisfy

$$\chi(a_1 + a_2, p) = \chi(a_1, p)\chi(a_2, p), \quad |\chi(a, p)| = 1. \tag{3.3}$$

The different "characters" $\chi(a, p) \in \hat{\mathbf{A}}$ are distinguished by the index p. The automorphism $a \rightleftarrows a^{\beta}$, $\beta \in \mathbf{B}$, associates with $\chi(a, p)$ the character

$$\chi(a^{\beta}, p) = \chi(a, \beta^{-1}(p)). \tag{3.4}$$

By this definition $\mathbf{B}$ becomes a transformation group of $\hat{\mathbf{A}}$:

$$\beta_1(\beta_2(p)) = \beta_1\beta_2(p), \quad 1(p) = p, \quad \beta_1, \beta_2 \in \mathbf{B}, \tag{3.5}$$

and we can partition $\hat{\mathbf{A}}$ in orbits of associate characters

$$\mathbf{P}^i = \cup_{\beta}\beta(p_i), \qquad \beta \in \mathbf{B}. \tag{3.6}$$

The index i labels the distinct equivalence classes of associate characters. The "little group" of a point of an orbit $\mathbf{P}^i$ is defined as the following subgroup $\mathbf{W}_p^i$ of $\mathbf{B}$:

$$\beta \in \mathbf{W}_p^i \text{ if and only if } \chi(a, p) = \chi(a^{\beta}, p). \tag{3.7}$$

Little groups of different points of the same orbit $\mathbf{P}^i$ are conjugate subgroups of $\mathbf{B}$: $\mathbf{W}_{\beta(p)}^i = \beta\mathbf{W}_p^i\beta^{-1}$, and therefore the same abstract groups. Thus, there is a unique little group associated with each orbit.

The main theorem about the representations of semidirect products of groups states: The orbits $\mathbf{P}^i$ of $\hat{\mathbf{A}}$, together with all the irreducible representations of the corresponding little groups, determine uniquely all the nonequivalent irreducible representations of $\mathbf{BA}$. The arguments run as follows. Given an irreducible, unitary representation of $\mathbf{BA}$: $(\beta,a) \to U(\beta, a)$ in $\mathbf{H}$, we can introduce a basis $\{|p, \eta\rangle\}$ of $\mathbf{H}$ in which the representation of $\mathbf{A}$ by $U(1, a) \equiv U(a)$ is completely reduced:

[7] For the concepts of group theory, see M. Hall, Ref. 2; J. S. Lomont, *Applications of finite groups* (Academic Press Inc., New York, London, 1959).

[8] E. P. Wigner, Ann. Math. 40, 149 (1939); G. W. Mackey, Acta Math. 99, 265 (1958), and previous papers (University of Chicago mimeographed notes, 1955).

$$U(a)\,|p,\eta\rangle = \chi(a,p)\,|p,\eta\rangle; \tag{3.8}$$

η is a degeneration parameter. Now we calculate

$$\begin{aligned} U(a)U(\beta^{-1})\,|p,\eta\rangle &= U(\beta^{-1})U(a^{\beta})\,|p,\eta\rangle \\ &= \chi(a,\beta^{-1}(p))U(\beta^{-1})\,|p,\eta\rangle. \end{aligned} \tag{3.9}$$

Therefore, an irreducible representation of **BA** restricted to **A** contains all the characters $\chi(a,p)$ of one certain orbit $\mathbf{P}^i$. Applying formula (3.9) to $\rho \in \mathbf{W}_p^i$ we see that $U(\rho)$ transforms the subspace with fixed p in itself and therefore we get

$$U(\rho)\,|p,\eta\rangle = \sum_{\eta'} |p,\eta'\rangle\, D_{\eta'\eta}(\rho), \tag{3.10}$$

where $D_{\eta'\eta}(\rho)$ is a representation of the little group $\mathbf{W}_p^i$. There is, according to a standard procedure, a 1–1 correspondence between the points p of an orbit $\mathbf{P}^i$ and the left cosets of $\mathbf{W}_p^i$ in **B**, since from $\beta_1(p^i) = \beta_2(p^i)$ follows $\beta_1 = \beta_2\rho$ with $\rho \in \mathbf{W}_{pi}^i$. Let us take a representative $b(p,p^i)$ from each coset,

$$b(p,p^i)(p^i) = p; \tag{3.11}$$

then (3.9) gives

$$U(a)U(b(p,p^i))\,|p^i,\eta\rangle = \chi(a,p)U(b(p,p^i))\,|p^i,\eta\rangle.$$

Therefore, $U(b(p,p^i))$ maps the subspace with character p^i onto the subspace with character p. By redefining the degeneration parameter η for the different p, we get a base in **H** such that

$$U(b(p,p^i))\,|p^i,\eta\rangle = |p,\eta\rangle. \tag{3.12}$$

We remark that the element defined by

$$\rho(\beta,p) = b^{-1}(\beta(p),p^i)\beta b(p,p^i) \tag{3.13}$$

belongs to $\mathbf{W}_p^i$. We calculate with help of (3.10), (3.12), and (3.13),

$$\begin{aligned} U(\beta)\,|p,\eta\rangle &= U(b(\beta(p),p^i))U(\rho(\beta,p)) \\ &\qquad \times U(b^{-1}(p,p^i))\,|p,\eta\rangle \\ &= \sum_{\eta'} |\beta(p),\eta'\rangle\, D_{\eta'\eta}(\rho(\beta,p)), \end{aligned}$$

for all $\beta \in \mathbf{B}$. It is evident that, for an irreducible representation of **BA**, $D(\rho)$ has to be an irreducible representation of $\mathbf{W}_p^i$. This equation, together with (3.8), gives

$$\begin{aligned} U(a)\,|p,\eta\rangle &= \chi(a,p)\,|p,\eta\rangle, \qquad p \in \mathbf{P}^i, \\ U(\beta)\,|p,\eta\rangle &= \sum_{\eta'} |\beta(p),\eta'\rangle\, D_{\eta'\eta}(\rho(\beta,p)), \end{aligned} \tag{3.14}$$

the explicit construction of the irreducible representations of **BA** in terms of the orbits $\mathbf{P}^i$, and the irreducible representations of the corresponding little groups $\rho \to D(\rho)$, $\rho \in \mathbf{W}_p^i$.

We use now this construction to get the representations of the groups ${}^q\mathbf{C}_n = {}^q\mathbf{L}_n\mathbf{T}$, $(n = \prime, \uparrow, +)$. The characters of the translation group are

$$\chi(a,p) = \exp\,[(2\pi i/q)(a_0p_0 - a_1p_1)], \tag{3.15}$$

a_μ, p_ν residue classes mod q, i.e., elements of $GF(q)$. Under the transformations $\Lambda(p)$ defined by $\chi(\Lambda^{-1}a,\,p) = \chi(a,\,\Lambda(p))$, $\Lambda \in {}^qL^n$, $p = (p_0,\,p_1)$ transforms like a vector of **M**: $\Lambda(p) = \Lambda p$. We discussed the orbits for these transformations in the preceding section, [Eq. (2.8) and (2.9)]. The little groups of the orbits are generated by certain reflections depending on the group ${}^q\mathbf{C}_n$ and the orbit, with the exception of the orbit $\mathbf{P}^0$: $p = 0$, for which the little group is ${}^q\mathbf{L}_n$. With these remarks we get immediately the irreducible representations of ${}^q\mathbf{C}_n$ from (3.14). For ${}^q\mathbf{C}^\uparrow$ we have, therefore, the following representations:

$$U(a)\,|p\rangle = \exp\,[(2\pi i/q)\cdot ap]\,|p\rangle$$

(i) $p^2 = M$, $\epsilon(M,q) = +1$, $\sigma = \pm 1$,

$$\epsilon(p_+,q) = \epsilon(p_-,q) = \rho = \pm 1,$$

$$U(\Lambda)\,|p\rangle = \sigma^r\,|\Lambda p\rangle \text{ for } \Lambda = \Lambda(\lambda)\Pi^r \in {}^q\mathbf{L}^\uparrow; \tag{3.16a}$$

(ii) $p^2 = M$, $\epsilon(M,q) = -1$,

$$U(\Lambda)\,|p\rangle = |\Lambda p\rangle \text{ for } \Lambda \in {}^q\mathbf{L}^\uparrow; \tag{3.16b}$$

(iii) $p^2 = 0$, $p \neq 0$,

$$\epsilon(p_+,q) = \epsilon(p_-,q) = \rho = \pm 1,$$

$$U(\Lambda)\,|p\rangle = |\Lambda p\rangle \text{ for } \Lambda \in {}^q\mathbf{L}^\uparrow; \tag{3.16c}$$

iv $p \equiv 0$; $U(\Lambda)$ is an irreducible representation of ${}^q\mathbf{L}^\uparrow$. (3.16d)

In order to complete the discussion of the irreducible representations of ${}^q\mathbf{C}^\uparrow$, we have to construct the representations of ${}^q\mathbf{L}^\uparrow$. As ${}^q\mathbf{L}^\uparrow$ is generated by an $\Lambda(\lambda_0)$ of order $\frac{1}{2}(q-1)$ and Π with the relations (2.3), ${}^q\mathbf{L}^\uparrow$ is isomorphic to the dihedral group of order $(q-1)$. There is another subgroup of ${}^q\mathbf{C}^\uparrow$, which will later interest us, and which has the structure of a dihedral group, namely the "Euclidean group ${}^q\mathbf{E}$" generated by the translation τ: $x_0' = x_0$, $x_1' = x_1 + 1$, and the reflection Π. The transformations τ and Π satisfy the relations $\tau^q = 1$, $\Pi^2 = 1$, and $\Pi\tau\Pi = \tau^{-1}$. This group ${}^q\mathbf{E}$ is therefore isomorphic to the dihedral group of order $2q$. The general dihedral group of order $2n$ generated by S and T with the relations

$$S^2 = 1, \quad T^n = 1, \quad STS = T^{-1}, \tag{3.17}$$

is a semidirect product of the cyclic group of order n generated by T, and the cyclic group of order 2 generated by S. With the well-known characters of the cyclic groups: $\chi(T^r,\,m) = \exp\,(2\pi irm/n)$,

$m = 0, 1, \cdots, n-1$, we get from (3.14) the irreducible representations of the dihedral groups:

$$\text{(i)}\quad U(T^\tau S^\rho) = 1; \tag{3.18a}$$

$$\text{(ii)}\quad U(T^\tau S^\rho) = (-1)^\rho; \tag{3.18b}$$

$$\text{(iii)}\quad U(T^\tau S^\rho) = \begin{bmatrix} \exp(2\pi i r m/n) & 0 \\ 0 & \exp -(2\pi i r m/n) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^\rho, \tag{3.18c}$$

$m = 1, 2, \cdots, \frac{1}{2}(n-1)$ for n odd,

$m = 1, 2, \cdots, \frac{1}{2}(n-2)$ for n even;

$$\text{(iv)}\quad U(T^\tau S^\rho) = (-1)^{\tau+\rho}, \quad \text{for even } n,$$

$$\text{(v)}\quad U(T^\tau S^\rho) = (-1)^\tau, \quad \text{for even } n. \tag{3.18d}$$

With $n = \frac{1}{2}(q-1)$, these are the representations of the ${}^q\mathbf{C}^\uparrow$ needed in (3.16d).

The groups which we have considered until now are all finite groups. Therefore, we have at hand all the standard tools of the representation theory of such groups, as for instance, the orthogonality and completeness relations of the representation matrices and characters. We do not want to discuss these for our groups now, but we only give the characters of the representations of the ${}^q\mathbf{C}^\uparrow$, as listed in (3.16a–c). The matrix elements $\langle p'|\ U(\Lambda, a)\ |p\rangle$ are, according to (3.16),

$$\langle p'|\ U(\Lambda, a)\ |p\rangle = \sigma^\tau \delta(\Lambda p - p^i) \exp\left[(2\pi i/q) a \Lambda p\right] \quad \text{for} \quad \Lambda = \Lambda(\lambda)\Pi^\tau;$$

therefore we get, for the characters $\Delta(\Lambda, a; \xi) = \sum_p \langle p|\ U(\Lambda, a)\ |p\rangle$ of the different types of representations,

(a) $\epsilon(M, q) = +1$, $p \in \mathbf{V}^\rho$, parity σ:

$$\Delta(\Lambda, a; M, \rho, \sigma) = \sum_p \exp(2\pi i \cdot pa/q), \quad p^2 = M, \quad p \in \mathbf{V}^\rho, \quad \text{for} \quad \Lambda = 1,$$

$$= 0 \quad \text{for} \quad \Lambda \in {}^q\mathbf{L}', \quad \Lambda \neq 1,$$

$$= \sigma \exp\left[(i\pi\mu/q)(\lambda a_- + \lambda^{-1} a_+)\right] \quad \text{for} \quad \Lambda = \Lambda(\lambda^2)\Pi, \quad \mu^2 = M.$$

(b) $\epsilon(M, q) = -1$:

$$\Delta(\Lambda, a; M) = \sum_p \exp(2\pi i \cdot pa/q), \quad p^2 = M, \quad \text{for} \quad \Lambda = 1,$$

$$= 0 \quad \text{for} \quad \Lambda \neq 1.$$

(c) $p^2 = 0$, $p \in \mathbf{V}^\rho$:

$$\Delta(\Lambda, a; 0, \rho) = \sum_p \exp(2\pi i \cdot pa/q), \quad p^2 = 0, \quad p \in \mathbf{V}^\rho, \quad \text{for} \quad \Lambda = 1,$$

$$= 0 \quad \text{for} \quad \Lambda \neq 1.$$

For the arguments $\Lambda = 1$, $a = x$, these characters correspond to the invariant functions $\Delta^+(x)$ and $\Delta^-(x)$ used in quantum field theory.[9]

4. THE AXIOMS OF LOCAL FIELDS

We formulate now the axioms for local fields in our model of space–time along the lines discussed in the introduction. With the notation of the preceding sections, these axioms are:

(i) *Quantum mechanical description.* There is a mapping $x \to A(x)$ of ${}^q\mathbf{M}$ in the set of linear, self-adjoint operators of a complex Hilbert space $\mathbf{H}$.

(ii) *Transformation Law of the Field.* $(\Lambda, a) \to U(\Lambda, a)$ is a unitary representation $U({}^q\mathbf{C}^\uparrow)$ of the orthochronous inhomogeneous Lorentz group over $GF(q)$ by transformations of $\mathbf{H}$, which transform the $A(x)$ covariantly:

$$U(\Lambda, a) A(x) U^{-1}(\Lambda, a) = A(\Lambda x + a).$$

(iii) *Existence of the Vacuum; Spectrum Condition.*

(a) There is precisely one independent vector $\Omega \in \mathbf{H}$ which is invariant under all $U(\Lambda, a)$: $U(\Lambda, a)\Omega = \Omega$.

(b) The characters p of the translations $U(1, a)$ are contained in the "forward cone": $p \in \mathbf{V}^+$; $p \neq 0$ except for Ω.

(iv) *Locality.*

$$[A(x), A(x')]_- = 0 \quad \text{if} \quad \epsilon((x-x')^2, q) = -1. \tag{4.1}$$

Before we discuss the algebraic structure which is characterized by (i) to (iv), we would like to make a remark on the problem of defining an S matrix in such a field theory. In physical field

[9] J. Schwinger, Phys. Rev. **75**, 651 (1949); D. Hall and A. S. Wightman, Phys. Rev. **99**, 674 (1955).

theories there is defined[10] a "free field" $A_{in}(x)$ and an "interpolating field" $A_1(x)$, in the same space of states $\mathbf{H}$ and with the same representation $U(\Lambda, a)$ of the inhomogeneous, orthochronous Lorentz group. The TCP theorem assures[11] the existence of an uniquely defined antiunitary TCP operator Θ_{in} (or Θ_1) for both fields, with the properties

$$\Theta_s A_s(x)\Theta_s = A_s(-x), \quad \Theta\Omega = \Omega, \quad \Theta^2 = 1, \quad s = in, 1,$$
$$\Theta_s U(\Lambda)\Theta_s = U(\Lambda), \quad \Theta_s U(a)\Theta_s = U(-a). \tag{4.2}$$

The unitary S Matrix which transforms the $A_{in}(x)$ into the $A_{out}(x) = \Theta_1 A_{in}(-x)\Theta_1$ is then $S = \Theta_{in}\Theta_1$.

This definition can be adapted to our models. Let $A_1(x)$ and $A_2(x)$ be local fields of ${}^q\mathbf{M}$ which are defined in the same $\mathbf{H}$, and transform covariantly under the same representation $U(\Lambda, a)$ of ${}^q\mathbf{C}^\uparrow$. We assume further the existence of TCP operators Θ_1, Θ_2 for these fields which satisfy the conditions (4.2). Then we may call the unitary operator $S_{12} = \Theta_1\Theta_2$ the relative S matrix of the two fields $A_1(x)$ and $A_2(x)$. If the algebras generated by the fields $A(x)$ are irreducible, the Θ_s and therefore S_{12} are defined uniquely by (4.2) if they exist. S_{12} is unitary and as a consequence of (4.2) invariant under the transformations of ${}^q\mathbf{C}$,

$$U(\Lambda, a)S_{12}U^{-1}(\Lambda, a) = S_{12}.$$

The free fields of our models, which can be defined by the commutation relation

$$[A(x), A(y)]_- = i\Delta(x - y, M),$$

violate the "spectrum condition" (iii) of (4.1), as we shall see in Sec. 6. This fact prevents a close analogy of the definition of the S matrix.

5. THE STRUCTURE OF LOCAL FIELDS

In this Section we make some general remarks about the problem of the construction of local fields. If we are interested in fields $A(x)$ of finite-dimensional matrices, we may consider fields as certain unitary representations of a group, which we shall describe now.

With a finite local matrix field $A(x)$, every polynomial pol $(A(x))$ is also such a field. Suppose $A(0)$ has n distinct real eigenvalues α_m. Then there is a polynomial such that

$$\text{pol}\,(\alpha_m) = \exp(2\pi i m/n) \quad \text{for} \quad m = 1, 2, \cdots, n. \tag{5.1}$$

The field $B(x) = \text{pol}\,(A(x)) = U(x)[\text{pol}\,(A(0))]U^{-1}(x)$ is then unitary and satisfies the equation $B^n(x) = 1$. We consider now the groups $\mathbf{B}(q, n)$ and $\mathbf{CB}^\uparrow(q, n)$ defined by generating elements with relations.[7] The group $\mathbf{B}(q, n)$ is generated by the elements $\mathfrak{B}(x)$, $x \in {}^q\mathbf{M}$ with the relations

$$\mathfrak{B}^n(x) = 1, \qquad \mathfrak{B}(x)\mathfrak{B}(y)\mathfrak{B}^{-1}(x)\mathfrak{B}^{-1}(y) = 1 \quad \text{for} \quad \epsilon((x - y)^2, q) = -1 \tag{5.2}$$

The generating elements of $\mathbf{CB}^\uparrow(q, n)$ are the $\mathfrak{B}(x)$ with the relations (5.2) together with the elements of the inhomogeneous Lorentz group $(\Lambda, a) \rightleftarrows \mathfrak{C}_i \in {}^q\mathbf{C}$ which satisfy besides the relations of ${}^q\mathbf{C}^\uparrow$ also

$$\mathfrak{C}_i\mathfrak{B}(x)\mathfrak{C}_i^{-1} = \mathfrak{B}(c_i x)$$
$$\text{with} \quad c_i x = \Lambda x + a \quad \text{if} \quad \mathfrak{C}_i \to (\Lambda, a). \tag{5.3}$$

${}^q\mathbf{C}$ is a subgroup and $\mathbf{B}(q, n)$ is a normal subgroup of $\mathbf{CB}^\uparrow(q, n)$. As a consequence of (5.3), we can write every element of $\mathbf{CB}^\uparrow(q, n)$ in the form $\mathfrak{B}\mathfrak{C}$ with $\mathfrak{B} = \mathfrak{B}^{r_1}(x_1) \cdots \mathfrak{B}^{r_s}(x_s) \in \mathbf{B}(q, n)$, $r_i = 0, 1, \cdots, n - 1$, and ${}^q\mathbf{C}^\uparrow$. We have

$$\mathfrak{B}\mathfrak{C}\mathfrak{B}'\mathfrak{C} = \mathfrak{B}\mathfrak{B}'^{\mathfrak{C}}\mathfrak{C}\mathfrak{C}', \quad \text{with} \quad \mathfrak{B}^{\mathfrak{C}} = \mathfrak{C}\mathfrak{B}\mathfrak{C}^{-1},$$
$$\mathfrak{B}^{\mathfrak{C}} = \mathfrak{B}^{r_1}(cx_1) \cdots \mathfrak{B}^{r_s}(cx_s) \in \mathbf{B}(q, n). \tag{5.4}$$

Therefore $\mathbf{CB}^\uparrow(q, n)$ is the semidirect product of ${}^q\mathbf{C}$ with $\mathbf{B}(q, n)$ according to the definition (3.2).

It is an immediate consequence of the relations (5.2) and (5.3) that every unitary representation of $\mathbf{CB}^\uparrow(q, n)$ is a local field provided its restriction to ${}^q\mathbf{C}$ satisfies the spectrum condition (iii) of (4.1). According to our preliminary remarks, there corresponds to every finite-dimensional matrix field such a unitary representation of $\mathbf{CB}^\uparrow(q, n)$ with a certain n. The problem of finding all finite-dimensional local matrix fields is therefore identical with the problem of constructing all unitary representations of the group $\mathbf{CB}^\uparrow(q, n)$ which satisfy the spectrum condition.

From the point of view of representation theory, this is not a simple problem. As the nontrivial classes of conjugate elements in $\mathbf{CB}^\uparrow(q, n)$ have infinite order, the regular representation is[12] of type II_1, i.e., it cannot be uniquely decomposed in irreducible unitary representations. Therefore, representation theory gives no straight forward procedure for the solution of our problem. On the other hand, representation theory contributes some interesting points of view, which we plan to study more systematically in a future paper. In our present exposition we can only indicate some of the ideas by constructing an explicit example of a local field.

There are some interesting subgroups of $\mathbf{CB}^\uparrow(q, n)$

[10] R. Haag, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **29**, No. 12 (1955); H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **6**, 319 (1957).

[11] R. Jost, Helv. Phys. Acta **30**, 409 (1957).

[12] I. Kaplansky, Tohoku Math. J. **3**, 249 (1951).

which are defined similarly to $\mathbf{B}(q, n)$ and $\mathbf{CB}^{\uparrow}(q, n)$. Let $\mathbf{D}$ be a subgroup of ${}^{q}\mathbf{C}^{\uparrow}$ and let $\mathbf{N}$ be an orbit of $\mathbf{D}$ in ${}^{q}\mathbf{M}$; then we denote by $\mathbf{B}(q, n, \mathbf{N})$ the subgroup generated by the $\mathfrak{B}(x)$ with $x \in \mathbf{N}$ and by $\mathbf{DB}(q, n, \mathbf{N})$ the subgroup generated by $\mathfrak{C} \in \mathbf{D}$ and $\mathbf{B}(q, n, \mathbf{N})$. As in (5.4), the group $\mathbf{DB}(q, n, \mathbf{N})$ is the semidirect product of $\mathbf{D}$ with $\mathbf{B}(q, n, \mathbf{N})$. For $\mathbf{D} = {}^{q}\mathbf{L}^{\uparrow}$ and $\mathbf{N} = \{0\}$, the group $\mathbf{B}(q, n, \mathbf{N})$ is the cyclic group of order n generated by $\mathfrak{B}(0)$, and $\mathbf{DB}(q, n, \mathbf{N})$ is the direct product of this cyclic group with ${}^{q}\mathbf{L}^{\uparrow}$.

We study now another example with $D = E =$ "euclidean group" $\{(R, z)\}$, $R = 1, \Pi$, $z = (0, z)$, and $\mathbf{N} = \{(0, y)\}$. According to (5.2) all $\mathfrak{B}((0, y)) = \mathfrak{B}_{\cdot}(y)$ commute; therefore $\mathbf{B}(q, n, \mathbf{N})$ is the direct product of q cyclic groups of order n generated by the $\mathfrak{B}_{\cdot}(y)$. The group $\mathbf{EB}(q, n, \mathbf{N})$ is then the semidirect product of $\mathbf{E}$ with this Abelian group. Therefore we may apply the Wigner–Mackey theory explained in Sec. 3 in order to construct the irreducible unitary representations of $\mathbf{EB}(q, n, \mathbf{N})$. The characters of $\mathbf{B}(q, n, \mathbf{N})$ are

$$\chi[\prod_{y} \mathfrak{B}^{r(y)}(y), b(y')] = \exp\,[(2\pi i/n) \sum_{y} b(y)r(y)],$$
$$y, y' = -\tfrac{1}{2}(q-1), -\tfrac{1}{2}(q-1)+1, \cdots, \tfrac{1}{2}(q-1). \tag{5.4}$$

The indices p of the characters are the functions $b(y)$, $y \in GF(q)$, $b(y) = 0, 1, \cdots, n-1$, and $\mathbf{E}$ is a transformation group of these functions according to (3.4), (5.3), and (5.4):

$$(R, z)[b(y)] = b(R(y - z)). \tag{5.5}$$

The q^n functions $b(y)$ can be partitioned into orbits $\mathbf{P}^i$ with respect to these transformations. We classify these orbits according to possible little groups $\mathbf{W}^i \subset E$:

(i) $\mathbf{W}^i = 1$: all functions $(R, z)[b(y)] \in P^i$ are different; (5.6a)

(ii) $\mathbf{W}^i = \{1, \Pi\}$ or, conjugate to this group. There is a $b(y) \in \mathbf{P}^i$ with $b(-y) = b(y)$; (5.6b)

(iii) $\mathbf{W}^i = \{(1, z)\}$: it follows $b(y) =$ constant and therefore $\mathbf{W}^i = \mathbf{E}$. (5.6c)

As the irreducible representations of $\mathbf{EB}(q, n, \mathbf{N})$ depend on the orbits $\mathbf{P}^i$ and the irreducible representations of the $\mathbf{W}^i$, we get a corresponding classification for the representations of $\mathbf{EB}(q, n, \mathbf{N})$. The explicit form of these representations $\mathfrak{B}_{\cdot}(y) \to B_{\cdot}(y)$, $(R, z) \to U(R, z)$, follows from (3.14):

(i) $B_{\cdot}(y)\,|b(y')\rangle = \exp\,(2\pi i b(y)/n)\,|b(y')\rangle$,
$U(R, z)\,|b(y')\rangle = |b(R(y' - z))\rangle$,
$b(y') \in \mathbf{P}^i$ of type (5.6a); (5.7a)

(ii) $B_{\cdot}(y)\,|b(y')\rangle = \exp\,(2\pi i b(y)/n)\,|b(y')\rangle$,
$U(R, z)\,|b(y')\rangle = \sigma^r\,|b(R(y' - z))\rangle$,
$\sigma = \pm 1$, $r = 0, 1$ according to $R = 1, \Pi$,
$b(y') \in \mathbf{P}^i$ of type (5.6b); (5.7b)

(iii) $B_{\cdot}(y)\,|\xi\rangle = \exp\,(2\pi i b_0/n)\,|\xi\rangle$,
$U(R, z)\,|\xi\rangle = \sum_{\xi'} |\xi'\rangle\, D_{\xi\xi'}(R, z)$,
$b_0 = 0, 1, \cdots, n-1$, and $D(R, x)$ is an irreducible representation of $\mathbf{E}$ as discussed in (3.18). (5.7c)

We shall use these representations of $\mathbf{EB}(q, n, \mathbf{N})$ for the construction of our example of a local field.

6. EXAMPLES

Finally we shall show that there is a local field which satisfies the axioms (4.1), i.e., there is a representation of the group $\mathbf{CB}^{\uparrow}(q, n)$, which, restricted to ${}^{q}\mathbf{C}^{\uparrow}$, satisfies the spectrum condition.

Let us first make the following preliminary remarks. The restriction of a representation of $\mathbf{CB}^{\uparrow}(q, n)$ to $\mathbf{EB}(q, n, \mathbf{N})$ is equivalent to a direct sum of irreducible representations of type (5.7). On the other hand, a representation of $\mathbf{EB}(q, n, \mathbf{N})$ must satisfy the following conditions, in order that it can be extended to a representation of a local field:

(a) The restriction of $U(\mathbf{EB}(q, n, \mathbf{N}))$ to ${}^{q}\mathbf{E}$ must be a restriction of a representation $U({}^{q}\mathbf{C}^{\uparrow})$ of ${}^{q}\mathbf{C}^{\uparrow}$ which satisfies the spectrum condition.

(b) The operator $B_{\cdot}(0)$ must be invariant under the restriction of $U({}^{q}\mathbf{C}^{\uparrow})$ to ${}^{q}\mathbf{L}^{\uparrow}$.

If these two conditions are satisfied for a representation of $\mathbf{EB}(q, n, \mathbf{N})$, then the representation $U({}^{q}\mathbf{C}^{\uparrow})$ together with $B(x) = U(x)[B_{\cdot}(0)]U^{-1}(x)$ define a local field.

For the following representation of $\mathbf{EB}(q, 2, \mathbf{N})$ these conditions can be satisfied:

$$B_{\cdot}(y)\,|y'\rangle_{\cdot} = (2\delta(y - y') - 1)\,|y'\rangle_{\cdot},$$
$$U(z)\,|y'\rangle_{\cdot} = |y' - z\rangle_{\cdot}, \quad U(\Pi)\,|y'\rangle_{\cdot} = |-y'\rangle_{\cdot},$$
$$y' = -\tfrac{1}{2}(q-1), \cdots, \tfrac{1}{2}(q-1). \tag{6.1}$$

This representation is of type (5.7b) with $\sigma = 1$, and the orbit $\mathbf{P}$ which contains the function $b(y) = 1$ for $y = 0$, $b(y) = -1$ for $y \neq 0$. We decompose the representation of ${}^{q}\mathbf{E}$ contained in (6.1) by "Fourier transformation",

$$|p_1\rangle_F = q^{-\frac{1}{2}} \sum_y \exp(-2\pi i p_1 y/q)\,|y\rangle_e,$$
$$p_1 = -\tfrac{1}{2}(q-1), \cdots, \tfrac{1}{2}(q-1).$$

From

$$U(z)\,|p_1\rangle_F = \exp(2\pi i p_1 z/q)\,|p_1\rangle_F$$

and (6.3)

$$U(\Pi)\,|p_1\rangle_F = |-p_1\rangle_F,$$

follows that $\{|0\rangle_F\}$ and $\{|p_1\rangle_F, |-p_1\rangle_F\}$ are the bases of the irreducible subrepresentations of ${}^q\mathbf{E}$ in the form (3.18). The restriction to ${}^q\mathbf{E}$ of the representation of ${}^q\mathbf{C}$, consisting of the direct sum of the trivial representation

$$U(\Lambda, a)\Omega = \Omega = |0, 0\rangle, \tag{6.4}$$

and the representation (3.16c) with $p^2 = 0$, $p \neq 0$, $p \in \mathbf{V}^+$,

$$U(a)\,|p\rangle = \exp(2\pi i \cdot pa/q)\,|p\rangle, \qquad U(\Lambda)\,|p\rangle = |\Lambda p\rangle, \tag{6.4'}$$

is equivalent to the representation defined by (6.1) and (6.3) as is shown by the equivalence transformation

$$\Omega = \eta_0\,|0\rangle_F, \qquad |p_0, p_1\rangle = \eta(p_1)\,|p_1\rangle_F, \qquad \eta(p_1) = \eta(-p_1). \tag{6.5}$$

Therefore condition (a) is satisfied.

Let $B_e(0) = E^+ - E^-$ be the spectral decomposition of $B_e(0)$. In order to satisfy condition (b), the range of E^+ or E^- has to be invariant under the transformations of ${}^q\mathbf{L}$. We get from (6.1), (6.2), and (6.5):

$$E^+\Omega = q^{-\frac{1}{2}}\eta_0\,|0\rangle_e, \quad E^-\Omega = q^{-\frac{1}{2}}\eta_0 \sum_{y\neq 0} |y\rangle_e,$$
$$E^+\,|p\rangle = q^{-\frac{1}{2}}\eta(p_1)\,|0\rangle_e,$$
$$E^-\,|p\rangle = q^{-\frac{1}{2}}\eta(p_1) \sum_{y\neq 0} \exp(-2\pi i p_1/q)\,|y\rangle_e. \tag{6.6}$$

The necessary and sufficient condition for

$$U(\Lambda)E^+\Omega = E^+\Omega, \quad U(\Lambda)E^+\,|p\rangle = E^+\,|\Lambda p\rangle \tag{6.7}$$

is, therefore, $\eta((\Lambda p)_1) = \eta(p_1) = \eta_1$ for all $\Lambda \in {}^q\mathbf{L}^\uparrow$. With this choice for the phases η, we determine the transformation of the $|y\rangle_e$, $y \neq 0$, from the condition

$$U(\Lambda)E^-\,|p\rangle = E^-\,|\Lambda p\rangle. \tag{6.8}$$

We get as a consequence

$$U(\Lambda)E^-\Omega = -(\eta_0/\eta_1)U(\Lambda)\sum_{p\neq 0} E^-\,|p\rangle = E^-\Omega. \tag{6.9}$$

Therefore we satisfy condition (b) by choosing $\eta(p_1) = \text{constant} = \eta_0$ in (6.5), and we get finally, as the matrix elements of a local field,

$$\langle p|\,B(x)\,|p'\rangle = (2/q)\exp(2\pi i(p-p')/q) - \delta(p-p'). \tag{6.10}$$

As another example, we discuss the "free fields" $A(x)$ which are defined in analogy to the physical case by the "field equations"

$$\sum_y \Delta^{(1)}(x-y, M)A(y) = A(x), \qquad \epsilon(M, q) \geq 0, \tag{6.11}$$

and the c-number commutation relations

$$[A(x), A(y)]_- = i\Delta(x-y, M). \tag{6.12}$$

The invariant function $\Delta(x)$ and $\Delta^{(1)}(x)$ are linear combinations of characters (3.19) of representations of ${}^q\mathbf{C}$,

$$\begin{aligned}\Delta^{(1)}(x) &= q^{-2}[\Delta(1, x; M, +, +) + \Delta(1, x; M, -, +)] \\ &= q^{-2}\sum_p \exp(2\pi i px/q), \quad p^2 = M, \\ \Delta(x) &= -iq^{-2}[\Delta(1, x; M, +, +) - \Delta(1, x; M, -, +)] \\ &= -iq^{-2}\sum_p \epsilon(p_+, q)\exp(2\pi i px/q), \quad p^2 = M.\end{aligned} \tag{6.13}$$

Equations (6.11) and (6.12) imply, for the Fourier transformations of the fields

$$\tilde{A}(p) = q^{-1}\sum_x \exp(-2\pi i px/q)A(x), \qquad (p^2 - M)\tilde{A}(p) = 0, \tag{6.14}$$

and

$$[A(p), A(p')]_- = \delta(p+p')\delta(p^2 - M)\epsilon(p_+, q), \qquad A^+(p) = A(-p). \tag{6.15}$$

These are essentially the canonical commutation relations of a system of as many degrees of freedom, as there are points p with $p^2 = M$, $p \in \mathbf{V}^+$, i.e., $(q-1)/2$ for $p^2 \neq 0$, and $q-1$ for $p^2 = 0$. The representation of the operators $A(p)$ satisfying (6.15) is therefore up to equivalence transformations uniquely determined. Let us describe the vectors Φ of the representation space $\mathbf{H}$ of the free fields by the Fock expansion

$$\Phi = \sum_{p^1,\ldots,p^n} \varphi_n(p^1, \cdots, p^n)A^+(p^1)\cdots A^+(p^n)\Omega. \tag{6.16}$$

Then the covariance condition

$$U(\Lambda, a)A(x)U^{-1}(\Lambda, a) = A(\Lambda x + a),$$

or $U(\Lambda, a)A(p)U^{-1}(\Lambda, a) = \exp(-2\pi i a\Lambda p/q)A(\Lambda p)$ determines the representation of ${}^q\mathbf{C}$ which is connected with the free fields,

$$U(\Lambda)A^+(p^1)\cdots A^+(p^n)\Omega = A^+(\Lambda p^1)\cdots A^+(\Lambda p^n)\Omega,$$

$$U(a)A^+(p^1)\cdots A^+(p^n)\Omega = \exp\left[(2\pi i/q)\sum_i (p^i a)\right]A^+(p^1)\cdots A^+(p^n)\Omega,$$

$$n = 0, 1, 2, \cdots . \qquad (6.17)$$

The characters of the translation group p include arbitrary sums $\sum_i p^i$ of p^i with $(p^i)^2 = M$, $p^i \in \mathbf{V}^+$. As a consequence of the cyclic character of the addition group of $GF(q)$, these p are not all contained in $\mathbf{V}^+$. The representation $U({}^q\mathbf{C}^\uparrow)$ of the free fields violates therefore the spectrum condition.

We believe the discussion of these two examples gives a fair impression about the possibilities of getting insight in the structure of quantum field theory by the study of group theoretical models.[13]

ACKNOWLEDGMENTS

The author wishes to thank Professor J. R. Oppenheimer for the hospitality extended to him at the Institute for Advanced Study, and the U. S. Educational Commission for a Fulbright travel grant. He is indebted to Professor A. S. Wightman for many stimulating and helpful discussions.

[13] Y. Ahmavaara, Ann. Acad. Sci. Fennicae **Ser. A, 95** (1962), studied independently field theories with finite geometries.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 2 FEBRUARY 1964

On the Derivation of Statistical Thermodynamics from Purely Phenomenological Principles

BENOIT MANDELBROT
Harvard University, Cambridge, Massachusetts and
I.B.M. Research Center, Yorktown Heights, New York

Szilard pointed out in 1925 that it is possible to base the foundations of statistical thermodynamics upon "phenomenological" principles, analogous to those of the non-statistical "classical" theory. This approach is discussed and developed.

1. INTRODUCTION

THERE is a widespread belief among physicists that statistical thermodynamics cannot be derived from "phenomenological" principles relative to macroscopic experiments, however idealized. Actually, in an admirable paper, Leo Szilard[1] has shown the two aspects to be quite compatible. Unfortunately, the paper is difficult, complicated and even confusing, and is unduly pessimistic about its own scope. But its general idea should be revived and further developed. Moreover, while Szilard's mathematical and conceptual framework is quite isolated within physics, we have earlier shown[2] that it leads itself readily to the introduction of certain powerful methods of probability theory and mathematical statistics as yet unexploited in theormodynamics. (Refs. 1 and 2 have been utilized in a recent treatment of the statistical thermodynamics of equilibrium[3]). The present paper will summarize[2] and will discuss several problems related to the place within phenomenological statistical thermodynamics of the second principle and of several alternative concepts of entropy.

It will be noted that, since a part of the laws of statistical mechanics can be obtained without any mechanical consideration, the term "statistical thermodynamics" will be used to designate the results of the statistical theory, without implying anything about the method used to derive them.

We insist, with the classical thermodynamicists, on a strict separation between the results linked to the zeroth and first principles, and those also requiring the second. For example, our zeroth principle differs sufficiently from that of Szilard to make the second principle unnecessary in order

[1] L. Szilard, Z. Physik **32**, 753 (1925). Do not confuse this reference with Z. Physik, **53**, 840 (1929).

[2] B. Mandelbrot, Ann. Math. Stat. **33**, 1021 (1962). For an early announcement of this result, see Compt. Rend. **243**, 1835 (1956).

[3] L. Tisza and P. M. Quay, Ann. Physics **25**, 48 (1963).

$$U(\Lambda)A^+(p^1)\cdots A^+(p^n)\Omega = A^+(\Lambda p^1)\cdots A^+(\Lambda p^n)\Omega,$$

$$U(a)A^+(p^1)\cdots A^+(p^n)\Omega = \exp\left[(2\pi i/q)\sum_i (p^i a)\right]A^+(p^1)\cdots A^+(p^n)\Omega,$$

$$n = 0, 1, 2, \cdots . \qquad (6.17)$$

The characters of the translation group p include arbitrary sums $\sum_i p^i$ of p^i with $(p^i)^2 = M$, $p^i \in \mathbf{V}^+$. As a consequence of the cyclic character of the addition group of $GF(q)$, these p are not all contained in $\mathbf{V}^+$. The representation $U({}^q\mathbf{C}^\uparrow)$ of the free fields violates therefore the spectrum condition.

We believe the discussion of these two examples gives a fair impression about the possibilities of getting insight in the structure of quantum field theory by the study of group theoretical models.[13]

ACKNOWLEDGMENTS

The author wishes to thank Professor J. R. Oppenheimer for the hospitality extended to him at the Institute for Advanced Study, and the U. S. Educational Commission for a Fulbright travel grant. He is indebted to Professor A. S. Wightman for many stimulating and helpful discussions.

[13] Y. Ahmavaara, Ann. Acad. Sci. Fennicae **Ser. A, 95** (1962), studied independently field theories with finite geometries.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 2 FEBRUARY 1964

On the Derivation of Statistical Thermodynamics from Purely Phenomenological Principles

BENOIT MANDELBROT
Harvard University, Cambridge, Massachusetts and
I.B.M. Research Center, Yorktown Heights, New York

Szilard pointed out in 1925 that it is possible to base the foundations of statistical thermodynamics upon "phenomenological" principles, analogous to those of the non-statistical "classical" theory. This approach is discussed and developed.

1. INTRODUCTION

THERE is a widespread belief among physicists that statistical thermodynamics cannot be derived from "phenomenological" principles relative to macroscopic experiments, however idealized. Actually, in an admirable paper, Leo Szilard[1] has shown the two aspects to be quite compatible. Unfortunately, the paper is difficult, complicated and even confusing, and is unduly pessimistic about its own scope. But its general idea should be revived and further developed. Moreover, while Szilard's mathematical and conceptual framework is quite isolated within physics, we have earlier shown[2] that it leads itself readily to the introduction of certain powerful methods of probability theory and mathematical statistics as yet unexploited in theormodynamics. (Refs. 1 and 2 have been utilized in a recent treatment of the statistical thermodynamics of equilibrium[3]). The present paper will summarize[2] and will discuss several problems related to the place within phenomenological statistical thermodynamics of the second principle and of several alternative concepts of entropy.

It will be noted that, since a part of the laws of statistical mechanics can be obtained without any mechanical consideration, the term "statistical thermodynamics" will be used to designate the results of the statistical theory, without implying anything about the method used to derive them.

We insist, with the classical thermodynamicists, on a strict separation between the results linked to the zeroth and first principles, and those also requiring the second. For example, our zeroth principle differs sufficiently from that of Szilard to make the second principle unnecessary in order

[1] L. Szilard, Z. Physik **32**, 753 (1925). Do not confuse this reference with Z. Physik, **53**, 840 (1929).

[2] B. Mandelbrot, Ann. Math. Stat. **33**, 1021 (1962). For an early announcement of this result, see Compt. Rend. **243**, 1835 (1956).

[3] L. Tisza and P. M. Quay, Ann. Physics **25**, 48 (1963).

to derive Gibbs' canonical distribution, and the concept of qualitative temperature.

Being parallel to statistical mechanics and thermodynamics, our method stresses more than is usual the parallelism existing between the two classical approaches. As a matter of fact, we shall stress the following: from the viewpoint of statistical mechanics, the physical statements upon which it is based may be considered as being "principles", and it happens that they can be associated one-to-one with the classical "principles" of phenomenological thermodynamics.

The random variable "energy" is designated by U and Gibbs' canonical distribution—as applied to isolated states—is written as

$$\exp(-\beta u)/Z(\beta).$$

However, from our purely phenomenological viewpoint, there is no sense in considering the states themselves; we rather have to consider surfaces of given energy, or sets made up of a number of such surfaces. As a result, we need a less precise form of the canonical law. Let us write $F(u \mid \beta)$ for the probability that the system be found in a state of energy equal to u or smaller. Then, designating by $G(u)$ the equivalent of the number of states of energy equal to u or smaller, we write Gibbs' distribution as follows:

$$dF(u \mid \beta) = d[\Pr(U \leq u)] = dG(u) \exp(-\beta u)/Z(\beta).$$

This law plays the central role in our considerations. It is more than a convenient auxiliary mathematical device.

We shall need to repeat certain well-known arguments, in order to insert them within our development; we hope that these repetitions will not be too bothersome.

2. DERIVATION OF GIBBS' CANONICAL DISTRIBUTION FROM "ZEROTH" AND "FIRST" PRINCIPLES: TERMINOLOGICAL COMMENTS UPON THE STATISTICAL-MECHANICAL APPROACH

It is well known that Gibbs' law can be obtained, as an asymptotic approximation, from the distribution of a small part of a large physical system. The latter distribution is itself deduced from the following assumptions:

(0) A system known to have an energy U contained between u and $u + du$ can be found with equal probabilities in either of $dG(u)$ "states", where $G(u)$ is a nondecreasing function of u.

(1a) Energy is the unique invariant of certain physical transformations, those resulting from "thermal interaction".

(1b) When thermal interactions are weak, the energy is an additive expression.

(1a) and (1b) have an obvious connection with the "first principle of thermodynamics" and (0) serves to first introduce the concept of temperature, and can therefore be associated with the "zeroth principle". That is, even if one succeds in deriving this principle from mechanics by some "ergodic" argument, such a derivation would be foreign to thermodynamics; from the viewpoint of that science, (0), (1a) and (1b) should rather be considered as "principles". The only purpose of the present section is to recommend such a nomenclature.

3. DERIVATION OF GIBBS' CANONICAL DISTRIBUTION FROM A "ZEROTH" AND A "FIRST" PRINCIPLE: A PURELY PHENOMENOLOGICAL APPROACH

Let us now summarize the results of our previous paper.[2] Its purpose is to show that, the zeroth and first principles can be rephrased so as to remain fully "phenomenological", while permitting a derivation of the canonical distribution, without any mechanical interpretation of $G(u)$.

The concept of a representative space, or "A-space".

By definition, the set of all possible outcomes of all the macroscopic measurements performed upon a system will constitute an "A-space" for that system. It will depend upon the physical structures of both the system and its environment and will not require or provide any microscopic mechanical substratum.

A stochastic concept of thermal equilibrium.

We assume that, when a system is in thermal equilibrium with its environment, the position of the corresponding "A-point" is random.

Conditioning.

A physical conditioning is any set of operations, which can be realized by purely macroscopic physical operations, and which impose some mathematical relation upon the A-point of a physical system.

For example, let the system S be made up of M parts S_m, and let the combinations of the energies U_m of these parts be the only macroscopically measurable characteristics of S. If S is in contact with an infinite environment, its A-point, of coordinates U_m, is any point of the positive hyperquadrant of M-dimensional Euclidean space. If the environment is finite and of total energy u, the A-space is the domain in which all $U_m \geq 0$

and $\sum U_m \leq u$. If S is "isolated", the A-space is the domain in which all $U_m \geq 0$ and $\sum U_m = u$.

The physical content of our "zeroth principle".

When appropriate mathematical restrictions are imposed upon the A-space and upon conditioning relations, the following two paragraphs will express a physical principle that can be referred to as being the "zeroth", although its relation with the statement that usually goes under this name may only transpire gradually.

The nature of "thermal equilibrium" is such that when a system is in equilibrium under a condition R', more stringent than an earlier condition R'', the distribution of the A-point no longer depends upon the parameters that have characterized R''.

Moreover, the equilibrium distribution under the physical condition R' may be obtained as follows: begin by the distribution relative to any less-strict condition R'', consider R' as being a mathematical relation and apply the rules of probability theory relative to conditioning.

We may consider this statement of the zeroth principle as referring to a kind of equilibrium between a system and its parts.

The first principle.

We shall express the first principle by the *unicity* portion of the statement made in Sec. 2. The *additivity* portion of that statement will rather be replaced by the following statement.

The auxiliary principle of the existence of heat reservoirs.

There exists a family of physical systems, to be called the "heat-reservoirs", which can simultaneously serve as environments for several distinct systems S_m, and are such that the corresponding phase points are statistically independent random variables.

Derivation of the canonical law.

It was shown in Ref. 2 that the canonical law can be derived from the combination of the zeroth, first, and auxiliary principle, with the help of certain known theorems.

Reference to statistical sufficiency.

The preceding statements are somewhat isolated in the context of physics, but they happen to be intimately related to a branch of mathematical statistics, called the theory of sufficiency. For technical details, we must refer to the textbooks of statistics,[4] but we may say that a probability distribution is said to possess a sufficient statistic if the following is true: Suppose that one wants to "estimate" β from a sample of M values u_m of U; this means that, given a finite sample, one wants to "guess" reasonably the value of the parameter of an infinite population from which the sample has been drawn. It is obvious that such guessing cannot be performed without some ambiguity and that even the best-trained people may disagree about the best method of estimation. Therefore, if one does not want to prejudice the statistician's procedure, it is in general necessary to furnish him with the complete collection of values of u_m. This is of course unwieldy, so that data collectors are very pleased when—irrespectively of their statistician's preferences—the data can be summarized exhaustively and without loss by giving a small number of functions $R_i(u_1, \cdots u_M)$. For example, if the parent population is Gaussian, the set of u_m can with no loss be summarized by $\sum u_m$ and $\sum (u_m)^2$. Such a set of numbers R_i is what is called a "sufficient statistic".

Clearly, such a set of functions exists if, and only if, the probability distribution of the random variables U_m, given the R_i, is independent of the value of the parameter β. This has a close counterpart in physics; indeed, the joint distribution of the energies of M systems, given their total energy, is known to be independent of the temperature of the heat reservoir with which the M systems used to be in contact. The purpose of our previous paper[2] was to exploit this general idea more fully and more rigorously, in particular in deriving the canonical distribution.

4. DERIVATION OF THE CONCEPT OF ENTROPY FROM A "SECOND PRINCIPLE": THE STATISTICAL-MECHANICAL APPROACH

This section purports to stress a conceptual viewpoint that completes that of Sec. 2, and to suggest the corresponding terminology. Once the canonical law has been derived, and the concept of heat has been introduced, it is well known that the path-integral of the expression "$\oint \beta\, dQ$" can be written as:

$$\Delta[\beta E(U)] - \oint \left\{ (\partial/\partial\beta) \log Z(\beta, \bar{V})\, d\beta - \beta \sum \left[\int dF(u \mid \beta, \bar{V})\, (\partial u/\partial V_k) \right] dV_k \right\},$$

[4] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, (John Wiley & Sons, Inc., New York, 1952). R. B. Hoag and A. T. Craig, *Introduction to Mathematical Statistics* (The Macmillan Company, New York, 1959).

where $\bar{V}$, of coordinates V_k, is the vector of external parameters of the system, and $(\partial u/\partial V_k)$ is the "adiabatic change" of the energy of a system when the external parameter V_k is varied alone (in Sec. 7, we shall return to this concept).

For this "$\oint \beta\, dQ$" to be path-independent, it is sufficient that the following be true:

$$(*) \quad -\beta \int dF(u \mid \beta, \bar{V})\, (\partial u/\partial V_k) = (\partial/\partial V_k) \log Z(\beta, \bar{V}).$$

If this is indeed the case, one can write

$$\oint \beta\, dQ = h_{\text{final state}} - h_{\text{initial state}},$$

where h is the entropy, defined as

$$h = \beta E(U) + \log Z(\beta, V).$$

In turn, condition (*) is satisfied if $G(u)$ satisfies the condition of "adiabatic invariance". This statement, seldom as much emphasized as in the writings of Paul Ehrenfest,[5] can of course be proven from mechanics. However, we think that—from the viewpoint of thermodynamics—it should rather be considered as being a basic "principle", namely the "second principle of statistical mechanics". (It is in fact somewhat surprising that, in view of the popularity of "conservation principles," the present one should have been discussed so little.) The simplest case is that of nondegenerately quantized energy, where $G(u)$ varies only for a denumerable number of values u_i of energy, and where $dG(u_i) = 1$ for all i. Let us recall that adiabatic invariance means that every energy level is a function of V over the same range of values of $\bar{V}$; no energy level is either created or annihilated by a change of volume. More generally, the function $G(u)$ is defined only up to multiplication by an arbitrary function of $\bar{V}$ (which vanishes from the distribution of u); this multiplier must be susceptible of being chosen in such a fashion that, when u^+ and u^- perform a free adiabatic transformation, $G(u^+) - G(u^-)$ remains invariant.

The above classical result has an important, but less well known, classical partial converse: Ehrenfest (Ref. 5, p. 347) has indeed shown that, in order that $\oint \beta\, dQ$ be path-independent and equal to Δh, adiabatic invariance is not only sufficient but also necessary.

[5] P. Ehrenfest, *Collected Scientific Papers* (Interscience Publishers, Inc., New York, 1959).

5. DERIVATION OF THE CONCEPT OF ENTROPY—AND OF ADIABATIC INVARIANCE—FROM THE USUAL SECOND PRINCIPLE OF THERMODYNAMICS

Of course, adiabatic invariance could also be added to the zeroth and first principles, as stated in Sec. 3. However, it would have little meaning because $G(u)$ has no mechanical interpretation there. It is therefore fortunate that our zeroth and first principles can be continued by any of the classical forms of the phenomenological second principle.

For example, we can postulate directly that the Pfaffian form dQ has an integrating multiplier B. This quantity cannot be a function of the external parameters, and can therefore only be a function of the "qualitative temperature" $1/\beta$ introduced by Gibbs' canonical law. [Or else, the existence of the multiplier B can be deduced from Carathéodory's principle, as applied to the mean values of the random variables in question: "in the neighborhood of any mean state of a system (as defined by β and by $\bar{V}$) there exist states that cannot be reached—on the average—by any transformation in which the initial and final state of the environment are indentical."]

Let us now show that adiabatic invariance is a necessary consequence of the fact that the integrating multiplier of dQ is a function $B(\beta)$; and, moreover, that this integrating multiplier must be β itself. This result will of course be stronger than Ehrenfest's classical statement referred to in Sec. 4, because we shall not require a priori that β itself be the integrating multiplier, and we shall not require that the integral be equal to the variation of h, as defined classically.

The smooth case.

First suppose that all the quantities of interest are continuous and derivable as often as required. Then, if the integrating multiplier is independent of $\bar{V}$, the following must hold for every couple of parameters V' and V'':

$$\frac{\partial}{\partial V''} \frac{\int \frac{\partial u}{\partial V'} G'(u) e^{-\beta u}\, du}{\int G'(u) e^{-\beta u}\, du} = \frac{\partial}{\partial V'} \frac{\int \frac{\partial u}{\partial V''} G'(u) e^{-\beta u}\, du}{\int G'(u) e^{-\beta u}\, du}.$$

This in turn leads to the requirement that

$$\frac{\int \frac{\partial u}{\partial V'} \frac{\partial G'}{\partial V''} e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du} - \frac{\int \frac{\partial u}{\partial V'} G' e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du}$$

$$\frac{\int \frac{\partial G'}{\partial V''} e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du} = \frac{\int \frac{\partial u}{\partial V''} \frac{\partial G'}{\partial V'} e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du} - \frac{\int \frac{\partial u}{\partial V''} G' e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du} \cdot \frac{\int \frac{\partial G'}{\partial V'} e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du}. \tag{1}$$

Every term of this identity is the canonical average of some function $f(u)$; if the systems are very large, one can write:

$$\frac{\int f(u) G' e^{-\beta u}\, du}{\int G' e^{-\beta u}\, du} \sim f[E(U)] + \tfrac{1}{2} f''[E(U)] D,$$

where D is the variance of U. Naturally, in order that (1) be satisfied, it is necessary (and we shall see that it is also sufficient) that it be satisfied up to terms of second order, for every value of $E(U)$, i.e., of β. This yields the requirement

$$\frac{\partial}{\partial u}\left[\frac{\partial \log G'}{\partial V''}\right] \cdot \frac{\partial}{\partial u}\left[\frac{\partial u}{\partial V'}\right] = \frac{\partial}{\partial u}\left[\frac{\partial \log G''}{\partial V'}\right] \cdot \frac{\partial}{\partial u}\left[\frac{\partial u}{\partial V''}\right],$$

so that the following ratio must be independent of β and of the V:

$$\frac{(\partial/\partial u)[\partial \log G'/\partial V]}{(\partial/\partial u)[\partial u/\partial V]} = \text{constant} = C^0.$$

This in turn requires the existence of a function $W^0(V)$ such that

$$\frac{\partial \log G'}{\partial V} = C^0 \frac{\partial u}{\partial V} + W^0(\bar{V}).$$

The term in C^0 can be eliminated by a trivial renormalization in which G' is multiplied by $\exp(-C^0 u)$ and the origin of β is translated by C^0.

The term in $W^0(\bar{V})$ amounts to the multiplication by $\exp[\int W^0(\bar{V})\, dV]$ of both the numerator and the denumerator of the canonical distribution. Hence, this term vanishes.

To sum up, the second principle requires that—up to trivial renormalization—the function G' be independent of V, which means that it is adiabatically invariant. We know already that this condition is also sufficient to prove the second law, and that the integrating multiplier of dQ is in that case equal to β.

6. PERFECT GASES

In order to measure the β derived in the preceding sections, an instrument is needed. The usual procedure is to use a perfect gas thermometer, which is a physical system such that

$$\partial u/\partial V = u f(V).$$

7. GENERALIZATION OF THE SECOND LAW TO INDIVIDUAL CANONICAL SYSTEMS: THE RANDOM FORM OF THE CONCEPTS OF ENTROPY AND OF HEAT

Let us return to the situation at the end of Sec.3, where the canonical law has been established from either set of zeroth and first principles. The concept of "heat", as used in Secs. 4 and 5, was assumed to have been obtained by the usual method, which averages everything very early over a canonical distribution.

As a result, both Secs. 4 and 5 involve the *nonrandom* entropy

$$h = \beta E(U) + \log Z(\beta).$$

However, one frequently wishes to interpret "entropy" by the *random* expression:

$$h = \beta U + \log Z(\beta).$$

For example, h is necessary to give meaning to "Boltzmann's principle", that

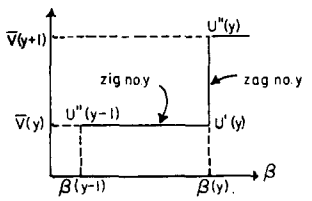
"entropy = $-$log (probability of a state)."

This principle is in turn necessary to generalize the concept of entropy beyond its original context relative to equilibrium (that is, to the canonical law). Similarly, the methods based upon averaging raise difficulties concerning heat. In the verbal explanations of what is thermodynamics, one states that "heat" is a noncontrollable and presumably random portion of energy, while "work" is controllable and presumably nonrandom. However, by defining $\oint \beta E(dQ)$ with the sign E, that is by defining it for ensembles and not for systems, one immediately cancels out this distinction.

The above remarks describe the motivation of this section. We shall study the expression $\oint \beta\, dQ$ for an individual system, *without* averaging from the outset. "Heat" will remain a random quantity, but "work" will not; however, the averaging involved in the concept of "work" will be based on different grounds than is usual. The second law will be shown to apply to the *unaveraged* entropy and the Boltzmann's principle will be fully meaningful. Part of our discussion is closely related to that of Refs. 1 and 6.

[6] L. Rosenfeld, Physica **27**, 67 (1961).

Let us consider a physical system for which the path, that is, the sequence of its values of β and of the V_k, is an alternation of "zigs" during which β alone varies, and of "zags" during which β remains invariant, and let us parametrize the path by an index y, taking successive integral values at the points where one passes from one zig-zag to the next. The following graph summarizes various definitions.



Variation of temperature at fixed volume.

During a "zig", let our system be successively put in contact with a series of heat reservoirs, of slowly varying β, and let each contact be very long. The energy exchanged during each contact is of course a random variable and it will be interpreted as "heat". Clearly, the averaged "path-integral" $\oint \beta\, dE(Q) = \oint \beta\, dE(U)$ is trivially path-independent. But the unaveraged $\oint \beta\, dU$ might have depended upon intermediate random elements. Our main purpose here will be to show that such is not the case. For that, begin by dividing the "zig" into steps corresponding to equal increments of β. Then

$$\oint \beta\, dU = \sum \beta(x)[U(x) - U(x - dx)]$$
$$= U'(y)\beta(y) - U''(y - 1)\beta(y - 1 - dx) - \sum U(x)[\beta(x) - \beta(x - 1)],$$

the sum on the preceding line being carried over all the values of x except the first and last one. Let all the steps become infinitesimally small, while the contact with each successive heat reservoir remains sufficiently slow for the $U(x)$ to remain independent random variables. An easy application of a form of the strong law of large numbers shows that, with probability one,

$$\oint \beta\, dU \Rightarrow U'(y)\beta(y) - U''(y - 1)\beta(y - 1) - \oint E(U \mid \beta)\, d\beta,$$

the averaging being due to a theorem and not to the fact that one has decided *a priori* to take account of ensemble averages exclusively.

Free and normal adiabatic changes of volume; the pressure.

Similarly, during the yth zag, the system acquires the random energy $U''(y) - U'(y) = U[\beta(y), \bar{V}(y + 1)] - U[\beta(y), \bar{V}(y)]$. But this is not the whole story, since only part of this energy was contributed by heat reservoirs, the rest being contributed by the outside forces that changed $\bar{V}$. These two parts are respectively called "heat" and "work". Suppose that each zag is further subdivided into small steps, between which the system is put back into contact with a heat reservoir of temperature $1/\beta(y)$. Each minute change of volume requires an addition of energy which depends upon the initial and final volume and upon the initial energy, but not upon earlier values of energy; in particular, it is independent of the earlier temperature, and it can be designated by $(\partial u/\partial V_k)dV_k$. Such a change of volume is called "free adiabatic" and it does not in general preserve the canonical distribution: That is, let the same dV_k be applied to the members of a family of systems with a canonical energy; the energy of the perturbed systems needs not be canonical. Therefore, one replaces the "free adiabatic" changes by "normal adiabatic" changes, in which contact is recurrently reestablished with the heat reservoir of temperature $1/\beta$. Thanks to this canonical averaging, the exchange of energy due to changing V_k becomes independent of the intermediate energies of the system. By a further easy application of a variant of the strong law of large numbers, one finds that (with probability one) the exchanges of energy directly traceable to the changes dV_k are

$$\sum \oint \int dF(u \mid \beta, \bar{V})\, (\partial u/\partial V_k)\, dV_k = \sum p_k\, dV_k = dW.$$

The p_k defined by this equality are called "generalized pressures", and dW defines the concept of "work", an asymptotically nonrandom part of the energy communicated to the system from the outside.

Consider now the rest of the energy exchanges, that is the exchanges during the successive intermittent contacts with the heat reservoir. This part is random and uncontrollable and it is natural to identify it as "heat"

Note that the above argument derives the concept of pressure and of work for individual canonical systems going through a series of small transformations. The usual verbal distinction

between the "disorderly" character of heat energy and the "orderly" character of work is formalized as the difference between random and nonrandom.

Now, putting the zigs and the zags together, one finds that

$$\oint \beta\, dQ = U'(F)\beta(F) - U''(0)\beta(0) - \oint \left\{ E(U \mid \beta, \bar{V})\, d\beta + \beta \sum_k \left[\int dF(u \mid \beta, \bar{V})\, (\partial u/\partial V_k) \right] \right\} dV_k.$$

Conclusion.

This leads us to the point where the usual discussion begins. The last integral is path-invariant if and only if its expected value is path-independent, If so,

$$dQ = H_{\text{final}} - H_{\text{initial}},$$

where H is the random entropy

$$H = \beta U + \log Z(\beta).$$

(This shows the invalidity of Khinchin's[7] assertion, that H does not satisfy the "second law") Note that, when the path is closed, the entropy change does not necessarily vanish, because the initial and final values of U need not be equal.

Noncanonical systems, for which an entropy can be defined without using additional axioms.

Szilard[1] has shown that, when entropy is written as $\sum p_{\text{state}} \log p_{\text{state}}$, it can also apply to systems obtained from canonical systems by a free adiabatic transformation.

8. INCREASE PROPERTIES OF THE CANONICAL ENTROPY

We want to stress that the concept of entropy requires the second principle, which is unnecessary to derive the canonical law. This is why we insist upon avoiding any derivation of the canonical law that uses anything resembling entropy (or Shannon's "information"). More precisely, even if the use and the maximization of "log W" or "$-p \log p$" is motivated on some axiomatic grounds, the maximum values of these expressions *cannot* be identified with entropy unless one introduces some additional statement equivalent to the second principle. But, if a second principle is used, one can derive the canonical law and the form of entropy. Let us show that one can also derive some "increasing" properties of this entropy.

Irreversible changes of temperature.

Returning to the expression for $\oint \beta\, dU$; before the steps of β are made infinitely small $\oint \beta\, dU$ depends upon intermediate energy exchanges and can be greater or smaller than its limit for continuously varying β. But, considering expected values, one has

$$E\left[\sum \beta \Delta U - \int \beta\, dU \right] = \int E(U)\, d\beta - \sum E[U(x)][\beta(x) - \beta(x-1)],$$

where $E[U(\beta)]$ is a decreasing function of β. Then, if β increases the above expression is the difference between the integral and a *lower* Riemann sum of a decreasing function. If β decreases, the above expression is the difference between an *upper* Riemann sum and an integral. Hence, for a closed loop, the expected value of $\sum \beta \Delta U - \int \beta\, dU$ is the difference between an upper and lower Riemann sum and it *is positive.*

Irreversible variation of volume.

The same argument holds, assuming that pressure is a decreasing function of every V_k.

9. GENERALIZATION OF THE SCOPE OF THERMODYNAMICS TO SYSTEMS WHICH ARE NOT IN EQUILIBRIUM AND CANONICAL

From the viewpoint of the core of results applicable to macroscopic systems, the results of the preceding sections are not a substantial improvement over the results of Sec. 5, relative to means. Moreover, for large systems, the fluctuations due to contact with a heat reservoir are so small in relative value, that the canonical theory gives acceptable predictions concerning most characteristics of isolated systems of fixed energy. However, to be able later to generalize thermodynamics beyond the results linked with the basic principles, it is necessary first to explicitly define a temperature and an entropy for isolated systems. For that, one must distinguish between work and heat. The path of the system, as sequence of values of $(u, V_1, \cdots V_K)$, will again be approximated by zig-zags. The "zigs" are free adiabatic and the energy change is the "work" $\sum (\partial u/\partial V_k) dV_k$. The "zags" correspond to an energy addition of $du - \sum (\partial u/\partial V_k) dV_k$, which is all heat. Temperature would be defined as a function of the V_k and of u, constituting an integrating divisor for heat; un-

[7] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, Inc., New York, 1949).

fortunately, the existence of such a multiplier requires a condition upon the expressions $(\partial u/\partial V_k)$, which was not necessary in the canonical case and has no reason of being generally satisfied. Hence, this method of generalizing the scope of thermodynamical concepts fails.

Of course, the definition of temperature for isolated systems is usually approached very differently. This temperature was discussed in reference 2. One knows that it loses all meaning when the energy u is known, but it is found convenient to give the same name to either one of a variety of functions of u, which converge for very large systems but differ for small ones. In reference 2 we have analyzed these functions and have noted that the choice of a definition of temperature bears the closest connexions with the basic problem of mathematical statistics: knowing that the quantity u is a sample value of a random quantity U, and that the distribution of U depends upon a parameter β, "estimate" the value of β from the value of u. It is intuitively true, and is confirmed by the theory, that estimation is a kind of guessing and is indeterminate except if one has a large number of sample values. But both the practice and the apparent intent of the operation "to define a temperature for an isolated thermodynamical system" can be interpreted as really meaning "to estimate the temperature of a heat reservoir, with which the isolated system should be presumed to have been in contact". For large systems, many estimates are equally good.

Let us now examine entropy. For isolated systems, the distinction between work and heat lacks here the clarity which it had in the canonical case. In order to define heat or work, one must define pressure and this is also done by an estimation procedure. By choosing appropriately a set of definitions of temperature and of pressure, one can arrange for the integral $\oint \beta\, dq$ to be path-invariant and thus define H. As a result, the number of useful definitions of entropy will be at least as large as the number of useful temperatures-in-isolation. The best known groups of definitions are the following:

Boltzmann's definitions. Temperature $1/\hat{\beta}_b$ is such that

$$u = E[U(\hat{\beta}_b)] = -\partial \log Z(\hat{\beta}_b)/\partial\hat{\beta}_b.$$

The pressures are $-(1/\hat{\beta}_b)\ \partial \log Z(\beta_b)/\partial V_k$, and entropy is

$$\hat{h}_b = \hat{\beta}_b u + \log Z(\hat{\beta}_b) = \min_{\beta} \{\beta u + \log Z(\beta)\}.$$

Gibbs' differential definitions. Temperature is given by

$$\hat{\beta}_{gd} = (\partial/\partial u) \log [dG(u)].$$

The pressures are $(\partial/\partial V_k) \log [dG(u)]$, and entropy is

$$\hat{h}_{gd} = \log [dG(u)]$$

Gibbs' integral definitions. One replaces $dG(u)$ in the differential definitions by $G(u)$.

Note that nothing can be said in general concerning the sign of the difference between the Boltzmann and the Gibbs' entropies, the reason being that $\log [dG(u)]$ depends heavily upon the local regularity properties of $G(u)$, while the Boltzmann entropy does not.

Generalization of entropy. The usual generalizations of the concept of entropy are based upon a formal broadening of the conditions of applicability of either of the two relations:—Boltzmann's: "entropy = min $\{-\log$ [Pr (each "state" of energy u)]$\}$"—Gibbs' differential: "entropy = log (number of "states" of energy u)".

Either of these methods (and presumably other methods as well) will lead to a generalized theory. But the choice between them is largely arbitrary and hence controversial (they represent two methods of describing the role of the observer and of "information" in thermodynamics). One knows that, adding the right maximization criteria, either of the generalized definitions can replace the "zeroth principle" either in the classical sense or in our phenomenological sense. Hence, one may say that the generalization of thermodynamics hinges upon the zeroth principle.

Asymptotic Properties of the Electromagnetic Field

J. N. GOLDBERG* AND R. P. KERR†
Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio
(Received 30 August 1963)

From the integral form of the general solution for the retarded electromagnetic field of a localized charge–current distribution, the asymptotic field is shown to have the behavior $F_{\mu\nu} = N_{\mu\nu}/R + III_{\mu\nu}/R^2 + {}_2J_{\mu\nu}/R^3$, where the coefficients satisfy $N_{\mu\nu}k^\nu = 0$, $III_{\mu\nu}k^\nu = Ak_\mu$, and $k_\mu k^\nu = 0$. The remainder ${}_2J_{\mu\nu}$ is shown to be bounded by using the second-mean-value theorem. Thus the algebraically special character of the asymptotic electromagnetic field is exhibited.

1. INTRODUCTION

IN studying gravitational radiation, recent work has been primarily concerned with the asymptotic field.[1–3] The reason for this, clearly, is that a general solution of the nonlinear Einstein field equations does not yet exist. With appropriate assumptions about "asymptotic flatness", however, the behavior of the distant field can be studied and the properties of gravitational radiation in the wave zone may be examined. Although one finds that an asymptotically flat metric (gravitational field) allows an outgoing (incoming) radiation field, this result has not yet been connected with the sources of the field.

Because of the simplicity of its structure, the electromagnetic field has been used as a model for studying the formal as well as the physical properties of the gravitational field. The purpose of this paper is to look at the retarded electromagnetic field from the point of view used in discussing the asymptotic Riemann tensor.[1,4] However, instead of relying on the field equations to propagate the field components along the assumed characteristic null directions,[5] the electromagnetic field from an isolated charge–current distribution is expressed explicitly in integral form. Then the asymptotic existence of hypersurface-orthogonal characteristic null directions can be shown explicitly. Specifically, the following will be proven in the next section:

Theorem. The retarded electromagnetic field from an isolated, but extended, charge–current distribution has the asymptotic form

$$F_{\mu\nu} = N_{\mu\nu}/R + III_{\mu\nu}/R^2 + O(R^{-3}), \qquad (1.1)$$

where R is a suitably normalized affine parameter along the null rays, k^μ, and the coefficients satisfy the algebraic conditions

$$N_{\mu\nu}k^\nu = 0, \quad III_{\mu\nu}k^\nu = Ak_\mu, \quad k_\mu k^\mu = 0. \qquad (1.2)$$

Also, k_μ is hypersurface-orthogonal,

$$k_\mu = -u_{,\mu}. \qquad (1.3)$$

The null surfaces of constant u will be defined later.

These results agree with previous calculations using a multipole expansion for the field[6] as well as calculations based on Sach's outgoing radiation condition in general relativity.[1,4] The principal advantage of the present work is that only an analysis of the asymptotic field is used, and no series expansions are involved. Also, we shall find explicit expressions for the coefficients in Eq. (1.1) as integrals over the intersection of certain null planes with the support of the charge–current distribution.

2. PROOF OF THE THEOREM

Maxwell's equations for the electromagnetic field due to a localized charge–current density $j^\mu(x)$ are[7]

$$F^{\mu\nu}{}_{,\nu} = 4\pi j^\mu, \qquad (2.1a)$$

$$F_{[\mu\nu,\rho]} = 0, \qquad (2.1b)$$

where $F_{\mu\nu}$ is the skew-symmetric tensor representing the electromagnetic field, and the square brackets mean complete antisymmetrization for all indices enclosed. That $j^\mu(x)$ is localized means that the 4-current vanishes outside a timelike world tube of finite diameter.

Introducing the vector potential $A_\mu(x)$ such that

$$F_{\mu\nu}(x) = 2A_{[\mu,\nu]}(x),$$

* Present address: Syracuse University, Syracuse, N. Y.
† Present address: University of Texas, Austin, Texas.
[1] R. K. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).
[2] H. Bondi, M. G. J. van den Berg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[3] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).
[4] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[5] J. L. Synge, *Relativity: The Special Theory* (North-Holland Publishing Company, Amsterdam, 1956).
[6] R. K. Sachs, Chapter in *Recent Developments in General Relativity* (Panstwowe Wydawnictwo Naukowe, PWN-Polish Scientific Publishers, Warsaw, and Pergamon Press, London, 1962).
[7] Indices are raised and lowered with the Minkowski metric $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$, (1, 1, 1, −1).

one finds the solution[5,8]

$$A_\mu(x) = \int_N j_\mu(z)\,\frac{d^3z}{|\mathbf{x}-\mathbf{z}|} = \int_N j_\mu(z)\,d\omega.$$

The domain of integration N is the intersections of the past null cone from the field point x and the timelike world tube containing the current. Synge calls $d\omega$ the absolute 2-content of a 3-cell on a null cone.

In what follows, the assumption is made that j^μ is everywhere continuous together with its first three derivatives; that is, it is of class C^3. This assumption could be weakened to allow j^μ to be discontinuous on a finite number of timelike hypersurfaces, however, the increase in computational difficulty in treating the discontinuities would obscure the result. It is sufficient to note that the results go through in this case too. With the above assumption on the continuity and differentiability of j^μ, one obtains immediately

$$F_{\mu\nu}(x) = 2\int_N j_{[\mu,\nu]}(z)\,d\omega. \tag{2.2}$$

In order to examine the geometric structure of the asymptotic electromagnetic field, it will be necessary to give an exact definition for what we mean by a localized charge–current distribution. It is not sufficient that the intersection of all spacelike planes with the support of the source vector should be of finite radius, since this would permit null geodesics which never left the charge distribution. We shall assume that there exists a timelike world line $\xi^\mu(s)$ and a function $a(s)$, such that

$$j_\mu(x) = 0 \text{ for all } x \text{ such that}$$

$$(x^\mu - \xi^\mu(s))(x_\mu - \xi_\mu(s)) \geq \max\{0,\, a^2(s) - [v_\alpha(x^\alpha - \xi^\alpha(s))]^2\}. \tag{L}$$

In the instantaneous rest system of the world line at time s, (i.e., $\dot\xi^\mu = v^\mu \overset{*}{=} \delta^\mu_0$)[9], this condition says that the charge–current vector is zero in the region extending to spatial infinity which is bounded by the forward and backward null cones from $\xi^\mu(s)$, and the cylinder of radius $a(s)$ defined by

$$|\mathbf{x} - \boldsymbol{\xi}(s)|. \tag{L'}$$

Two scalar functions $u(x)$ and $R(x)$, and the null-vector field $k_\mu(x)$ are defined by the following equations:

$$(x^\mu - \xi^\mu(u))(x_\mu - \xi_\mu(u)) = 0, \quad x^0 - \xi^0(u) \geq 0, \tag{2.3}$$

$$x^\mu = \xi^\mu(u) + Rk^\mu. \tag{2.4}$$

The null vector k_μ is normalized by the equation

$$v^\mu k_\mu = -1, \tag{2.5}$$

so that $R(x)$ is given by

$$R = -v_\mu(x^\mu - \xi^\mu(u)). \tag{2.6}$$

It will be observed that the surface of constant u is the future null cone with vertex at $\xi^\mu(u)$. By differentiation of Eq. (2.3) with respect to x^μ, and using Eq. (2.5) and (2.6), Eq. (1.3) is obtained.

In the expression for the electromagnetic field [Eq. (2.2)], the point z is restricted to lie on the backward null cone from x^μ, and so z^0 can be expressed as a function of the 3-vector $\mathbf{z}$,

$$z^0 = x^0 - |\mathbf{x} - \mathbf{z}| = Rk^0 + \xi^0 - |R\mathbf{k} - \boldsymbol{\zeta}|,$$

where for convenience we have introduced $\zeta^\rho = z^\rho - \xi^\rho(u)$. At the time u, in the instantaneous rest system of the particle, we have the following relations:

$$v^\mu(u) \overset{*}{=} \delta^\mu_0, \quad k^0 \overset{*}{=} 1, \quad R \overset{*}{=} |\mathbf{x} - \boldsymbol{\xi}|, \quad |\mathbf{k}| \overset{*}{=} 1.$$

If ρ is defined by

$$\rho = R^{-1},$$

then z^0 can be written as

$$z^0 = \xi^0 + \mathbf{k}\cdot\boldsymbol{\zeta} + \varphi(\rho), \tag{2.7}$$

where

$$\varphi(\rho) \overset{*}{=} \rho^{-1}[1 - \rho\mathbf{k}\cdot\boldsymbol{\zeta} - (1 - 2\rho\mathbf{k}\cdot\boldsymbol{\zeta} + \rho^2|\boldsymbol{\zeta}|^2)^{\frac{1}{2}}], \quad \varphi(0) = 0. \tag{2.8}$$

The function $\varphi(\rho)$ is an analytic function of ρ in the region where

$$\rho\,|\boldsymbol{\zeta}| < 1.$$

Since $\mathbf{z}$ is restricted to lie inside the sphere of radius $a(u)$, and $\mathbf{x}$ lies outside this sphere, the above condition is always satisfied. We shall also define another analytic function, $g(\rho)$, by

$$g(\rho) \overset{*}{=} [1 - 2\rho\mathbf{k}\cdot\boldsymbol{\zeta} + \rho^2|\boldsymbol{\zeta}|^2]^{-\frac{1}{2}}. \tag{2.9}$$

The electromagnetic field tensor can now be written as

$$F_{\mu\nu}(\xi^\mu(u) + Rk^\mu) \overset{*}{=} \frac{2}{R}\int j_{[\mu,\nu]}(\mathbf{z}, \xi^0 + \mathbf{k}\cdot\boldsymbol{\zeta} + \varphi(\rho))g(\rho)\,d^3z. \tag{2.10}$$

[8] The bold face symbols represent 3-vectors formed by the projection of the corresponding 4-vectors into the spacelike hypersurface $x^0 =$ constant. z is an arbitrary point inside the support of the current distribution. x is an arbitrary field point outside the tube.

[9] Equations which have an asterisk above the equality (or inequality) symbols are true only in the instantaneous rest system of the world line.

From the definition of ρ given above, it is clear that the asymptotic region is defined by the limit as ρ approaches zero. In this limit, the defining equation for z^0 [Eq. (2.7)] becomes

$$z^0 \stackrel{*}{=} \xi^0 + \mathbf{k}\cdot\boldsymbol{\zeta},$$

which is the equation of the null plane P containing the null line joining x to the world line. If the integrand of Eq. (2.8) were analytic in ρ, it might be possible to express $F_{\mu\nu}$ as a power series in ρ, with coefficients given as integrals over P, rather than N. However, in most physical situations the assumption of analyticity is not justified. Fortunately, the theorem as stated in Sec. 1 requires knowing only the first two terms of this expansion, and the remainder can be evaluated, using the second-mean-value theorem. This analysis is somewhat laborious, and therefore is carried out in the appendix. The result obtained there is that, in the rest system,

$$F_{\mu\nu} \stackrel{*}{=} \frac{N_{\mu\nu}}{R} + \frac{III_{\mu\nu}}{R^2} + \frac{{}_2J_{\mu\nu}}{R^3}, \tag{2.11}$$

where the remainder ${}_2J_{\mu\nu}$ is bounded by the following expression:

$$|{}_2J_{\mu\nu}| \leq \tfrac{4}{35}\Pi[14M_0 + 35M_1 a + 20M_2 a^2][a^5/(1-\rho a)^3]. \tag{2.12}$$

Here, the $M_i(u)$ are the maximum values assumed by the ith derivatives of $j_{\mu,\nu}$ between the two planes,

$$v_\alpha \zeta^\alpha = \pm a.$$

These maxima exist for $i \leq 3$, since the current vector was assumed to be three times *continuously* differentiable. The coefficients $N_{\mu\nu}$ and $III_{\mu\nu}$ are given by

$$N_{\mu\nu} \stackrel{*}{=} 2\int_P j_{[\mu,\nu]}\, d^3z, \tag{2.13a}$$

$$III_{\mu\nu} \stackrel{*}{=} 2\int_P \{j_{[\mu,\nu]}\mathbf{k}\cdot\boldsymbol{\zeta} + \tfrac{1}{2}j_{[\mu,\nu]0}[(\mathbf{k}\cdot\boldsymbol{\zeta})^2 - |\boldsymbol{\zeta}|^2]\}\, d^3z. \tag{2.14a}$$

Up to this point, the calculations have all been carried out in the rest system of the particle at time u, there being no Lorentz frame in which $v^\mu(\mu) \stackrel{*}{=} \delta^\mu_0$ *for all* u, when the world line is curved. However, the coefficients $N_{\mu\nu}$ and $III_{\mu\nu}$ may be written covariantly as

$$N_{\mu\nu} = 2\int_P j_{[\mu,\nu]}\, d\Omega, \tag{2.13}$$

$$III_{\mu\nu} = -2\int_P \{v_\rho \zeta^\rho j_{[\mu,\nu]} + \tfrac{1}{2}j_{[\mu,\nu]\alpha}v^2\zeta_\rho\zeta^\rho\}\, d\Omega, \tag{2.14}$$

where $d\Omega$ is an invariant measure on the plane P:

$$d\Omega = (1/3!)\,\epsilon_{\alpha\beta\gamma\delta}v^\alpha d_{(1)}x^\beta d_{(2)}x^\gamma d_{(3)}x^\delta.$$

Clearly, in the instantaneous rest system, Eqs. (2.13) and (2.14) reduce to the expressions given earlier.

To establish the theorem stated in Sec. 1, we must now determine the algebraic properties of the coefficients in Eq. (1.1). For this purpose, we shall need the following relations [*Note added in proof.* In Eq. (2.15) the total differential is to be understood as

$$dF(\mathbf{z}, z^0, u) = (\partial F/\partial z^\mu)\, dz^\mu + (\partial F/\partial u)\, du.$$

For clarity we define

$$[F(\mathbf{z}, z^0, u)] = (F(\mathbf{z}, z^0, u))z^0 = \xi^0 + \tilde{\mathbf{k}}\cdot\boldsymbol{\zeta}, \quad \tilde{\mathbf{k}} = \mathbf{k}/k^0.$$

Then,

$$[F_{,\mu}] = [F]_{,s} - \tilde{k}_\mu[F_{,0}]. \tag{*}$$

Using $D\int d\Omega = 0$, we have

$$D\int [F]\, d\Omega = \int \left\{(k^0)^{-1}[F_{,0}] + \left[\frac{\partial}{\partial u}F\right]\right\} d\Omega. \tag{**}$$

Equation (2.15) now follows by integrating (*), noting that the divergence term vanishes by Gauss' theorem, and substituting for the remaining term from (**)]:

$$\int_P F_{,\mu}\, d\Omega = -k_\mu\left\{D\int_P F\, d\Omega - \int_P \frac{\partial F}{\partial u}\, d\Omega\right\}, \tag{2.15}$$

where F may be a function of u, as well as z^μ, and the differential operator D is given by

$$D = \partial/\partial u + \kappa, \qquad \kappa = \dot{v}_\mu k^\mu.$$

Applying the above equation to

$$F = j_\mu \prod_{i=1}^{n} \zeta^{\alpha_i},$$

and using the conservation law for charge, $j^\mu_{\ ,\mu} = 0$, we obtain

$$k_\mu D\mathcal{J}^{\mu:\alpha_1\cdots\alpha_n} + n[k_\mu \mathcal{J}^{\mu:(\alpha_2\cdots\alpha_n}v^{\alpha_1)} + \mathcal{J}^{(\alpha_1:\alpha_2\cdots\alpha_n)}] = 0, \tag{2.16}$$

where the moments are defined by

$$\mathcal{J}^{\mu:\alpha_1\cdots\alpha_n} = \int_P j^\mu \prod_{i=1}^{n} \zeta^{\alpha_i}\, d\Omega,$$

$$\mathcal{J}^{\mu:\alpha_1\cdots\alpha_n}k_{\alpha_1} = 0.$$

Using Eq. (2.15), we find from Eq. (2.13)

$$N_{\mu\nu} = 2k_{[\mu}D\mathcal{J}_{\nu]}, \tag{2.17a}$$

which clearly satisfies

$$N_{\mu\nu}k^\nu = 0, \tag{2.17b}$$

because of Eq. (2.15), with n equal to zero. Similarly, from Eq. (2.13),

$$III_{\mu\nu} = 2k_{[\nu}B_{\mu]} + 2v_{[\nu}\mathcal{J}_{\mu]} + 2D\mathcal{J}_{[\mu:\nu]},$$
$$B_\mu = \tfrac{1}{2}(\kappa + D)D\mathcal{J}^\rho_{\mu:\rho} + v_\rho(\kappa + 2D)\mathcal{J}_\mu{}^\rho_{:} + \dot{v}_\rho\mathcal{J}_\mu{}^\rho_{:} - \mathcal{J}_\mu. \qquad (2.18a)$$

Again using Eq. (2.15) with $n = 1, 2$, we obtain

$$III_{\mu\nu}k^\nu = Ak_\mu, \qquad A = v^\mu\mathcal{J}_\mu + D\mathcal{J}_\mu{}^\mu_{:} \qquad (2.18b)$$

Eqs. (2.11), (2.12), (2.17), and (2.18), together with the definition of k^μ given in Eqs. (2.3) through (2.5), complete the proof of the theorem.

3. CONCLUSION

We have shown explicitly how the classification of the electromagnetic field shows up in the asymptotic outgoing radiation field. According to Synge,[5] the real eigenvectors of the electromagnetic field are necessarily null, and a null field is characterized by having a zero eigenvalue. From the theorem, it is clear that although in general the rather arbitrarily chosen null vector k^μ is not an eigenvector, asymptotically it becomes one. The leading behavior of the asymptotic field is that of a null electromagnetic field. Furthermore, this behavior is explicitly tied to a localized charge–current distribution, but is independent of the velocity vector v^μ. If we expand around a different world line $\xi'^\mu(s)$, we find

$$N'^{\mu\nu}/R' = N^{\mu\nu}/R + O(R^{-2}).$$

The calculation presented here strongly suggests that a similar relationship exists between the gravitational field and its sources. However, until we are able to find a satisfactory relationship between the sources and the distant field, the asymptotic results obtained by Bondi[2] and Newman[3] are undoubtedly the best we can hope for.

This problem arose in the course of a discussion with Professor J. L. Synge.

APPENDIX

For the purpose of establishing an upper bound for the remainder in Eq. (2.11), we first observe that if $j_{[\mu,\nu]}$ is n times continuously differentiable, then so is $\Phi_{\mu\nu}(\rho)$, where[10]

$$\Phi_{\mu\nu}(\rho) = 2j_{[\mu,\nu]}(\mathbf{z}, \xi^0 + \mathbf{k}\cdot\boldsymbol{\zeta} + \varphi(\rho))g(\rho), \qquad (A1)$$

and $\varphi(\rho)$ and $g(\rho)$ are defined in Eqs. (2.8) and (2.9). From the nth-mean-value theorem, we then have

$$F_{\mu\nu}(\xi^\mu + Rk^\mu) \stackrel{*}{=} \int \Phi_{\mu\nu}(\rho)\, d^3z = \sum_{s=0}^{n-1} \frac{1}{s!\,R^{s+1}} \int \Phi^{(s)}_{\mu\nu}(0)\, d^3z + \frac{1}{R^{n+1}}\, {}_nJ_{\mu\nu}, \qquad (A2)$$

where the remainder is given by

$${}_nJ_{\mu\nu} = \frac{1}{n!}\int \Phi^{(n)}_{\mu\nu}(\sigma)\, d^3z; \qquad 0 < \sigma < \rho, \qquad (A3)$$

and so we find a bound for ${}_nJ_{\mu\nu}$ if we can find a bound for $\Phi^{(n)}_{\mu\nu}(\sigma)$. It is easily shown that all points of the set

$$\{z, \xi^0 + \mathbf{k}\cdot\boldsymbol{\zeta} + \varphi(\sigma)\}$$

lie between the forward and backward null cones from the point $\xi^\mu(u)$, and therefore, from the localization condition (L), the domain of integration in Eqs. (A2) and (A3) is the sphere S, defined by

$$S = |\zeta| \leq a(u).$$

Since the region R, bounded by the forward and backward null cones from ξ^μ, and the sphere S is compact, there exist constants $M_r(u)$, (dependent on u) such that, for all (μ, ν),

$$\left|\left(\frac{\partial}{\partial\varphi}\right)^r j_{[\mu,\nu]}(\varphi(\sigma))\right| = \left|\left(\frac{\partial}{\partial z^0}\right)^r j^{(s)}_{[\mu,\nu]}\right| \leq M_r(u); \qquad r \leq n. \qquad (A4)$$

It is clear that

$$\Phi^{(n)}_{\mu\nu}(\sigma) = 2\sum_{r=0}^{n} j^{(r)}_{[\mu,\nu]}\mathfrak{F}^n_r(\varphi^{(s+1)}(\sigma), g^{(t)}(\sigma); s, t \leq n), \qquad (A5)$$

where the $\mathfrak{F}^n_r$ are certain polynomials, *with positive coefficients*, in the $\varphi^{(s+1)}$ and $g^{(t)}$. Furthermore, if we can find an $\alpha(\sigma)$ such that

$$|\varphi^{(s+1)}(\sigma)| \leq 2\alpha^{(s)}(\sigma)\,|\zeta|^2, \qquad |g^{(t)}(\sigma)| \leq \alpha^{(t)}(\sigma), \qquad (A6)$$

then the following inequality will hold:

$$|\Phi^{(n)}_{\mu\nu}(\sigma)| \leq \sum_{r=0}^{n} M_r\mathfrak{F}^n_r(2\alpha^{(s)}\,|\zeta|^2, \alpha^{(t)}).$$

By definition, $g(\sigma)$ is given by

$$g(\sigma) = (1 - 2\mathbf{k}\cdot\boldsymbol{\zeta}\sigma + |\zeta|^2\sigma^2)^{-\frac{1}{2}} = \sum_{r=0}^{\infty} P_r(\gamma)(\sigma\,|\zeta|)^r,$$
$$\gamma = \mathbf{k}\cdot\boldsymbol{\zeta}/|\zeta|.$$

[10] In the appendix, all calculations are carried out in the rest frame determined by the world line at the point $\xi^\mu(\mu)$. As there is no danger of confusion, the asterisks are omitted over the equality (inequality) symbols.

If we use the inequalities

$$|P_r(\gamma)| \leq 1, \tag{A7}$$

we find that

$$|g^{(r)}(\sigma)| \leq [(1 - \sigma\,|\zeta|)^{-1}]^{(r)}, \qquad r \geq 0.$$

Similarly, from Eq. (2.8), we have

$$\varphi^{(1)}(\sigma) = \frac{1}{\sigma^2}[(1 - \sigma\mathbf{k}\cdot\zeta)g(\sigma) - 1]$$

$$= |\zeta|^2 \sum_{r=0}^{\infty} [P_{r+2}(\gamma) - \gamma P_{r+1}(\gamma)][\sigma\,|\zeta|]^r,$$

and so, from Eq. (A7),

$$|\varphi^{(r+1)}(\sigma)| \leq 2\,|\zeta|^2\,[(1 - \sigma\,|\zeta|)^{-1}]^{(r)}.$$

Consequently, Eq. (A6) is satisfied, with

$$\alpha(\sigma) = (1 - \sigma\,|\zeta|)^{-1}. \tag{A8}$$

The $\mathfrak{F}_r^n(\alpha^{(s+1)}, \alpha(t))$ are defined inductively by

$$\mathfrak{F}_r^{n+1} = \frac{\partial}{\partial\sigma}\mathfrak{F}_r^n + 2\,|\zeta|^2\,\alpha(\sigma)\mathfrak{F}_{r-1}^n,$$

where we have omitted the arguments $\alpha^{(s)}(\sigma)$. From this equation, and

$$\alpha^{(1)}(\sigma) = |\zeta|\,\alpha^2,$$

it can be shown inductively that

$$\mathfrak{F}_r^n = A_r^n\,|\zeta|^{n+r}\,\alpha^{n+1}, \tag{A9}$$

where the A_r^n are numerical coefficients satisfying the following recurrence relations:

$$A_r^{n+1} = (n + 1)A_r^n + 2A_{r-1}^n, \tag{A10}$$

with

$$A_0^n = n!, \qquad A_n^n = 2^n. \tag{A11}$$

These coefficients cannot be expressed in a simple closed form. However, they can be written as multiple series,

$$A_r^n = 2^n \sum_{\{h_1, h_2 \cdots h_{n-r}\}} h_1 h_2 \cdots h_{n-r}, \tag{A12}$$

where the summation is over all subsets $\{h_1 \cdots h_{n-r}\}$ of the set of integers $\{1, 2, \cdots, n\}$. (All products which differ only in the order of the factors are to be taken only once.)

Inserting these expressions into Eq. (A3), we see that

$$n!\,|{}_nJ_{\mu\nu}| \leq 2\sum_{r=0}^{n} M_r A_r^n \int \frac{|\zeta|^{n+r}}{(1 - \sigma\,|\zeta|)^{n+1}}\,d^3z$$

$$\leq 2\sum M_r A_r^n (1 - \rho a)^{-n-1} \int |\zeta|^{n+r}\,d^3z,$$

and so

$$|{}_nJ_{\mu\nu}| \leq \frac{8\pi}{n!}\sum M_r\left(\frac{A_r^n}{n + r + 3}\right) \times a^{n+r+3}(1 - \rho a)^{-n-1}. \tag{A13}$$

This proves that ${}_nJ_{\mu\nu}$ is asymptotically $O(1)$. In particular, if j_μ is three times continuously differentiable, then ${}_2J_{\mu\nu}$ will be $O(1)$, and (A13) becomes

$$|{}_2J_{\mu\nu}| \leq \tfrac{4}{315}\pi[14M_0 + 35M_1 a + 20M_2 a^2][a^5/(1 - \rho a)^3].$$

The functions $N_{\mu\nu}$ and $III_{\mu\nu}$ are obtained by differentiating $\Phi(\rho)$ with respect to ρ, and are given in Eqs. (2.13a) and (2.14a), respectively.

Continuous-Representation Theory. III. On Functional Quantization of Classical Systems

JOHN R. KLAUDER
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received 6 August 1963)

The form of Schrödinger's equation in a continuous representation is indicated for general systems and analyzed in detail for elementary Bose and Fermi systems for which illustrative solutions are given. For any system, a natural continuous representation exists in which state vectors are expressed as continuous, bounded functions of the corresponding classical variables. The natural continuous representation is generated by a suitable set $\mathfrak{S}$ of unit vectors labeled by classical variables for which, for the system in question, the quantum action functional restricted to the domain $\mathfrak{S}$ is equivalent to the classical action. When a classical action is viewed in this manner it contains considerable information about the quantum system. Augmenting the classical action with some physical significance of its variables, we prove that the classical theory virtually determines the quantum theory for the Bose system, while it uniquely determines the quantum theory for the Fermi system.

1. INTRODUCTION

In an earlier paper,[1] we introduced functional representations of a Hilbert space $\mathfrak{H}$ by continuous, bounded functions $\psi(\Phi) \equiv (\Phi, \Psi)$ defined for all $\Psi \in \mathfrak{H}$ and $\Phi \in \mathfrak{S}$, where $\mathfrak{S}$ is a subset of unit vectors having the properties of an overcomplete family of states (OFS). The richness of vectors in an OFS, its first defining property,[1] permits continuous paths $\Phi(t)$ to be constructed completely within $\mathfrak{S}$. This property enabled us to study in Part II[2] a generalized relationship between classical and quantum mechanics based on the quantum-action functional[3]

$$I_{\mathfrak{S}}\{\Phi(t)\} = \int [i(\Phi, d\Phi) - (\Phi, \mathcal{H}\Phi)\, dt], \qquad (1)$$

whose domain is restricted such that $\Phi(t) \in \mathfrak{S}$. It was shown in II that when, as permitted by the second property of an OFS,[1] a continuous labeling by labels l in some label space $\mathfrak{L}$ was introduced for the vectors in $\mathfrak{S}$ (i.e., $l \to \Phi[l]$), then (1) assumed the form of a classical action functional $I[l(t)]$. Explicit examples included the familiar Bose and Fermi action functionals, which were discussed in Secs. II.2 and II.4, respectively.[4]

In II, no essential use was made of the third property of an OFS[1] that postulates a measure $d\mu(\Phi)$ with which the unit operator in $\mathfrak{H}$ may be resolved as

$$1 = \int_{\mathfrak{S}} \Phi\, d\mu(\Phi)\Phi^{\dagger}. \qquad (2)$$

where $\Phi\Phi^{\dagger}$ denotes the projection operator onto Φ. Equation (2) forms the basis of a continuous representation[1] of $\mathfrak{H}$, for then the inner product may be expressed in the form

$$(\Psi', \Psi) = \int_{\mathfrak{S}} \psi'^{*}(\Phi)\, d\mu(\Phi)\psi(\Phi). \qquad (3)$$

In the present paper we shall examine the quantum equations of motion expressed in a continuous representation, and therefore we shall draw heavily on the properties (2) and (3) and on related formulas presented in I.

Among all the possible continuous representations, or among all the possible OFS's generating continuous representations, one stands out as being natural for a particular system. In the identification of (1) with a classical action, it is necessary to reinterpret the classical variables as labels for unit vectors in Hilbert space. In all cases examined, these vectors constitute an OFS,[2] and therefore they generate a continuous representation of $\mathfrak{H}$. Consequently we shall define the *natural continuous representation* for a specific system as that representation generated by the OFS $\mathfrak{S}$ which figures so significantly in the reinterpreted classical action (1) pertaining to that system. Since (1), by definition, depends on the classical variables, the natural continuous representation will be expressed as functions of the natural classical variables. For example, in a one-dimensional, single-particle problem, the classical variables are p and q; thus the corresponding natural continuous representation of Hilbert space,

[1] J. R. Klauder, J. Math. Phys. **4**, 1055 (1963), hereafter referred to as I. References to this paper carry the prefix I.
[2] J. R. Klauder, J. Math. Phys. **4**, 1058 (1963), hereafter referred to as II. References to this paper carry the prefix II.
[3] In the present paper we choose units such that $\hbar = 1$.
[4] In equating (1) with a classical action, we are permitting Planck's constant to enter the classical formalism. From the point of view of conventional quantum mechanics, factor-ordering ambiguity enters at this stage; for additional discussion see Sec. II.2.

a phase-space continuous representation, will be given in terms of suitable functions $\psi(p, q)$. This leads to a new formulation of the familiar Schrödinger equation which is discussed in Sec. 2. A fermion degree of freedom, whose "classical" variable may be taken as a complex number χ, $0 \leq |\chi| \leq 1$, is discussed in Sec. 3 in analogous terms.

In order to construct continuous representations of the quantum equations of motion, it is of course necessary to know the set $\mathfrak{S}$ generating the representation and, in general, also the Hamiltonian operator $\mathfrak{H}$.

An extremely useful characterization of $\mathfrak{S}$, apart from unitary equivalences, is determined by the function

$$\mathfrak{K}(l'; l) \equiv (\Phi[l'], \Phi[l]) \tag{4}$$

for all pairs in $\mathfrak{L}$. This function forms the reproducing kernel of the continuous representation [cf. Eq. (I4)].

In special cases it is not necessary to know the Hamiltonian operator $\mathfrak{H}$ to determine the quantum dynamics. Suppose that $\exp(-it\mathfrak{H})\mathfrak{S} = \mathfrak{S}$ for all t. Then the Euler–Lagrange equations of (1)—the classical equations of motion—are *exact*, and such systems are so named. The classical solution $l_{cl}(t)$, expressed as a label-space path, correctly characterizes the evolution of $\Phi[l] \in \mathfrak{S}$ by $\Phi[l_{cl}(t)]$, or in the natural continuous representation by

$$\begin{aligned} \varphi_l(l', t) &= (\Phi[l'], \Phi[l_{cl}(t)]) \\ &= \mathfrak{K}(l'; l_{cl}(t)) \\ &= \mathfrak{K}(l'_{cl}(-t); l), \end{aligned} \tag{5}$$

where $l_{cl}(0) \equiv l$, $l'_{cl}(0) \equiv l'$. The last form of (5) follows from the unitarity of the evolution operator *and* because we have expressed $\Phi[l_{cl}(t)]$ in the natural continuous representation. The evolution of *any* initial state $\psi(l') = (\Phi[l'], \Psi)$ is determined from superposition of the solutions (5) to be $\psi(l'_{cl}(-t))$. Thus for exact systems the form of the quantum dynamics is completely specified by $\mathfrak{K}$ and the classical solutions. Oscillator examples with this property are given in Secs. 2 and 3.

For inexact systems we require the quantum Hamiltonian. A convenient form in which to express $\mathfrak{H}$ is by means of

$$\mathfrak{H}(l'; l) \equiv (\Phi[l'], \mathfrak{H}\Phi[l]), \tag{6}$$

its matrix elements in the natural continuous representation. Given (6) we can immediately construct the complete quantum action functional

$$\begin{aligned} I = \int [&i\psi^*(l)\delta_l\, \partial\psi(l)/\partial t \\ &- \psi^*(l')\delta l'\mathfrak{H}(l'; l)\delta l\psi(l)]\, dt, \end{aligned} \tag{7}$$

the result of a maximal domain enlargement of (1), as well as Schrödinger's equation

$$i\, \partial\psi(l')/\partial t = \int \mathfrak{H}(l'; l)\delta l\psi(l), \tag{8}$$

the extremal equation of (7), both expressed in the natural continuous representation. In (7) and (8), $\delta l \equiv d\mu(\Phi[l])$.

It is to be noted that partial information regarding $\mathfrak{K}$ is contained in (1) in the expression

$$i(\Phi[l], d\Phi[l]), \qquad \Phi[l] \in \mathfrak{S}, \tag{9}$$

which we call the canonical kinematical form (CKF). Similarly, partial information regarding $\mathfrak{H}$ is contained in (1) in the "classical" Hamiltonian

$$H(l) \equiv \mathfrak{H}(l; l) \equiv (\Phi[l], \mathfrak{H}\Phi[l]). \tag{10}$$

The last question to which we address ourselves in this paper is the following: *To what extent does the classical theory*—in the form of Eqs. (9) and (10) plus some knowledge of the meaning of the classical variables—*determine the reproducing kernel* $\mathfrak{K}(l'; l)$ *and the Hamiltonian matrix elements* $\mathfrak{H}(l'; l)$?

For the single-particle example, we find that $\mathfrak{K}$ is determined by the classical theory up to a function of one variable. For any $\mathfrak{K}$, the matrix elements of a polynomial Hamiltonian are uniquely determined by their diagonal elements, while if $\mathfrak{K}$ never vanishes the matrix elements of any Hamiltonian are uniquely determined by their diagonal elements. For the single fermion degree of freedom, both $\mathfrak{K}$ and $\mathfrak{H}$ are completely determined by the classical theory.

In both examples the reproducing kernel is found by roughly the same method, which involves, in part, identifying the CKF with a linear sum of Maurer–Cartan differential forms belonging to some Lie group.[5] The relevant ideas of this aspect of our calculation are outlined in general terms in the Appendix. The determination of the Hamiltonian operator is somewhat special to each case.

In conclusion we find the interesting and nontrivial result that the reinterpreted classical action functional (1) combined with the meaning of the classical variables, essentially determines the quantum action and quantum dynamics for the cases studied (and for their obvious generalizations to systems with additional degrees of freedom). These examples have their analogue in field quantization, as will be shown in a subsequent paper.

[5] P. M. Cohn, *Lie Groups*, (Cambridge University Press, London, 1961), Chaps. IV and V; C. Chevalley, *Theory of Lie Groups I* (Princeton University Press, Princeton, New Jersey, 1946), Chap. V.

2. PARTICLE QUANTIZATION IN THE NATURAL CONTINUOUS REPRESENTATION

We consider a one-dimensional, single-particle example whose classical action is taken in the form

$$I = \int [p\dot{q} - H(p, q)]\, dt. \tag{11}$$

Let P and Q denote irreducible, self-adjoint quantum mechanical momentum and position operators which fulfill $[Q, P] = i$. Then it was shown in Sec. II.2 that for any fiducial unit vector Φ_0, the set $\mathfrak{S}$ of vectors

$$\Phi[p, q] \equiv e^{-iqP}e^{ipQ}\Phi_0 \tag{12}$$

for all real p and q formed an OFS, and that such sets were appropriate to reinterpret the classical action (11) in the manner of Eq. (1). To yield this correspondence, Φ_0 need only satisfy

$$(\Phi_0, P\Phi_0) = (\Phi_0, Q\Phi_0) = 0 \tag{13}$$

in order that

$$i(\Phi[p, q], d\Phi[p, q]) = p\, dq. \tag{14}$$

The Hamiltonian H is determined from the operator $\mathcal{H}(P, Q)$ by

$$H(p, q) = (\Phi[p, q], \mathcal{H}(P, Q)\Phi[p, q]), \tag{15a}$$

$$= (\Phi_0, \mathcal{H}(P + p, Q + q)\Phi_0), \tag{15b}$$

which for any Φ_0 and polynomial $\mathcal{H}$ differs from $\mathcal{H}(p, q)$ by a term $\mathcal{O}$ of order $\hbar$. In cases where $\mathcal{H} = \frac{1}{2}P^2 + V(Q)$, Φ_0 may be chosen so that $\mathcal{O}$ is as small as desired apart from a constant $c \equiv \frac{1}{2}(\Phi_0, P^2\Phi_0)$ [cf. the discussion in Sec. II.2]. However, we shall leave the choice of Φ_0 open.

The natural continuous representation for the present example[6] is given in terms of

$$\psi(p, q) \equiv (\Phi[p, q], \Psi) \tag{16}$$

for all $\Psi \in \mathfrak{H}$. The inner product of two vectors is expressed by

$$(\Psi', \Psi) = \int \psi'^*(p, q)(dp\, dq/2\pi)\psi(p, q). \tag{17}$$

The reproducing kernel is

$$\mathcal{K}(p', q'; p, q) \equiv (\Phi[p', q'], \Phi[p, q])$$
$$= e^{-ip'(q-q')}(\Phi_0, \Phi[p - p', q - q']), \tag{18}$$

while the matrix elements of the Hamiltonian are determined by

$$\mathcal{H}(p', q'; p, q) = (\Phi[p', q'], \mathcal{H}(P, Q)\Phi[p, q])$$
$$= e^{-ip'(q-q')}(\Phi_0, \mathcal{H}(P + p', Q + q')$$
$$\times \Phi[p - p', q - q']). \tag{19}$$

In terms of these quantities, Schrödinger's equation [Eq. (8)] may be formulated as

$$i\, \partial\psi(p', q')/\partial t$$
$$= \int \mathcal{H}(p', q'; p, q)(dp\, dq/2\pi)\psi(p, q). \tag{20}$$

The functions $\psi(p, q)$ in (20) are not arbitrary but must be expressible in the form (16) for all t. The necessary and sufficient condition[6] is Eq. (I4); namely,

$$\psi(p', q') = \int \mathcal{K}(p', q'; p, q)(dp\, dq/2\pi)\psi(p, q), \tag{21}$$

which actually need only be verified initially. As initial conditions for Eq. (20), it suffices to consider

$$\psi(p, q) = \mathcal{K}(p, q; \bar{p}, \bar{q}) \tag{22}$$

for arbitrary $\bar{p}$, $\bar{q}$. That (22) fulfills (21) is ensured by Eq. (I6), and follows from (16), (18), and (21).

A more convenient form may be given for (20) by the following observations. The matrix elements of P,

$$P(p, q; p', q') \equiv (\Phi[p, q], P\Phi[p', q']),$$

may, on reference to (12), be expressed as

$$P(p, q; p', q') = (-i\partial/\partial q)\mathcal{K}(p, q; p', q').$$

Similarly, for the matrix elements of Q,

$$Q(p, q; p', q') = (q + i\partial/\partial p)\mathcal{K}(p, q; p', q').$$

More generally,[7] it follows that $\mathcal{H}(P, Q)$ has the matrix elements

$$\mathcal{H}(p, q; p', q')$$
$$= \mathcal{H}(-i\partial/\partial q, q + i\partial/\partial p)\mathcal{K}(p, q; p', q'), \tag{23}$$

which, together with (21), permit Eq. (20) to be reexpressed as

$$i\partial\psi(p, q)/\partial t = \mathcal{H}(-i\partial/\partial q, q + i\partial/\partial p)\psi(p, q). \tag{24}$$

Clearly, this form also follows from (16) and the

[6] A rigorous analysis of phase-space continuous representations characterized by (12), (16), and (17), and their relation to more familiar representations will be dealt with in a careful study in collaboration with J. McKenna as Part IV of the present series.

[7] Differentiability of reproducing kernels will be discussed in Part IV, where it will be shown that there exists a dense set of $\Phi_0 \in \mathfrak{H}$ such that $\mathcal{K}$ is C^∞ in all variables.

relation $i\,\partial\Psi/\partial t = \mathcal{H}\Psi$. Solutions to (24) must still fulfill (21) for all t. Again it is sufficient to satisfy (21) at one time, e.g., by imposing the initial condition (22). Equation (24) constitutes the phase-space form of Schrödinger's equation reported earlier.[8]

The fiducial unit vector Φ_0 does not appear in the operations called for in (24) but enters the solution nonetheless because $\psi(p, q)$ must fulfill (21). This is as it should be since the precise physical interpretation of $\psi(p, q)$ depends on the choice of Φ_0. Roughly speaking, $\psi(p, q)$ represents the probability amplitude that Ψ equals the state Φ_0 translated by q in configuration space and by p in momentum space. This may be seen more clearly perhaps in terms of a Schrödinger representation of (16), where

$$\psi(p, q) = \int \Phi_0^*(x - q)e^{-ip(x-q)}\Psi(x)\,dx. \qquad (25)$$

The flexibility in selecting Φ_0 is a reflection of the usual freedom of choosing distinct but unitarily equivalent representations. A clever choice of Φ_0 may simplify the analysis of a particular problem, or may make some perturbation solution more rapidly convergent.

It is interesting that, in a certain sense, Schrödinger's equation is obtained as a limiting form of (24). The uniform bound

$$|(i\partial/\partial p)^m\psi(p, q)| \leq ||Q^m\Phi_0||\,||\Psi||$$

based on Schwartz's inequality implies that the p derivatives in (24) may be made everywhere arbitrarily small by choosing Φ_0 sufficiently sharp about zero in configuration space. For purposes of (24), ψ then loses its p-dependence giving rise to Schrödinger's equation. In fact, the arbitrarily weak p dependence of ψ is important only for the inner product as defined by (17). However, by scaling each representative ψ by a suitable factor $N = N(\Phi_0)$—i.e., let $N\psi(p, q) \to \Psi(q)$ as Φ_0 becomes sharp—we can thereafter ignore the arbitrarily weak p dependence altogether and use the usual $L^2(q)$ norm. The scale factor N is conveniently determined by the requirement that

$$N^2\mathcal{K}(p, q; \bar{p}, \bar{q}) \to \delta(q - \bar{q}).$$

This type of transition to a Schrödinger representation is illustrated below in an example.[9] It is of course always possible to obtain a Schrödinger representation by using the unitary mapping $\psi(p, q) \leftrightarrow \psi(x)$ that exists between a continuous representation based on any Φ_0 and elements in $L^2(x)$.[6]

The usual momentum representation of Schrödinger's equation may be obtained from (24) by limiting procedures analogous to those used above. It is first necessary, however, to introduce the representative $'\psi(p, q)$,where

$$'\psi(p, q) \equiv e^{-ipq}\psi(p, q);$$

these functions again form a continuous representation defined by (16), but in which

$$\Phi[p, q] \equiv e^{ipQ}e^{-iqP}\Phi_0,$$

rather than (12). Equation (24) is transformed to

$$i\partial'\psi(p, q)/\partial t = \mathcal{H}(p - i\partial/\partial q, i\partial/\partial p)'\psi(p, q). \qquad (26)$$

The standard momentum representation of Schrödinger's equation can be obtained from (26) by choosing Φ_0 arbitrarily sharp about zero in momentum space, and proceeding in obvious analogy with the preceding discussion.

Phase-Space Form of Harmonic-Oscillator Solution

As an instructive, soluble example of Eq. (24), we discuss an harmonic oscillator for which we adopt

$$i\partial\psi(p, q)/\partial t = \tfrac{1}{2}[(-i\partial/\partial q)^2 + \omega^2(q + i\partial/\partial p)^2 - \omega]\psi(p, q). \qquad (27)$$

As initial condition we choose (22) in which Φ_0 is taken equivalent to

$$\Phi_0(x) = (\Omega/\pi)^{\frac{1}{4}} \exp(-\tfrac{1}{2}\Omega x^2),$$

the ground state of a *reference* harmonic oscillator with classical frequency Ω. The explicit solution to (27) is then

$$\psi(p, q, t) = C^{\frac{1}{2}}(t) \exp\{-\tfrac{1}{2}\Omega(q^2 + \bar{q}^2) + i(pq - \bar{p}\bar{q}) + \tfrac{1}{2}i\omega t + (4\omega\Omega)^{-1}C(t)(\omega\cos\omega t + i\Omega\sin\omega t)[(q\Omega - ip)^2 + (\bar{q}\Omega + i\bar{p})^2] + (2\Omega)^{-1}C(t)(q\Omega - ip)(\bar{q}\Omega + i\bar{p})\}, \qquad (28)$$

where

$$C^{-1}(t) \equiv \cos\omega t + i\zeta\sin\omega t,$$
$$\zeta \equiv \tfrac{1}{2}[(\omega/\Omega) + (\Omega/\omega)].$$

Equation (28) exhibits the evolution of any member of $\mathfrak{S}$, and constitutes an appropriate solution to (27) for any Ω.

[8] J. R. Klauder, Bull. Am. Phys. Soc. **8**, 466 (1963).

[9] A related construction of the Schrödinger representation follows readily from Eq. (25). Let β be a point such that $\Phi_0(\beta) \neq 0$. Then

$$\Psi(q) = [2\pi\Phi_0(\beta)]^{-1}\int e^{ip\beta}\,\psi(p, q - \beta)\,dp$$

is the desired representation. In the present case, the integration converts any fiducial vector into a sharp state.

To illustrate the transition to a Schrödinger representation, we choose Ω arbitrarily large and set the scaling factor $N = (\Omega/4\pi)^{\frac{1}{2}}$ as follows from the form of the reproducing kernel [given by (28) when $t = 0$]. The Schrödinger representation of the evolution of a "state" that was initially a sharp "position eigenfunction" at $\bar{q}$ (whence N^2, and not N, below) is then given by

$$\Psi(q, t; \bar{q}, 0) \equiv \lim_{\Omega\to\infty} [(\Omega/4\pi)^{\frac{1}{2}}\psi(p, q, t)]$$
$$= (\omega \csc \omega t/2\pi i)^{\frac{1}{2}} \exp \{i\omega[\tfrac{1}{2}(q^2 + \bar{q}^2) \times \csc \omega t - q\bar{q} \cot \omega t + \tfrac{1}{2}t]\}, \quad (29)$$

which is the well-known result for an harmonic oscillator.[10]

In the special case $\Omega = \omega$, the set $\mathfrak{S}$ is exact for the oscillator apart from phase factors. When $\Omega = \omega$, Eq. (28) becomes

$$\psi(p, q, t) = \exp \{-\tfrac{1}{4}\omega[q_{cl}(-t) - \bar{q}]^2 - \tfrac{1}{4}\omega^{-1}[p_{cl}(-t) - \bar{p}]^2 + \tfrac{1}{2}i[q_{cl}(-t)\bar{p} - p_{cl}(-t)\bar{q} + pq - \bar{p}\bar{q}]\}, \quad (30)$$

where

$$q_{cl}(-t) \equiv q \cos \omega t - \omega^{-1}p \sin \omega t, \quad (31a)$$

$$p_{cl}(-t) \equiv \omega q \sin \omega t + p \cos \omega t. \quad (31b)$$

For all t, there exists a choice of p and q in (30) [such that $p_{cl}(-t) = \bar{p}$, $q_{cl}(-t) = \bar{q}$] for which $|\psi(p, q, t)| = 1$. But since

$$|\psi(p, q, t)| \leq ||\Psi(t)|| = ||\Psi(0)|| \equiv 1$$

holds, it follows that, up to a phase factor, $\psi(p, q, t)$ represents a vector in $\mathfrak{S}$. This phase factor may be eliminated (cf. Sec. II.2) by adopting the OFS

$$\Phi[p, q] \equiv e^{-i(qP - pQ)}\Phi_0$$

rather than (12). The net effect of this choice is to eliminate the term $pq - \bar{p}\bar{q}$ in (30), whereupon for all t there exists a p and q such that $\psi(p, q, t) = 1$. From the discussion in Sec. 1, we see we are dealing with an exact system for which, from (5), it is necessary that

$$\psi(p, q, t) = \mathcal{K}(p_{cl}(-t), q_{cl}(-t); \bar{p}, \bar{q}).$$

This is just the form of (30)—less the term $(pq - \bar{p}\bar{q})$—wherein the functions defined by (31) are indeed classical solutions based on the Hamiltonian $H(p, q) = \frac{1}{2}(p^2 + \omega^2q^2)$ as defined by (15).

[10] For example, R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); M. S. Bartlett and J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 545 (1949), Eq. (2.7).

Since the set $\mathfrak{S}$ is exact when $\Omega = \omega$, the solution of (27) for any initial condition $\psi(p, q, 0) \equiv \psi(p, q)$ satisfying (21) is simply

$$\psi(p, q, t) = \psi(p_{cl}(-t), q_{cl}(-t)),$$

where Eq. (31) still applies.

We note that forced oscillators are also exact systems when $\Omega = \omega$ apart from possible phase factors, which again may be eliminated.

The harmonic oscillator is, of course, a rather special example, soluble in any quantization technique, whose solutions are, in one way or another, related to the classical solutions. Our solutions are no exception. However, our last results should also be considered within the framework of Sec. 1 as one example of the class of all exact systems, some of which do not even possess classical analogues in the traditional sense.

Determination of the Reproducing Kernel and Hamiltonian from Classical Theory

Construction of the Reproducing Kernel

We examine the extent that the identification (14) combined with some significance of the variables p and q determine that $\mathcal{K}$ is based on an OFS of the form (12). By significance in the present context, we mean the following: We assume that variables p and q exist that admit a simultaneous interpretation as Cartesian coordinates, and we adopt such variables. Homogeneity and the probability significance of $|\mathcal{K}|^2$ then demand that

$$\mathcal{K}(p', q'; p, q) = f(p - p', q - q')e^{-i\beta}, \quad (32)$$

where $\beta = \beta(p', q'; p, q)$ is real.

The assumed form of (32) determines as well the form of the measure on phase-space points in the "resolution of unity" in the form (2). Let this measure be given by $d\mu(p, q) = \Delta(p, q)dp\,dq$. Then the requirement that $\mathcal{K}$ be the inner product of two unit vectors forming an OFS becomes

$$1 = \int |f(p - p', q - q')|^2\Delta(p, q)\,dp\,dq,$$

for all p' and q' leading to $\Delta(p, q) \equiv \Delta = \text{constant}$. We show below that an irreducible representation of P and Q requires that $\Delta = (2\pi)^{-1}$.

We follow the Appendix closely in our further study of Eq. (32). If p and q labeled a unitary representation of a two-parameter Lie group, it would necessarily be Abelian as required by the form assumed in (32). Hence $\beta = \beta(p - p', q - q')$. A computation of the CKF yields a total differential

$$[dp(\partial/\partial p) + dq(\partial/\partial q)][\beta(0, 0) + if(0, 0)],$$

rather than $p\,dq$. Thus p and q do not label a two-parameter group.[11]

In spite of this conclusion, p and q do label a set of elements $V[p, q]$ that generate a group $\mathcal{G}$, which we now seek. Unitary invariance of (32) suggests that

$$\tilde{p} = p_0 + p, \tag{33a}$$

$$\tilde{q} = q_0 + q \tag{33b}$$

are still appropriate combination laws in $\mathcal{G}$ but that they are not the whole story. Equations (A3) and (A5) (see Appendix) state that the CKF should be invariant against constant, left translations. However, from (33a) and (33b),

$$\tilde{p}\,d\tilde{q} = p\,dq + d(p_0 q) \neq p\,dq,$$

which shows the necessity of additional parameters. Repeated left translations, which by definition cover the group generated by $\{V[p, q]\}$, never add anything to $p\,dq$ but a total differential. Hence, one additional parameter, α, should suffice, and the augmented CKF is taken as $p\,dq + d\alpha$. The left translation of α is defined so that

$$\tilde{p}\,d\tilde{q} + d\tilde{\alpha} = p\,dq + d\alpha,$$

fulfilling (A5). It follows immediately that

$$\tilde{\alpha} = \alpha_0 + \alpha - p_0 q, \tag{33c}$$

which completes the (local) group composition law in coordinates consistent with (14) and (32).

The adoption of the group (33) has strong physical support as well. In classical particle problems, the addition of a total differential to $p\,dq$ in no way changes the physics of a particular system, but rather makes possible a different description in terms of other canonical variables. The transformations from one set of canonical variables to another set form a group, and Eq. (33) is an analogue of that group. Indeed, each classical canonical pair is characterized by some particular curve $\alpha = \alpha(p, q)$ among the elements $[p, q, \alpha]$ in the group space corresponding to (33) (cf. Sec. II.2).

From (33) we see that the unit element is parameterized by $p = q = \alpha = 0$, while the element inverse to $[p, q, \alpha]$ is $[-p, -q, -\alpha - pq]$. Next we observe that if any two parameters are zero, the third describes a one-parameter subgroup in terms of canonical coordinates.[5] Thus a unitary representation of this group in these coordinates $U[p, q, \alpha]$ necessarily corresponds to a representation in a generalized form of canonical coordinates, of which those of the first or second kind are special cases.[12] We now note that the infinitesimal element associated with α, say L_α, commutes with the other infinitesimal elements. [This is readily proved by showing $c^b_{a\alpha} = 0$ from (A7) (see Appendix), where a and b stand for p, q, or α.] In generalized canonical coordinates, this commutation property of the infinitesimal element means that the finite element factorizes, i.e.,

$$U[p, q, \alpha] = V[p, q]W[\alpha], \tag{34}$$

where

$$W[\alpha] = \exp(\alpha L_\alpha).$$

The factorization (34) has for the reproducing kernel the consequence that

$$\begin{aligned}\mathcal{K}(p', q'; p, q) &\equiv (V[p', q']\Phi_0, V[p, q]\Phi_0)\\ &= (\Phi_0, U[-p', -q', -p'q']V[p, q]\Phi_0)\\ &= (\Phi_0, V[p - p', q - q']W[p'(q - q')]\Phi_0).\end{aligned} \tag{35}$$

Completeness of the set $\Phi[p, q]$ and the assumed form of (32) then require that

$$W[\alpha]\Phi_0 = e^{-i\alpha}\Phi_0, \tag{36}$$

the eigenvalue $-i$ for L_α being necessary in order that (35) also conform with (14). In view of (35) and (36), the quantity β in (32) satisfies the functional relation

$$\begin{aligned}&\beta(p', q'; p, q)\\ &\qquad = \beta(0, 0; p - p', q - q') + p'(q - q').\end{aligned}$$

Hermitian symmetry of $\mathcal{K}$ and (with no loss of generality) also of f leads to β being odd under the interchange $(p', q') \leftrightarrow (p, q)$. From these properties it follows that

$$\begin{aligned}&\beta(p', q'; p, q)\\ &\qquad = \tfrac{1}{2}(p' + p)(q - q') + \gamma(p - p', q - q'),\end{aligned}$$

where the homogeneous term γ remains undetermined, and whose contribution we assume hereafter to be included in f. Hence we have determined that

$$\begin{aligned}\mathcal{K}(p', q'; p, q) &= f(p - p', q - q')\\ &\quad \times \exp\{\tfrac{1}{2}[i(p' + p)(q' - q)]\},\end{aligned} \tag{37}$$

where $f(0, 0) = 1$ and $f^*(p, q) = f(-p, -q)$.

[11] It may be noted that, apart from the Abelian group, the only other essentially different two-parameter group is the affine group whose combination law we take as $\tilde{p} = p_0 p$, $\tilde{q} = q_0 + p_0^{-1} q$, expressed in suitable units (cf. Ref. 5, p. 67; our notation differs from Cohn's in several respects). In these coordinates, a linear combination of Maurer–Cartan differential forms (see discussion in Appendix) does equal $p\,dq$ as needed, but the group composition law is incompatible with (32).

[12] Cf. Cohn, Ref. 5, p. 110.

We now impose the idempotent condition Eq. (I6) on $\mathcal{K}$, which implies that

$$f(p', q') = \Delta \int f(p' - p, q' - q) \times e^{\frac{1}{2}i(pq' - qp')} f(p, q)\, dp\, dq. \quad (38)$$

If we introduce

$$w(x, q) \equiv \Delta \int e^{ipx} f(p, q)\, dp, \quad (39)$$

then (38) becomes

$$w(x, q') = \int w(x - \tfrac{1}{2}q, q' - q) \times w(x + \tfrac{1}{2}q' - \tfrac{1}{2}q, q)\, dq. \quad (40)$$

Further, let

$$w(x, q) \equiv v(x - \tfrac{1}{2}q, x + \tfrac{1}{2}q), \quad (41)$$

in which case (40) becomes simply

$$v(y, z) = \int v(y, q)\, dq v(q, z), \quad (42)$$

i.e., $v(y, z)$ is idempotent. The general solution to (42) is given by

$$v(y, z) = \sum_{r=1}^{R} a_r(y) b_r(z), \quad (43)$$

where

$$\int a_r(x) b_s(x)\, dx = \delta_{rs}.$$

It follows that R, the numbers of terms in (43), is given by

$$R = \int v(x, x)\, dx = \int w(x, 0)\, dx$$

$$= (2\pi\Delta) f(0, 0) = 2\pi\Delta,$$

in view of Eqs. (39) and (41). Therefore we find that $\Delta = R/2\pi$, where R is a positive integer. We investigate further the case $R = 1$, for which

$$\Delta = (2\pi)^{-1}.$$

The cases $R > 1$ can be shown to correspond to fully reducible representations of the operator algebra, and will not be discussed here.

In the case $R = 1$, $v(y, z)$ is simply $a(y)b(z)$, which leads to

$$f(p, q) = \int a(x - \tfrac{1}{2}q) b(x + \tfrac{1}{2}q) e^{-ixp}\, dx.$$

From $f^*(p, q) = f(-p, -q)$ it follows that $a^*(x) = b(x) \equiv \Phi_0(x)$. To recover the reproducing kernel we use (37) which, after a slight change of variable, yields

$$\mathcal{K}(p', q'; p, q) = \exp[ip(q' - q)] \int \Phi_0^*(x) \times \exp[-ix(p' - p)] \Phi_0(x + q' - q)\, dx. \quad (44)$$

The condition $f(0, 0) = 1$ requires of Φ_0 that

$$\int |\Phi_0(x)|^2\, dx = 1. \quad (45)$$

Equation (44) is recognized as a Schrödinger representation of (18), the sought-for reproducing kernel based on the OFS defined in (12) for irreducible P and Q. Equation (45) expresses the fact that Φ_0 is a unit vector.

Suppose we now (schematically) calculate the CKF from $\mathcal{K}$ as given in (44). By itself, the exponent outside the integral gives the proper answer $p\, dq$, but the computation would yield more. For example, a change of p in the integrand of (44) generates a contribution to the CKF equal to

$$-dp \int |\Phi_0(x)|^2\, x\, dx,$$

which, although a nontroublesome total differential, we set equal to zero. A similar consideration to annul the contribution of a variation of q in the integrand of (44) leads to yet another condition on Φ_0, both of which are summarized in Eq. (13). [It is to be noted that these conditions determine that linear combination of the Maurer–Cartan differential forms (see Appendix) which we identify as the CKF.]

Our derivation of the phase-space reproducing kernel from the classical information is thus complete; all but the fiducial vector $\Phi_0(x)$ is explicitly determined.

Uniqueness of the Hamiltonian Operator

To what extent do the diagonal elements (15) determine the full matrix (19)? Assume that two operators $\mathcal{H}_1$ and $\mathcal{H}_2$ lead to the same diagonal elements. Then if $\mathfrak{D} \equiv \mathcal{H}_1 - \mathcal{H}_2$, we have

$$(\Phi[p, q], \mathfrak{D}\Phi[p, q]) = 0 \quad (47)$$

for all p and q. Let $\mathfrak{D}$ be written in Weyl-like form:

$$\mathfrak{D} = (2\pi)^{-1} \int e^{-iq'P} e^{ip'Q} D(p', q')\, dp'\, dq'.$$

Then (47) becomes

$$0 = \int (\Phi_0, e^{-iq'P} e^{ip'Q} \Phi_0) \times e^{-ipq' + iqp'} D(p', q')\, dp'\, dq'. \quad (48)$$

The expectation value in (48) is just $\mathcal{K}(0, 0; p', q')$, and the vanishing of (48) means that

$$\mathcal{K}(0, 0; p', q')D(p', q') = 0$$

almost everywhere.

Clearly $D(p', q')$ can be considered a distribution with support on the set $\mathcal{U}$ where $\mathcal{K}(0, 0; p', q')$ vanishes. Since all of our reproducing kernels are continuous functions of their arguments and $\mathcal{K}(0, 0; 0, 0) = 1$, it follows that $\mathcal{U}$ does not include some neighborhood of the origin $p' = q' = 0$. Hence, for a $\mathfrak{D} \neq 0$ to satisfy (47), it is necessary that the support of $D(p', q')$ exclude a neighborhood of the origin. In particular, since polynomials in P and Q are supported *wholly* at the origin, we have

Theorem 1. A polynomial in the operators P and Q is uniquely determined by its diagonal matrix elements in any phase-space continuous representation.

If the set $\mathcal{U}$ where $\mathcal{K}$ vanishes is *empty*, then $D(p', q') = 0$ almost everywhere, and therefore $\mathfrak{D} = 0$. We state this as

Theorem 2. An operator is uniquely determined by its diagonal matrix elements in a phase-space continuous representation whose reproducing kernel never vanishes.

Nonvanishing reproducing kernels are generated by Gaussian fiducial vectors $\Phi_0(x)$ [set $t = 0$ in (30)], but this is by no means the only possibility. Let $r > 0$ and $\rho(s)$ be a nonnegative real function such that

$$\Phi_0(x) = \exp(-ibx)(x - ir)^{-1} \int_r^\infty (x - is)^{-1}\rho(s)\, ds$$

is in $L^2(-\infty, \infty)$ and has unit norm. If this function is taken as the Schrödinger representation of a fiducial vector, the reproducing kernel so generated never vanishes. The constant b is picked so that (13) is satisfied. The author is indebted to J. McKenna for this class of examples.

A convenient algorithm to compute the off-diagonal matrix elements of any operator is given in (23). Of course, in all our discussions we have tacitly assumed that the set $\mathfrak{S}$ is in the domain of the operators of interest.

3. QUANTIZATION OF FERMION DEGREES OF FREEDOM IN THE NATURAL CONTINUOUS REPRESENTATION

A single fermion degree of freedom may be described by a complex number χ bounded by the unit circle, $0 \leq |\chi| \leq 1$, and by the classical action

$$I = \int [\tfrac{1}{2}i(\chi^* \overleftrightarrow{\partial}_t \chi) - H(\chi)]\, dt, \tag{49}$$

where $A\overleftrightarrow{\partial}_t B \equiv A(\partial B/\partial t) - (\partial A/\partial t)B$.

Let Φ_0 and Φ_1 be any orthonormal pair of vectors spanning a two-dimensional Hilbert space $\mathfrak{H}$. Then it was shown in Sec. II.4 that the OFS whose vectors are

$$\Phi[\chi] \equiv (1 - |\chi|^2)^{\frac{1}{2}}\Phi_0 + \chi\Phi_1 \tag{50}$$

is an appropriate set to reinterpret (49) in the manner of (1). The classical Hamiltonian is given by

$$H(\chi) = (\Phi[\chi], \mathcal{H}\Phi[\chi]) \tag{51}$$

for some Hermitian operator $\mathcal{H}$.

The natural continuous representation for the present example involves

$$\psi(\chi) \equiv (\Phi[\chi], \Psi) \tag{52}$$

for all $\Psi \in \mathfrak{H}$, in terms of which the inner product reads

$$(\Psi', \Psi) = \int \psi'^*(\chi)(2\, d\chi_r\, d\chi_i/\pi)\psi(\chi), \tag{53}$$

where χ_r and χ_i are the real and imaginary parts of χ, respectively. The reproducing kernel of the representation follows from (50) as

$$\mathcal{K}(\chi'; \chi) = (1 - |\chi'|^2)^{\frac{1}{2}}(1 - |\chi|^2)^{\frac{1}{2}} + \chi'^*\chi. \tag{54}$$

The matrix elements of the most general Hermitian $\mathcal{H}$ may be expressed in the form

$$\begin{aligned}\mathcal{H}(\chi'; \chi) &= a(1 - |\chi'|^2)^{\frac{1}{2}}(1 - |\chi|^2)^{\frac{1}{2}} \\ &+ b(1 - |\chi'|^2)^{\frac{1}{2}}\chi + b^*\chi'^*(1 - |\chi|^2)^{\frac{1}{2}} + c\chi'^*\chi\end{aligned} \tag{55}$$

for arbitrary real a and c, and complex b. The diagonal elements

$$\begin{aligned}\mathcal{H}(\chi; \chi) &= a(1 - |\chi|^2) \\ &+ 2(1 - |\chi|^2)^{\frac{1}{2}}\,\mathrm{Re}\,(b\chi) + c\,|\chi|^2\end{aligned} \tag{56}$$

for all χ, clearly suffice to determine the parameters a, b, and c. Hence the operator $\mathcal{H}$ is uniquely specified by the classical Hamiltonian $H(\chi) \equiv \mathcal{H}(\chi; \chi)$.

Schrödinger's equation (8) may be formulated as

$$i\partial\psi(\chi')/\partial t = \int \mathcal{H}(\chi'; \chi)(2\, d\chi_r\, d\chi_i/\pi)\psi(\chi), \tag{57}$$

whose solution must simultaneously fulfill

$$\psi(\chi') = \int \mathcal{K}(\chi'; \chi)(2\, d\chi_r\, d\chi_i/\pi)\psi(\chi). \tag{58}$$

It is only necessary to verify (58) at one time, since continued verification is ensured by (57). An

appropriate initial condition for (57) is just

$$\psi(\chi') = \mathcal{K}(\chi'; \tilde{\chi}) \tag{59}$$

for arbitrary $\tilde{\chi}$.

Fermion Oscillator

The classical fermion-oscillator Hamiltonian $\omega\chi^*\chi$ corresponds to $c = \omega$, $a = b = 0$ in (56), and thus in (55) as well. In this case, Eq. (57) reads

$$i\partial\psi(\chi', t)/\partial t = \omega\chi'^*\lambda(t),$$

$$\lambda(t) \equiv \int \chi(2\, d\chi_r\, d\chi_i/\pi)\psi(\chi, t).$$

Consistency of this solution requires that

$$i\dot{\lambda} = \lambda \int \omega\, |\chi|^2\, (2\, d\chi_r\, d\chi_i/\pi) = \omega\lambda,$$

i.e., $\lambda(t) = \exp(-i\omega t)\lambda(0)$. Therefore

$$\psi(\chi', t) = \chi'^* e^{-i\omega t}\lambda(0) + f(\chi'),$$

and imposing the initial condition (59) yields

$$\psi(\chi', t) = (1 - |\chi'|^2)^{\frac{1}{2}}(1 - |\tilde{\chi}|^2)^{\frac{1}{2}} + \chi'^* e^{-i\omega t}\tilde{\chi}. \tag{60}$$

We note that this solution may be expressed as

$$\psi(\chi', t) = \mathcal{K}(\chi'_{cl}(-t); \tilde{\chi}), \tag{61a}$$

where

$$\chi'_{cl}(-t) \equiv e^{i\omega t}\chi'$$

represents the classical solution of (49) when $H = \omega\chi^*\chi$. More generally, for any initial condition $\psi(\chi', 0) \equiv \psi(\chi')$ satisfying (58), the solution becomes

$$\psi(\chi', t) = \psi(\chi'_{cl}(-t)). \tag{61b}$$

Thus the classical fermion oscillator provides another example of an exact system.

A forced fermion oscillator with classical Hamiltonian $H = [\omega + \gamma(t)]\chi^*\chi$ is likewise an exact system whose solutions (61) involve $\chi'_{cl}(-t) = \exp[-i\int_0^t (\omega + \gamma)\, dt]\chi'$.

The oscillator ground state is represented by $(1 - |\chi|^2)^{\frac{1}{2}}$ in the natural continuous representation [as may be deduced from the limit $\tau \equiv it \to +\infty$ applied to (60), in which only the ground-state contribution survives]. Therefore, Φ_0 in (50) is the ground state of the particular oscillator under study, while similar reasoning shows Φ_1 to be the corresponding occupied state. Physically, χ is the probability amplitude for occupation of the relevant oscillator since $|(\Phi[\chi], \Phi_1)|^2 = |\chi|^2$.

The present description of a single oscillator is easily extended to a finite number of degrees of freedom, and a related formal treatment of fermion fields was discussed earlier.[13] A rigorous analysis of fermion fields will be treated elsewhere.

Determination of the Reproducing Kernel from Classical Theory

We shall show for a fermion degree of freedom that the canonical kinematical form and the physical significance of its variables determine the reproducing kernel, which is also shown to span a two-dimensional Hilbert space. We concentrate on establishing the form (50) for $\Phi[\chi]$, since it was already remarked that the Hamiltonian operator is determined by the classical Hamiltonian (56).

The CKF which the sought-for OFS fulfills is

$$i(\Phi[\chi], d\Phi[\chi]) = \tfrac{1}{2}i(\chi^* d\chi - d\chi^*\chi), \tag{62}$$

where χ is any complex number satisfying $0 \leq |\chi| \leq 1$. This restriction follows because we adopt for χ the significance of a probability amplitude corresponding to some measurable property. Probability amplitudes being the inner product of unit vectors, it follows that there exists a normalized vector Φ_1 such that

$$|(\Phi[\chi], \Phi_1)|^2 = |\chi|^2, \tag{63}$$

and furthermore that $\Phi_1 \in \mathfrak{S}$ since all χ on the unit circle are allowed.

Our further analysis is patterned after the Appendix. Let $\Phi[\chi] \equiv V[\chi]\,\Phi_0$, and assume first that the unitary transformations $V[\chi]$ form a two-parameter Lie group. This assumption however is readily disproved, for neither the Abelian nor the affine two-parameter group (discussed in Sec. 2 and Footnote 11, respectively) are consistent with both (62) and (63). We therefore envisage the group $\mathcal{G}$ generated by the set $\{V[\chi]\}$.

We shall follow the guide set in Sec. 2 and add to (62) a total differential as the simplest generalization of the CKF that has as a physical consequence no alteration of the classical dynamics. More precisely, we seek a three-parameter unitary group $V[\chi, \alpha]$ for which the OFS $\Phi[\chi, \alpha] = V[\chi, \alpha]\Phi_0$ has

$$i(\Phi[\chi, \alpha], d\Phi[\chi, \alpha]) = \tfrac{1}{2}i(\chi^* d\chi - d\chi^*\chi) + d\alpha \tag{64}$$

for its CKF.

A cursory examination of the four essentially different three-parameter Lie groups shows that all but the rotation group are obviously incompatible with (63) and (64). In suitable Eulerian angles φ, θ, ψ, the group elements of the rotation group are

$$V[\varphi, \theta, \psi] = e^{-i\varphi M_z}e^{-i\theta M_y}e^{-i\psi M_z}, \tag{65}$$

[13] J. R. Klauder, Ann. Phys. (N. Y.) **11**, 123 (1960).

where

$$[M_x, M_y] = iM_z$$

plus cyclic permutations. A computation of the CKF for the set $\Phi[\varphi, \theta, \psi] \equiv V[\varphi, \theta, \psi]\Phi_0$ yields

$$\begin{aligned} i(\Phi, d\Phi) &= d\varphi[\cos\theta\langle M_z\rangle \\ &\quad - \sin\theta(\cos\psi\langle M_x\rangle - \sin\psi\langle M_y\rangle)] \\ &\quad + d\theta[\cos\psi\langle M_x\rangle + \sin\psi\langle M_y\rangle] + d\psi\langle M_z\rangle, \end{aligned} \tag{66}$$

where $\langle M_z\rangle \equiv (\Phi_0, M_z\Phi_0)$, etc.

Physically, ψ must be a cyclic coordinate in problems of spherical symmetry and cannot enter the CKF (66) explicitly. Thus it is necessary that $\langle M_x\rangle = \langle M_y\rangle = 0$, whereupon Eq. (66) reduces to

$$\begin{aligned} i(\Phi, d\Phi) &= (\cos\theta d\varphi + d\psi)\langle M_z\rangle, \\ &= [(\cos\theta - 1)d\varphi + d(\psi + \varphi)]\langle M_z\rangle. \end{aligned} \tag{67}$$

Comparison with (65), expressed in the form

$$i(\Phi, d\Phi) = |\chi|^2\, d\tan^{-1}(\chi_i/\chi_r) + d\alpha,$$

permits us to set $\Phi[\chi, \alpha] = \Phi[\varphi, \theta, \psi]$ if we identify

$$\tfrac{1}{2}(1 - \cos\theta) = |\chi|^2, \tag{68a}$$

$$\varphi = -\tfrac{1}{2}\langle M_z\rangle^{-1}\tan^{-1}(\chi_i/\chi_r), \tag{68b}$$

$$\psi + \varphi = \langle M_z\rangle^{-1}\alpha. \tag{68c}$$

Thus we have found a three-parameter group compatible with (64). When $\alpha = 0$, the OFS of interest is

$$\Phi[\chi] = V[\varphi, \theta, -\varphi]\Phi_0, \tag{69a}$$

and, in particular,

$$\Phi_1 \equiv \Phi[1] = V[0, \pi, 0]\Phi_0. \tag{69b}$$

We now invoke (63), which with the help of (69), states

$$|(e^{i\varphi M_z}e^{i(\pi-\theta)M_y}e^{i\varphi M_z}\Phi_0, \Phi_0)| = |\chi| = \sin\tfrac{1}{2}\theta. \tag{70}$$

The required φ independence arises only if

$$e^{i\varphi M_z}\Phi_0 = e^{im\varphi}\Phi_0, \tag{71}$$

so that ψ and similarly α appear in the OFS strictly as phase factors.

Armed with the knowledge that α is a phase factor, the proof that the Hilbert space is two-dimensional is straightforward. Theorem 1 of I asserts that a multiple of the group-invariant measure is always appropriate to resolve unity in the manner of (2) whenever $\mathfrak{S}$ is invariant under a transitive, compact unitary group. This condition holds in the present case, and, for the rotation group, the invariant measure is $d\cos\theta\ d\varphi\ d\psi$, which, from (68), is proportional to $d\chi_r d\chi_i d\alpha$. Since α enters as a phase factor, it plays no role in the resolution of unity other than a nonessential scaling. If we omit $d\alpha$, then the net proportionality factor Δ is determined by requiring

$$\int |(\Phi[\chi], \Phi_1)|^2\, \Delta\, d\chi_r\, d\chi_i = 1.$$

Employing (63) we find $\Delta = 2/\pi$. Now, from the trace of Eq. (2) it follows that $\int d\mu(\Phi)$ determines the Hilbert-space dimension. In our example, the dimension is two, since

$$\int (2/\pi)\, d\chi_r\, d\chi_i = 2$$

when integrated over the unit disc. Of course, a two-dimensional Hilbert space is completely consistent with

$$|(\Phi_0, e^{-i\theta M_y}\Phi_0)| = \cos\tfrac{1}{2}\theta,$$

because $M_y^2 = \frac{1}{4}$ in the spin-$\frac{1}{2}$ representation.

An explicit form for $\Phi[\chi]$ is now easily established since the orthonormal pair Φ_0 and Φ_1 span $\mathfrak{H}$. In particular,

$$\Phi[\chi] = a\Phi_0 + b\Phi_1,$$

where, from (69) and (70),

$$a = (\Phi_0, \Phi[\chi]) = \cos\tfrac{1}{2}\theta = (1 - |\chi|^2)^{\frac{1}{2}},$$

$$b = (\Phi_1, \Phi[\chi]) = \sin\tfrac{1}{2}\theta e^{-i2m\varphi} = \chi,$$

which is just (50). This completes the derivation of the fermion OFS and, thereby, of the fermion reproducing kernel from the classical theory, which was our aim.

ACKNOWLEDGMENTS

It is a pleasure to extend thanks to E. I. Blount, to H. Leutwyler and especially to J. McKenna, for numerous discussions regarding continuous representations and their applications.

APPENDIX

Here we outline the connection between the CKF in Eq. (9) and the Maurer–Cartan (M–C) differential forms for a Lie group. We also discuss the M–C differential equations, a knowledge of which enables the composition law of the group to be found, and point out their relevance in determining $\mathcal{K}$. General references remain the same as in Footnote 5.

Let $V[l^a]$, $a = 1, 2, \cdots, N$ be an N-dimensional Lie group $\mathcal{G}$ of unitary transformations in an n-dimensional (possibly infinite) Hilbert space $\mathfrak{H}$. The

element $V^{-1}[l^a]dV[l^a]$, where $dV[l^a] \equiv V[l^a+dl^a]-V[l^a]$, is an infinitesimal transformation and may be expanded according to

$$V^{-1}[l^a]dV[l^a] = \omega^b(l^a)L_b, \tag{A1}$$

where L_b are N skew-Hermitian operators that form a basis of the Lie algebra $\mathfrak{g}$ associated with $\mathcal{G}$. The real linear differential forms $\omega^b(l^a)$ are the M–C differential forms.

The connection of the M–C forms to the CKF arises as follows: For some fiducial unit vector $\Phi_0 \in \mathfrak{H}$, let

$$\Phi[l^a] \equiv V[l^a]\Phi_0, \tag{A2}$$

and require that the set of such unit vectors form an OFS. For example, if the V form an n-dimensional irreducible representation of a compact group $\mathcal{G}$, then the vectors (A2) automatically form an OFS for any Φ_0. The CKF based on (A2) follows from (A1) as

$$i(\Phi, d\Phi) = i(\Phi_0, V^{-1}dV\Phi_0) = \omega^b v_b, \tag{A3a}$$

where

$$v_b \equiv i(\Phi_0, L_b\Phi_0). \tag{A3b}$$

Thus the CKF is some linear combination of M–C forms, the particular linear combination being determined by the real constants v_b.

Although an OFS may generally be defined as in (A2), it frequently occurs that the set $\{V[l^a]\}$ of unitary transformations do *not* in fact form a group. Nevertheless, Eq. (A3) is still correct since we can always imagine the set $\{V\}$ imbedded into the group $\mathcal{G}$ it generates, and in which the parameters characterizing the additional elements have been kept constant.

We return now to Eq. (A1) and derive the M–C differential equations from the left invariance of the M–C forms. Equation (A1) is invariant under

$$\begin{aligned} &V^{-1}[l^a]V^{-1}[l_0^a]V[l_0^a]dV[l^a] \\ &\quad = (V[l_0^a]V[l^a])^{-1}d(V[l_0^a]V[l^a]) \\ &\quad \equiv \omega^b((l_0\cdot l)^a)L_b. \end{aligned} \tag{A4}$$

where $V[l_0^a]$ is an unvaried transformation, and where the dot denotes group multiplication. Independence of the basis elements leads to

$$\omega^b(\tilde{l}^a) = \omega^b(l^a), \tag{A5}$$

where

$$\tilde{l}^a \equiv (l_0 \cdot l)^a. \tag{A6}$$

The N equations (A5) are the M–C differential equations whose solution enables the group composition law (A6) to be determined.

The M–C equations are of significance in the determination of the reproducing kernel Eq. (4) from a knowledge of the M–C forms. Indeed, when $\{V\}$ forms a group $\mathcal{G}$, $\mathcal{K}$ is an idempotent function depending only on $(l'^{-1} \cdot l)^a$. There are two difficulties in the determination of an appropriate $\mathcal{K}$ by using the M–C equations, however. Generally the variables of the CKF are insufficient to parameterize a group; and, even when the variables do parameterize a group, the CKF gives only *one* linear combination of M–C forms, whereas we need N separate equations (A5).

To make up for these deficiencies, we must marshall any other information about $\mathcal{K}$ that we can to provide the missing forms or their equivalent. Perhaps the biggest source of additional information can come from independent knowledge of the physical significance of the classical variables, which in turn leads to corresponding functional conditions on the kernel. Some standard Lie groups can then be examined to see if one of them meets all known requirements on $\mathcal{K}$.

The meaning of the classical variables may possibly be used to deduce additional M–C forms. Assume $\{V\}$ fails to form a group. As a consequence, some differentials, corresponding to group parameters held fixed, could be missing from (A3). From (A5), some insight may be gained into the way the fixed parameters would have appeared in (A3), were they free to vary. A simple application of this idea is discussed in Sec. 2; this example also serves best to make clear possible generalizations.

Finally we remark that the consistency of a given set of M–C forms to determine a group by (A5) may be ascertained from the Maurer–Cartan relations

$$d\omega^c = \tfrac{1}{2}c_{ab}^c\omega^a \wedge \omega^b, \tag{A7}$$

where $\wedge$ denotes exterior multiplication,[5] and where c_{ab}^c satisfy the well-known relations for the structure constants of a Lie algebra.

Separation of Angular Variables in the Bethe–Salpeter Equation for Two Spin-½ Particles

MARIAN GÜNTHER
*Plasma Physics Laboratory, Boeing Scientific Research Laboratories, Seattle, Washington**
(Received 28 May 1963)

The Bethe–Salpeter equation for two fermions, in its usual differential form, is reduced here to one which involves a single spatial and a single temporal variable. Concomitantly, the original underlying 16-dimensional spinor space is reducible to one of 8 dimensions for angular momentum $l \neq 0$ and of 4 dimensions for $l = 0$. The possibility of this reduction derives from the existence of one (or two) normal divisors of the matrices operating on the original spinor space, in these respective cases. The procedure is illustrated with the positronium problem, but may easily be seen to be quite general.

I. INTRODUCTION

ALTHOUGH the relativistic two-body equation known as the Bethe–Salpeter equation had been proposed and subsequently rigorously derived from basic principles by several authors[1] quite some time ago, to date there has been no instance known where it permits of an exact solution corresponding to a realistic physical situation. The most noteworthy of the attempts to solve it has perhaps been that of Wick and Cutkowsky[2] who solved the relativistic two-body problem exactly for two scalar particles interacting through the exchange of a quantum of zero mass, but only for (in the usage of the notation of the present paper) the $\kappa = 0$ case.

Similar attempts for spin-½ particles have also been made by Goldstein and later by Scarf and Umezawa[3] for a very specialized situation as far as the spin variables are concerned (also $\kappa = 0$ case). The latter attempts come closest to the problem dealt with in this note although no general angular analysis is attempted. A more thorough analysis of the angular-variable problem for the two-fermion case, as attempted here, seems furthermore indicated by the recent revival of interest in the Bethe–Salpeter equation, e.g., from the standpoint of Regge analysis.[4] Such an analysis has so far been attempted only for scalar particles where the angular-variable problem is rather easy and straightforward. The scope of this note does not permit an attempt to guess if the analysis given here would be of more use in the nonrelativistic limit or completely relativistic case. However, if solving the Bethe–Salpeter equation could be expected to make any substantial progress in the future, the author is inclined to believe that the present analysis would be of more importance in the latter, as indeed was the original aim of the author—this notwithstanding the fact that the present note does not go beyond the $\kappa = 0$ case as far as actually solving the equation is concerned.

The interrelation between nonrelativistic and extremely relativistic limits can be discussed, including angular analysis, more fully for spin-zero particles as done already by many authors. The nonrelativistic limit has been investigated *ab initio* by some authors[5] employing, e.g., the Foldy–Woythusen transformation[6] to separate "large" and "small" spinor components (which further obscures its connection with the original relativistic problem) and only then angular analysis is carried out. The nonrelativistic limit of the solutions of Goldstein and Scarf and Umezawa can be investigated, of course, but they become limited to the special cases of their relativistic counterparts.

II. ANGULAR ANALYSIS

To the lowest order in $\alpha = ie^2/2\pi hc$, the differential form of the Bethe–Salpeter equation for two fermions, as exemplified by the positronium problem,

* This work had its inception during the author's affiliation with the Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland.

[1] Y. Nambu, Progr. Theoret. Phys. (Kyoto) **5**, 614 (1950); E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951); M. Gell-Mann and F. Low, *ibid.* **84**, 350 (1951); M. Günther, *ibid.* **88**, 1411 (1952); M. Günther, *ibid.* **94**, 1347 (1954).

[2] G. S. Wick, Phys. Rev. **96**, 1124 (1954); R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954); F. L. Scarf, Phys. Rev. **100**, 912 (1955).

[3] J. S. Goldstein, Phys. Rev. **91**, 1516 (1953); H. S. Green, *ibid.* **97**, 1135 (1955); F. L. Scarf and H. Umezawa, *ibid.* **109**, 1848 (1958). Compare also H. S. Green and S. N. Biswas, Progr. Theoret. Phys. (Kyoto) **18**, 121 (1957).

[4] B. W. Lee and R. F. Sawyer, Phys. Rev. **127**, 2266 (1962); N. Nakanishi, *ibid.* **130**, 1230 (1963). See also N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **24**, 1275 (1960).

[5] Z. V. Chraplyvy, Phys. Rev. **91**, 1310 (1953); M. Kawaguchi, Progr. Theoret. Phys. (Kyoto) **25**, 178 (1961); J. Reinfelds, *ibid.* **27**, 1165 (1962). See also R. Karplus and A. Klein, Phys. Rev. **87**, 848 (1952) for the application of the "nonrelativistic limit" approach to the positronium problem.

[6] L. L. Foldy and S. A. Woythusen, Phys. Rev. **78**, 29 (1950).

reads

$$\{[\gamma_\mu^{(1)}\partial/\partial_{x_\mu}^{(1)} + \kappa_0][\gamma_\nu^{(2)}\partial/\partial_{x_\nu}^{(2)} + \kappa_0] + \alpha\gamma_\mu^{(1)}\gamma_\mu^{(2)}[x_\nu^{(1)} - x_\nu^{(2)}]^{-2}\}\Psi = 0. \quad (1)$$

After transforming away the motion of the "composite particle" and going over to its rest frame, by putting

$$\Psi = \exp[i\kappa(x_0^{(1)} + x_0^{(2)})]\chi(x_\mu^{(1)} - x_\mu^{(2)}), \quad (2)$$

we obtain

$$\{[\gamma_\mu^{(1)}\partial/\partial_{x_\mu} + i\gamma_0^{(1)}\kappa + \kappa_0][-\gamma_\nu^{(2)}\partial/\partial_{x_\nu} + i\gamma_0^{(2)}\kappa + \kappa_0] + \alpha\gamma_\mu^{(1)}\gamma_\mu^{(2)}(x_\nu)^{-2}\}\chi = 0, \quad (3)$$

where

$$x_\nu = x_\nu^{(1)} - x_\nu^{(2)}.$$

In Eq. (2), κ plays the role of the mass eigenvalue parameter of the positronium.

A systematic way of transforming spinors into spatial polar coordinates is furnished by the transformation

$$\chi \to \exp\{\tfrac{1}{2}(\gamma_3^{(1)}\gamma_1^{(1)} + \gamma_3^{(2)}\gamma_1^{(2)})\vartheta\} \times \exp\{\tfrac{1}{2}(\gamma_1^{(1)}\gamma_2^{(1)} + \gamma_1^{(2)}\gamma_2^{(2)})\varphi\}\chi. \quad (4)$$

This corresponds to the standard way of transforming spinors into a Cartesian coordinate system locally coincident with the r, ϑ, φ axes. The expressions

$$\gamma_i^{(1)}\partial/\partial x_i \quad \text{and} \quad \gamma_i^{(2)}\partial/\partial x_i, \qquad i = 1, 2, 3$$

then become

$$\gamma_3^{(1)}\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) + \frac{Q(1,2)}{r} + \frac{1}{2r} \times (\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)})\gamma_3^{(2)}, \quad (5)$$

and

$$\gamma_3^{(2)}\left(\frac{\partial}{\partial r} + \frac{1}{2/}\right) + \frac{Q(2,1)}{r} + \frac{1}{2r} \times (\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)})\gamma_3^{(1)}, \quad (5')$$

where

$$Q(l,k) = \gamma_2^{(l)}\left(\frac{1}{\sin\vartheta}\frac{\partial}{\partial\varphi} - \tfrac{1}{2}\gamma_1^{(k)}\gamma_2^{(k)}\cot\vartheta\right) + \gamma_1^{(l)}\left(\frac{\partial}{\partial\vartheta} + \tfrac{1}{2}\cot\varphi\vartheta\right); \qquad k, l = 1, 2. \quad (6)$$

$\gamma_0\partial/\partial x_0$, $\gamma_0\kappa$, and the invariant $\gamma_\mu^{(1)}\gamma_\mu^{(2)}$ are of course unaffected by the transformation (4).

We now have

$$[Q(1,2), Q(2,1)] = 0, \quad (7)$$

and

$$Q^2 = [Q(1,2)]^2 = [Q(2,1)]^2. \quad (8)$$

Furthermore, if $Q(1, 2)$ and $Q(2, 1)$ are multiplied by $\gamma_1^{(1)}$ and $\gamma_1^{(2)}$, respectively, only the products $\gamma_1^{(1)}\gamma_2^{(1)}$ and $\gamma_1^{(2)}\gamma_2^{(2)}$ appear, and the problem of finding the angular eigenfunctions would seem to pose no further difficulty. This is not quite so, however, since $Q(1, 2)$ and $Q(2, 1)$ fail to commute with the remaining terms of (5) and (5′). We therefore propose to apply a somewhat less direct approach, and try to find a spinor *matrix* function Z, such that

$$Q(1,2)Z = Z\gamma_1^{(1)}j, \qquad Q(2,1)Z = Z\gamma_1^{(2)}j, \quad (9)$$

and in the hope that "pulling Z through" $\gamma_3^{(1,2)}$ and $[\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)}]\gamma_3^{(1,2)}$ will result in some simple relations of the type of (9).

The eigenvalue problem (9) is especially simple in the case of no φ dependence, since then the Q^2 of (8) becomes

$$Q^2 = \frac{\partial^2}{\partial\vartheta^2} + \cot\vartheta\frac{\partial}{\partial\vartheta} - \tfrac{1}{2}(1 - \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)})\frac{1}{\sin^2\vartheta}. \quad (10)$$

We note that (10) is just the Legendre operator, whose eigenfunctions are the Legendre functions

$$P_l^{*\mu}(\cos\theta), \quad \text{where} \quad \mu = \tfrac{1}{2}(1 - \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)}). \quad (11)$$

Now, since μ is an idempotent operator, we also have $\mu^2 = \mu$. It may now be shown that the $P_l^\mu(\cos\theta)$ [alone of the two possibilities $P_l^{*\mu}(\cos\theta)$], also satisfy (9), with

$$j = [l(l+1)]^{\frac{1}{2}}. \quad (12)$$

If there is no φ dependence, we therefore have for the desired Z,

$$Z_l^0(\cos\theta) = P_l^\mu(\cos\theta) = \tfrac{1}{2}(1 - \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)})P_l^1(\cos\theta) + \tfrac{1}{2}(1 + \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)})P_l^0(\cos\theta). \quad (13)$$

The following relations are then also valid:

$$[\gamma_3^{(1,2)}, Z_l^0] = 0, \qquad [\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)}, Z_l^0] = 0, \quad (14)$$

and thus the Z so constructed satisfies the required conditions with the simple properties stated in (14). These latter then give the behavior of Z upon "pulling it through"

$$\gamma_3^{(1,2)} \quad \text{and} \quad (\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)})\gamma_3^{(1,2)}.$$

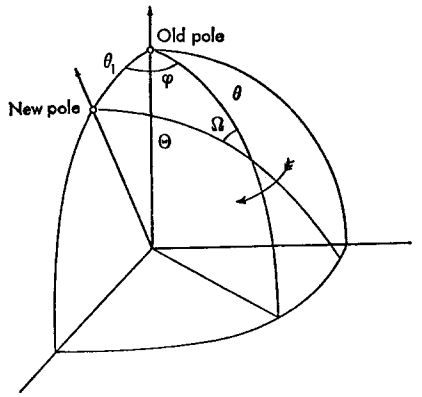


FIG. 1. Auxiliary drawing illustrating how φ-dependent eigenfunctions can be generated with help of φ independent ones.

Z_l^0 can now be used as a generating function for eigenfunctions in the most general case of φ dependence. Choosing a new pole on the unit sphere, at an angular distance θ_1 from the old one, we have the situation depicted in Figure 1.

We are now assured that the function[7]

$$Z_l = \exp\left[-\tfrac{1}{2}(\gamma_1^{(1)}\gamma_2^{(1)} + \gamma_1^{(2)}\gamma_2^{(2)})\Omega\right]P_l^\mu(\cos\vartheta) \tag{15}$$

will satisfy the general form of Eq. (9), and so also will all its Fourier components with respect to φ, since $\partial/\partial\varphi$ commutes with $Q(i, j)$. Upon Fourier-analyzing (15) with respect to φ, and using the spherical harmonic addition theorem, we then arrive at the most general set of solutions of (9):

$$\begin{aligned} Z_1^m(\vartheta\varphi) = \Big\{&\tfrac{1}{2}(1 + \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)}) \\ &+ [l(1+1)]^{-\frac{1}{2}}\,\tfrac{1}{2}(1 - \gamma_1^{(1)}\gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)}) \\ &\times \left(\frac{\partial}{\partial\vartheta} - \gamma_1^{(1)}\gamma_2^{(1)}\frac{1}{\sin\vartheta}\frac{\partial}{\partial\varphi}\right)\Big\}e^{im\varphi}P_l^m(\cos\vartheta). \end{aligned} \tag{16}$$

Because of (14) we then also have in general,

$$\begin{aligned} [\gamma_3^{(1,2)}, Z_l^m] &= 0, \\ [\gamma_1^{(1)}\gamma_2^{(1)} + \gamma_1^{(2)}\gamma_2^{(2)}, Z_l^m] &= 0. \end{aligned} \tag{16'}$$

When $Z_1^m(\varphi\vartheta)$ is "pulled through" the entire B–S equation (3), (and eventually dropped on the left side), the result is

$$\begin{aligned} \Big\{\Big[&\gamma_3^{(1)}\frac{\partial}{\partial r} + \frac{1}{r}(\gamma_3^{(1)} + \gamma_2^{(1)}\gamma_1^{(2)}\gamma_2^{(2)})[l(1+1)]^{\frac{1}{2}} \\ &+ \tfrac{1}{2}(\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)})\gamma_3^{(2)} \\ &+ \gamma_0\frac{\partial}{\partial x_0} + i\gamma_0\kappa + \kappa_0\Big] \\ &\times \Big[-\gamma_3^{(2)}\frac{\partial}{\partial r} - \frac{1}{r}(\gamma_3^{(2)} + \gamma_2^{(2)}\gamma_1^{(1)}\gamma_2^{(1)}) \\ &\times [l(1+1)]^{\frac{1}{2}} + \tfrac{1}{2}(\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)})\gamma_3^{(1)} \\ &- \gamma_0\frac{\partial}{\partial x_0} + i\gamma_0\kappa + \kappa_0\Big] \\ &+ \alpha\gamma_\mu^{(1)}\gamma_\mu^{(2)}[r^2 - x_0^2]^{-1}\Big\}\chi = 0 \end{aligned} \tag{17}$$

[where actually χ of Eq. (3) is equal to $Z\chi$ of Eq. (17)].

At this point we may effect the simplification of reducing the spinor space to a lesser dimensionality. We note that neither $\gamma_2^{(1)}$ nor $\gamma_2^{(2)}$ appears alone, but only in the combination $\gamma_2^{(1)}\gamma_2^{(2)}$. Consequently, there exists a normal divisor

$$N_l = \gamma_1^{(1)}\gamma_3^{(1)}\gamma_0^{(1)}\gamma_1^{(2)}\gamma_3^{(2)}\gamma_0^{(2)}, \tag{18}$$

which commutes with all terms of Eq. (17). The original 16-dimensional spinor space can therefore be split into two independent 8-dimensional subspaces wherein the normal divisor N_l plays the role of $+I$ and $-I$, respectively (I being the 8×8 identity matrix), and in each of which the quantities $\gamma_1^{(1)}$, $\gamma_3^{(1)}$, $\gamma_0^{(1)}$, $\gamma_1^{(2)}$, $\gamma_3^{(2)}$, $\gamma_0^{(2)}$, and the product $\gamma_2^{(1)}\gamma_2^{(2)}$ can be represented by three independent sets of Pauli σ matrices ($\sigma^{(1)}$, $\sigma^{(2)}$, Σ) as follows:

$$\begin{aligned} \gamma_{1,3}^{(1,2)} &\to \Sigma_3\sigma_{1,2}^{(1,2)}, \\ \gamma_0^{(1,2)} &\to i\Sigma_3\sigma_3^{(1,2)}, \\ \gamma_2^{(1)}\gamma_2^{(2)} &\to \Sigma_2. \end{aligned} \tag{19}$$

In the case $l = 0$, a further normal divisor N_0 exists, because now, again, neither $\gamma_1^{(1)}$ nor $\gamma_1^{(2)}$ appears alone in (17), but only in the combination

$$Q = \gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)}. \tag{20}$$

The normal divisor N_0 is

$$N_0 = \gamma_2^{(1)}\gamma_3^{(1)}\gamma_0^{(1)}\gamma_2^{(2)}\gamma_3^{(2)}\gamma_0^{(2)}. \tag{21}$$

The original 16-dimensional spinor space splits into four 4-dimensional subspaces, in each of which only two sets of Pauli σ matrices ($\sigma^{(1)}$, $\sigma^{(2)}$) suffice to represent the remaining operators, according to the correspondences[8]

$$\begin{aligned} &\gamma_s^{(1,2)} \to \sigma_1^{(1,2)}, \qquad \gamma_0^{(1,2)} \to \sigma_2^{(1,2)}, \\ &\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)} \to 2\sigma_3^{(1)}\sigma_3^{(2)}, \quad \text{or} \quad 0. \end{aligned} \tag{22}$$

If the wavefunction χ has nonvanishing com-

[7] The factor $\exp\left[-\tfrac{1}{2}(\gamma_1^{(1)}\gamma_2^{(1)} + \gamma_1^{(2)}\gamma_2^{(2)})\Omega\right]$ corresponds, in the spinor space, to the rotation by angle Ω about the r axis, and comes from readjusting the ϑ direction when going back to the old pole (Fig. 1).

[8] Since

$$\begin{aligned} \gamma_1^{(1)}\gamma_1^{(2)} + \gamma_2^{(1)}\gamma_2^{(2)} &= \gamma_1^{(1)}\gamma_1^{(2)}(1 - \gamma_1^{(1)}\gamma_1^{(2)}\gamma_2^{(1)}\gamma_2^{(2)}) \\ &= \gamma_1^{(1)}\gamma_1^{(2)}(1 - N_lN_0), \end{aligned}$$

and, in two of the subspaces, $N_lN_0 = 1$.

ponents only in the subspaces corresponding to $N_l N_0 = 1$, and if $l = 0$, then the B–S equation reduces to its simplest form, namely

$$\left\{\left[\sigma_1^{(1)}\frac{\partial}{\partial r} + i\sigma_2^{(1)}\frac{\partial}{\partial x_0} - \kappa_0\sigma_2^{(1)} + \kappa_0\right]\right.$$
$$\times\left[-\sigma_1^{(2)}\frac{\partial}{\partial r} - i\sigma_2^{(2)}\frac{\partial}{\partial x_0} - \kappa_0\sigma_2^{(2)} + \kappa_0\right]$$
$$\left. + \sigma(\sigma_1^{(1)}\sigma_1^{(2)} - \sigma_2^{(1)}\sigma_2^{(2)})(r_0^2 - x_0^2)^{-1}\right\}(r\chi) = 0. \quad (23)$$

III. A SOLUTION FOR $\kappa = 0$

Despite the great simplification, Eq. (23) is still very difficult to solve unless we put $\kappa = 0$, thereby artificially restoring the rotational symmetry in the (r, ix_0) plane, which was lost due to our having distinguished the "direction of flight" of the positronium system. This case, corresponding to the assumption of a zero-mass system, may actually be completely solved. The study of this case yields the conditions imposed on the interaction strength α, in order to yield a massless composite system. To conclude this work, we propose to give two examples of such solutions, one of which is identical with that studied by Goldstein. No criteria as to the physical admissibility of the wavefunctions will be studied here, however, because they could be worked out very much in the same fashion as done by the above authors. Besides, such discussion is best postponed until the complete variety of the $\kappa = 0$ solutions is arrived at. Work in this direction is being undertaken by the author.

With the substitutions

$$r = x/\kappa_0, \qquad -it = y/\kappa_0, \quad (24)$$

Eq. (23) for the $\kappa = 0$ case can be written as

$$\left\{\left[\sigma_1^{(1)}\frac{\partial}{\partial x} + \sigma_2^{(1)}\frac{\partial}{\partial y} + 1\right]\left[-\sigma_1^{(2)}\frac{\partial}{\partial x} - \sigma_2^{(2)}\frac{\partial}{\partial y} + 1\right]\right.$$
$$\left. + \frac{i\alpha}{2\pi}(\sigma_1^{(1)}\sigma_1^{(2)} + \sigma_2^{(1)}\sigma_2^{(2)})(x^2 + y^2)^{-1}\right\}\chi = 0. \quad (25)$$

Going over to polar coordinates in the (x, y)-plane, according to

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \quad (26)$$

and performing the transformation

$$\chi \to \chi' = \exp\left[(\tfrac{1}{2}i\varphi)(\sigma_3^{(1)} + \sigma_3^{(2)})\right]\chi, \quad (27)$$

we first get

$$\exp\left[+[\tfrac{1}{2}(i\varphi)(\sigma_3^{(1)} + \sigma_3^{(2)})]\left(\sigma_0^{(i)}\frac{\partial}{\partial x} + \sigma_2^{(i)}\frac{\partial}{\partial y}\right)\right.$$
$$\left.\times\exp\left[-(\tfrac{1}{2}i\varphi)(\sigma_3^{(1)} + \sigma_3^{(2)})\right]\right.$$
$$= \sigma_1^{(i)}\frac{\partial}{\partial\rho} + \sigma_2^{(i)}\frac{1}{\rho}\frac{\partial}{\partial\varphi} - \frac{i}{2\rho}\sigma_2^{(i)}$$
$$\times(\sigma_3^{(1)} + \sigma_3^{(2)}), \qquad i = 1, 2, \quad (28)$$

so that, if we assume no φ dependence of χ', Eq. (25) becomes

$$\left\{\left[\sigma_1^{(1)}\left(\frac{\partial}{\partial\rho} + \frac{\mu}{\rho}\right) + 1\right]\left[\sigma_1^{(2)}\left(\frac{\partial}{\partial\rho} + \frac{\mu}{\rho}\right) - 1\right]\right.$$
$$\left. + \frac{i\alpha}{2\pi}\frac{1}{\rho^2}\sigma_1^{(1)}\sigma_1^{(2)}(1 - \sigma_3^{(1)}\sigma_3^{(2)})\right\}\chi' = 0, \quad (29)$$

with

$$\mu = +\tfrac{1}{2}(1 + \sigma_3^{(1)}\sigma_3^{(2)}). \quad (30)$$

Since $\sigma_3^{(1,2)}$ does not appear alone in Eq. (29), but only in the combination $\sigma_3^{(1)}\sigma_3^{(2)}$, and $\sigma_2^{(1,2)}$ does not appear at all, there exists a further normal divisor

$$M = \sigma_1^{(1)}\sigma_1^{(2)} \quad (31)$$

with eigenvalues ± 1. The spinor dimensionality of (29) is therefore subject to further reduction. In the two irreducible 2-dimensional subspaces which thus result, we can make the following substitutions using *one* set of Pauli σ matrices:

$$\sigma_1^{(1)} \to \sigma_1, \qquad \sigma_1^{(2)} \to \pm\sigma_1, \qquad \sigma_3^{(1)}\sigma_3^{(2)} \to \sigma_3, \quad (32)$$

and, as a consequence of (30),

$$\mu \to \tfrac{1}{2}(1 + \sigma_3). \quad (33)$$

For the eigenvalue $+1$ of M, we therefore have

$$\left\{\left[\sigma_1\left(\frac{\partial}{\partial\rho} + \frac{\mu}{\rho}\right) + 1\right]\left[\sigma_1\left(\frac{\partial}{\partial\rho} + \frac{\mu}{\rho}\right) - 1\right]\right.$$
$$\left. + \frac{i\alpha}{2\pi\rho^2}(1 - \sigma_3)\right\}\chi' \equiv \left\{\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho}\right.$$
$$\left. + \frac{1}{\rho^2}\left[-\mu + \frac{i\alpha}{\pi}(1 - \mu)\right] - 1\right\}\chi' = 0, \quad (34)$$

whose solution is obviously

$$\chi' = H_{\pm\lambda_\pm}(\rho), \quad (35)$$

where $H_n(\rho)$ is any Bessel function, and where

$$\lambda_\pm = \gamma(1 - \mu) \pm \mu, \qquad \gamma = -(i\alpha/\pi)^{\frac{1}{2}} \quad (36)$$

with

(there are actually four values of λ since μ has two eigenvalues 1 and 0).

Equation (34) and two of its solutions correspond to the case studied by Goldstein[3] [compare Eq. (22) and the solution (23) of his paper if $(1-\mu)$ is replaced by 1 in the above Eq. (36).]

The equation which results for the eigenvalue -1 of M which reads

$$\left\{\left[\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)+1\right]\left[-\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)-1\right]-\frac{i\alpha}{2\pi\rho^2}(1-\sigma_3)\right\}\chi' \equiv -\left\{\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\times\left[-\mu+\frac{i\alpha}{\pi}(1-\mu)\right]+1+2\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)\right\}\chi^1=0 \tag{37}$$

is certainly more complicated however, at least because there appear two noncommuting matrices σ, and μ.

ACKNOWLEDGMENT

Cordial thanks are extended to Dr. Leo C. Levitt for his help in materially improving the clarity and conciseness of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 2 FEBRUARY 1964

Regge Behavior of Forward Elastic Scattering Amplitudes*

JAMES D. BJORKEN†
Stanford University, Stanford, California
(Received 7 August 1963)

A theorem concerning the asymptotic behavior of forward elastic scattering amplitudes in relativistic theories is stated and proved. The assumptions made are (1) identical spinless particles interact via $G\phi^3$ and $\lambda\phi^4$ couplings; (2) a cutoff of the propagators is introduced; (3) the forward scattering amplitude satisfies a Bethe–Salpeter equation in the crossed channel; (4) the kernel of the equation is an arbitrary finite subset of the Feynman graphs which compose the exact kernel. The theorem states that under these assumptions, the forward scattering amplitude exhibits Regge behavior, i. e., $A(s, 0) \to s^\alpha + O(1)$ as $s \to \infty$.

I. INTRODUCTION

IN this paper we state and prove a theorem concerning asymptotic behavior of forward elastic scattering amplitudes in relativistic theories.

The theorem, modeled closely upon the work of Regge[1] in potential scattering, predicts a similar behavior for the relativistic amplitude. However, restrictive assumptions must be made which do not allow the theorem to be applied to realistic (i.e., divergent) field theories, but only to appropriately cut-off versions.

More specifically, we shall study the elastic scattering amplitude by means of a Bethe–Salpeter equation in the crossed channel (see Fig. 1 and below). Our conclusion is that if the kernel is sufficiently nice—i.e., has properties sufficiently similar to those of potential scattering—Regge behavior of the scattering amplitude is guaranteed.[2] The theorem will quite precisely define what is meant by "sufficiently nice". Unfortunately, realistic field theories (aside possibly from $G\phi^3$) do not satisfy the criteria.

In Sec. II we state the theorem and outline the proof. Section III is devoted to details of the calculation. In Sec. IV we discuss the conclusions.

II. THE THEOREM

We first state the assumptions entering into the theorem:

(A) We consider an off-shell forward scattering amplitude $A(k, q)$ for identical spinless particles of incident momenta[3] k and $-q$. We assume these

* "This work is supported in part through the U. S. Atomic Energy Commission Contract AT(30-1)-2098, by funds provided by the U. S. Atomic Energy Commission."
† A. P. Sloan Foundation Fellow.
[1] T. Regge, Nuovo Cimento **14**, 951 (1959); **18**, 947 (1960).
[2] B. W. Lee and R. F. Sawyer, Phys. Rev. **127**, 2266 (1962), have established Regge behavior for the Bethe–Salpeter equation in ladder approximation.
[3] Our metric is $g_{\mu\mu} = \{-1, 1, 1, 1\}$; $\mu = 0, 1, 2, 3$.

Equation (34) and two of its solutions correspond to the case studied by Goldstein[3] [compare Eq. (22) and the solution (23) of his paper if $(1-\mu)$ is replaced by 1 in the above Eq. (36).]

The equation which results for the eigenvalue -1 of M which reads

$$\left\{\left[\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)+1\right]\left[-\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)-1\right]-\frac{i\alpha}{2\pi\rho^2}(1-\sigma_3)\right\}\chi' \equiv -\left\{\frac{\partial^2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial}{\partial\rho}+\frac{1}{\rho^2}\times\left[-\mu+\frac{i\alpha}{\pi}(1-\mu)\right]+1+2\sigma_1\left(\frac{\partial}{\partial\rho}+\frac{\mu}{\rho}\right)\right\}\chi^1=0 \tag{37}$$

is certainly more complicated however, at least because there appear two noncommuting matrices σ, and μ.

ACKNOWLEDGMENT

Cordial thanks are extended to Dr. Leo C. Levitt for his help in materially improving the clarity and conciseness of this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 2 FEBRUARY 1964

Regge Behavior of Forward Elastic Scattering Amplitudes*

JAMES D. BJORKEN†
Stanford University, Stanford, California
(Received 7 August 1963)

A theorem concerning the asymptotic behavior of forward elastic scattering amplitudes in relativistic theories is stated and proved. The assumptions made are (1) identical spinless particles interact via $G\phi^3$ and $\lambda\phi^4$ couplings; (2) a cutoff of the propagators is introduced; (3) the forward scattering amplitude satisfies a Bethe–Salpeter equation in the crossed channel; (4) the kernel of the equation is an arbitrary finite subset of the Feynman graphs which compose the exact kernel. The theorem states that under these assumptions, the forward scattering amplitude exhibits Regge behavior, i. e., $A(s, 0) \to s^\alpha + O(1)$ as $s \to \infty$.

I. INTRODUCTION

IN this paper we state and prove a theorem concerning asymptotic behavior of forward elastic scattering amplitudes in relativistic theories.

The theorem, modeled closely upon the work of Regge[1] in potential scattering, predicts a similar behavior for the relativistic amplitude. However, restrictive assumptions must be made which do not allow the theorem to be applied to realistic (i.e., divergent) field theories, but only to appropriately cut-off versions.

More specifically, we shall study the elastic scattering amplitude by means of a Bethe–Salpeter equation in the crossed channel (see Fig. 1 and below). Our conclusion is that if the kernel is sufficiently nice—i.e., has properties sufficiently similar to those of potential scattering—Regge behavior of the scattering amplitude is guaranteed.[2] The theorem will quite precisely define what is meant by "sufficiently nice". Unfortunately, realistic field theories (aside possibly from $G\phi^3$) do not satisfy the criteria.

In Sec. II we state the theorem and outline the proof. Section III is devoted to details of the calculation. In Sec. IV we discuss the conclusions.

II. THE THEOREM

We first state the assumptions entering into the theorem:

(A) We consider an off-shell forward scattering amplitude $A(k, q)$ for identical spinless particles of incident momenta[3] k and $-q$. We assume these

* "This work is supported in part through the U. S. Atomic Energy Commission Contract AT(30-1)-2098, by funds provided by the U. S. Atomic Energy Commission."
† A. P. Sloan Foundation Fellow.
[1] T. Regge, Nuovo Cimento **14**, 951 (1959); **18**, 947 (1960).
[2] B. W. Lee and R. F. Sawyer, Phys. Rev. **127**, 2266 (1962), have established Regge behavior for the Bethe–Salpeter equation in ladder approximation.
[3] Our metric is $g_{\mu\mu} = \{-1, 1, 1, 1\}$; $\mu = 0, 1, 2, 3$.

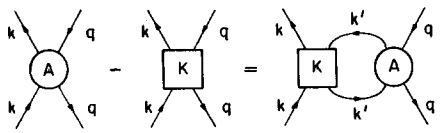


FIG. 1. Bethe–Salpeter equation for $A(k, q)$.

particles interact with themselves via $G\phi^3$ and $\lambda\phi^4$ couplings.

(B) The theory is rendered convergent, order by order, by replacing all propagators $(p^2 + m^2)^{-1}$ by $(p^2 + m^2)^{-1-\epsilon}$ with $\epsilon > 0$, and including a mass counterterm if $\epsilon < \frac{1}{3}$.

(C) The exact amplitude satisfies a Bethe–Salpeter equation in the crossed channel (see Fig. 1),

$$A(k, q) = \tilde{K}(k, q) - i \int \frac{d^4k' \tilde{K}(k, k')A(k', q)}{(2\pi)^4(k'^2 + m^2)^{2+2\epsilon}}. \quad (1)$$

Here $\tilde{K}(k, q)$ is the set of all Feynman graphs with no two-particle states in that crossed channel which has zero total energy. We shall approximate $\tilde{K}$ by any *finite* subset of the graphs composing the exact $\tilde{K}$. With these assumptions we can prove the following theorem:

Theorem. Let $s = (k - q)^2$, $u = (k + q)^2$. Then for $k^2 = q^2 = -m^2$,

$$A(s, k^2, q^2) \to O(1) + \sum_i \beta_i(k^2)\beta_i(q^2)[s^{\alpha_i} + u^{\alpha_i}] \quad (\alpha_i > 0), \quad (2)$$

when $s \to \infty$.

The terms in the sum will be identifiable as contributions from Regge poles.

We now outline the steps involved in the proof:

(1) Remove any point couplings (as in Fig. 2) from $\tilde{K}$. They will be put back in at the end.

(2) Choose the coupling constants G and λ sufficiently small so that there are no Regge poles in (2) and that the perturbation expansion of (1) converges. Strictly speaking, if $A \sim s^\alpha$ $(\alpha > 0)$ the integral on the right hand side of (1) diverges, and the equation does not make sense. The amplitude in general must be defined by analytic continuation in the coupling constants G and λ of the perturbation solution of (1).

(3) Continue equation (1) to values of k^2, q^2, and $k \cdot q$ corresponding to a Euclidean space[4] of spacelike vectors, i.e. $k^2 > 0$, $q^2 > 0$, and

$$-1 \leq \cos\theta = k \cdot q/[k^2 q^2]^{\frac{1}{2}} \leq +1 \quad (3)$$

This continuation may be done as follows: The $i\epsilon$'s in Feynman graphs allow any Feynman integral to be defined for all real k and q—in particular, when $k = (0, \mathbf{k})$ and $q = (0, \mathbf{q})$, which satisfy our specifications. To change the Lorentz to Euclidean metric, we need only rotate the contours of all intermediate energy integrations in the graphs up to the imaginary axis. This may be done graph by graph by (a) combining denominators using Feynman parameters; (b) rotating contours of the intermediate energy integrations to the imaginary axis using standard arguments (no completion of the square is needed because k and q are pure spacelike); and (c) undoing the Feynman parameterization. Provided perturbation theory converges (as it will in our model for small G and λ), this defines also $A(k, q)$ in the spacelike region by the equation

$$A(k, q) = K(k, q) + \int \frac{d^4k' K(k, k')A(k', q)}{(2\pi)^4(k'^2 + m^2)^{2+2\epsilon}}, \quad (4)$$

where the integral is now over 4-dimensional *Euclidean* space.

(4) Expand $A(k, q) = A(k^2, q^2, \cos\theta)$ in a series of orthogonal functions of $\cos\theta$ appropriate to the 4-dimensional symmetry of the problem; i.e., the 4-dimensional analogue of Legendre functions. These are Gegenbauer polynomials of the first kind:

$$C_n(\cos\theta) \equiv \sin(n + 1)\theta/\sin\theta.$$

So we write

$$A(k^2, q^2, \cos\theta) = \sum_{n \text{ even}} A_n(k^2, q^2)C_n(\cos\theta),$$
$$K(k^2, q^2, \cos\theta) = \sum_{n \text{ even}} K_n(k^2, q^2)C_n(\cos\theta). \quad (5)$$

The sum is over only even n provided $K(k, q) = K(k, -q)$, a property of the exact kernel due to crossing symmetry.

(5) Using this expansion, separate the integral equation (4) and solve for the $A_n(k^2, q^2)$ by the Fredholm method. This expresses $A_n(k^2, q^2)$ as the ratio of two power series, each of which is an entire function of G and λ.

(6) Continue the expansion back to $k^2 = q^2 = -m^2$ with $\cos\theta$ still in the interval $(-1, 1)$. By direct inspection it will be seen that the Fredholm

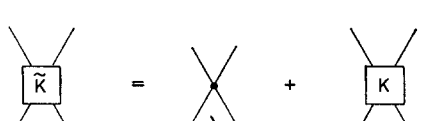


FIG. 2. Decomposition of the kernel K.

[4] G. C. Wick, Phys. Rev. **96**, 1124 (1954); R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954).

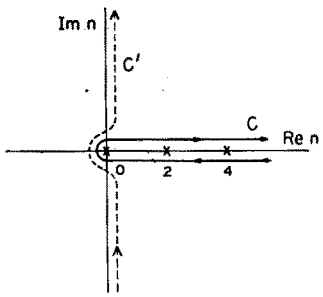


FIG. 3. Regge contour in the n plane.

expansion still exists and converges after this continuation.

(7) To get to high energies s, it is necessary to discuss the behavior of $A(\cos\theta)$ for large $\cos\theta$. For this we use the Watson transform *a la* Regge. By using a perturbation-theory type of integral representation[5] for $K(k, q)$, it is possible to explicitly extend the definition of $K_n(k^2, q^2)$ to complex values of n. Indeed we shall construct an interpolation of $K_n(k^2, q^2)$ for all n satisfying $\mathrm{Re}\ n \geq 0$ such that $K_n(k^2q^2) < M$ with M a fixed constant. Then using the Fredholm expansion, we construct $A_n(k^2, q^2)$ for all n in the right half-plane; in addition we show that

$$A_n(k^2, q)^2 \to K_n(k^2, q^2) \quad \text{as} \quad |n| \to \infty. \tag{6}$$

(8) Using the Watson transform and the above construction, we write (for $|\cos\theta| < 1$)

$$A(k^2, q^2, \cos\theta) = -\frac{1}{4i}\int_C \frac{dn A_n(k^2, q^2)}{\sin \pi n} \times [C_n(\cos\theta) + C_n(-\cos\theta)], \tag{7}$$

with the contour C illustrated in Fig. 3. For $|\cos\theta| < 1$ and $n \to \infty$

$$C_n(\cos\theta) \sim e^{|\mathrm{Im}\, n|\theta} \qquad (0 \leq \theta \leq \pi)$$
$$A_n < M$$

and therefore the contour C may be distorted to the imaginary axis and replaced by C'.

(9) With this replacement, we may let $\mathrm{Im}\ \theta$ run into the complex plane with $0 < \mathrm{Re}\ \theta < \pi$, and the integral (7) with C replaced by C' will still converge. Thus, $A(k^2, q^2, \cos\theta)$ is defined for all $\cos\theta$, with the exception of cuts running from $\cos\theta = \pm 1$ to ∞, and A is bounded by a constant in this region.

(10) We now let the coupling constants G and λ become large. The Fredholm expansions are seen to still converge for all n in the right half-plane; the only changes are possible zeros in the Fredholm determinant in the denominator of the expansion for A_n, giving rise to poles in the n plane which lie on the positive real axis. These are associated with bound-state solutions of the equation, and therefore behave like Regge poles, e.g., satisfy the factorization condition. They contribute terms to $A(k^2, q^2, \cos\theta)$ of the structure

$$\beta_i(k^2)\beta_i(q^2)[C_{n_i}(\cos\theta) + C_{n_i}(-\cos\theta)] \sim \beta_i(k^2)\beta_i(q^2)[s^{n_i} + u^{n_i}],$$

with $s = (k - q)^2$ and $u = (k + q)^2$.

(11) Finally, the additional s-wave point interactions may be included. Let us denote the complete kernel by $\bar{K}$, i.e.,

$$\bar{K}(k, q) = K(k, q) + \lambda.$$

Then, denoting A and $\bar{A}$ as the solutions of (1) with K and $\bar{K}$, respectively, as kernels, we may write

$$\bar{A} = A + A',$$

where A' satisfies the equation

$$A'(k, q) = \lambda + \int \frac{d^4k'[\lambda A(k', q) + \lambda A'(k', q) + K(k, k')A'(k', q)]}{(2\pi)^4(k'^2 + m^2)^{2+2\epsilon}}. \tag{8}$$

This equation is separable and admits the solution [easily constructed by drawing some pictures and verified by substitution into (8)]:

$$A'(k, q) = \lambda\Gamma(k^2)\Gamma(q^2)/(1 - \lambda J), \tag{9}$$

where

$$\Gamma(k^2) = 1 + \int \frac{d^4k' A(k, k')}{(2\pi)^4(k'^2 + m^2)^{2+2\epsilon}} = 1 + \int \frac{d^4k' A(k', k)}{(2\pi)^4(k'^2 + m^2)^{2+2\epsilon}},$$
$$J = \int \frac{d^4k\Gamma(k^2)}{(2\pi)^4(k^2 + m^2)^{2+2\epsilon}}. \tag{10}$$

A' is independent of s and therefore contributes only to $O(1)$ in (2). This completes the outline of the proof.

III. DETAILS OF THE PROOF

There are three areas in the proof which require

[5] Y. Nambu, Nuovo Cimento 6, 1064 (1957); K. Symanzik, Progr. Theoret. Phys. (Kyoto) 20, 690 (1957).

discussion: (a) The first is the question of the separation of the Bethe–Salpeter equation and the properties of the Gegenbauer polynomials. Then we discuss the radial equation and the formal construction of the Fredholm solution. (b) In order to justify the procedures of analytic continuation used in the proof and to establish the convergence of the Fredholm expansion, we shall need a number of properties of the kernel $K(k, q)$. These we derive from its Feynman parameter integral representation. (c) Finally using properties of K derived in (b) and the Fredholm expansion constructed in (a), we prove the properties of $A(k, q)$ used in the proof of Sec. II.

(a) It is straightforward to separate the Bethe–Salpeter equation (4) using the expansions (5) and the following lemma.

Lemma 1: Defining a solid angle $d\Omega$ by $\int d^4k = \int_0^\infty k^3\, dk \int d\Omega$, the C_n functions satisfy

$$I_{nm} = \int d\Omega_n C_n(\hat{k}_f \cdot \hat{k}_n) C_m(\hat{k}_n \cdot \hat{k}_i) = \frac{2\pi^2 \delta_{nm} C_n(\hat{k}_f \cdot \hat{k}_i)}{n+1}. \tag{11}$$

This is proved in analogy to the corresponding proof for Legendre functions; it is outlined in the Appendix. One then finds the radial equation

$$A_n(k^2, q^2) = K_n(k^2, q^2) + \frac{1}{n+1} \int_0^\infty \frac{dk'^2 k'^2 K_n(k^2, k'^2) A_n(k'^2, q^2)}{16\pi^2 (k'^2 + m^2)^{2+2\epsilon}}, \tag{12}$$

for which the formal solution is

$$A_n(k^2, q^2) = \frac{K_n(k^2, q^2) + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!\,(n+1)^m} \left[\prod_{i=1}^{m} \int_0^\infty \frac{k_i^2\, dk_i}{16\pi^2 (k_i^2 + m^2)^{2+2\epsilon}} \right] \bar{\Delta}_n(k, q; k_1, \cdots k_m)}{1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!\,(n+1)^m} \left[\prod_{i=1}^{m} \int_0^\infty \frac{k_i^2\, dk_i^2}{16\pi^2 (k_i^2 + m^2)^{2+2\epsilon}} \right] \Delta_n(k_1 \cdots k_m)}$$

$$\bar{\Delta}_n(k, q; k_1, \cdots k_m) = \begin{vmatrix} K_n(k, q) & K_n(k_1, q) & \cdots & K_n(k_m, q) \\ K_n(k, k_1) & & & \\ \vdots & & & \vdots \\ K_n(k, k_m) & & \cdots & K_n(k_m, k_m) \end{vmatrix}, \tag{13}$$

$$\Delta_n(k_1, \cdots k_m) = \begin{vmatrix} K_n(k_1, k_1) & \cdots & K_n(k_m, k_1) \\ \vdots & & \vdots \\ K_n(k_1, k_m) & \cdots & K_n(k_m, k_m) \end{vmatrix}.$$

We must still show this expansion converges for $k^2 \geq m^2$, $q^2 \geq m^2$ and Re $n \geq 0$. To do this we need properties of K_n.

(b) The information about K_n will be derived from the following integral representation:

$$K(k, q) = \sum_i \int_0^\infty d\zeta_1 \cdots d\zeta_4 \int_0^\infty dM^2 \times \frac{\tilde{\rho}_i(\zeta_1 \cdots \zeta_4, M^2)}{[\zeta_1 k^2 + \zeta_2 q^2 + \zeta_3 (k+q)^2 + \zeta_4 (k-q)^2 + M^2]^{e_i}}. \tag{14}$$

This representation was first derived by Nambu.[5]

The sum runs over a finite number of terms, each one corresponding to an individual Feynman graph. For simplicity, we shall hereafter suppress the sum. Symanzik[5] showed that when $k^2 = -m^2$, $q^2 = -m^2$, $|\cos\theta| \leq 1$, the denominator in (14) remains positive semidefinite; therefore (14) may be rewritten as

$$K = \int_0^\infty d\zeta_1 \cdots d\zeta_4\, dM^2 \times \frac{\rho(\zeta_1 \cdots \zeta_4, M^2)\theta(\zeta_3 - \zeta_4)}{[(\zeta_1 + \zeta_3 + \zeta_4)(k^2 + m^2) + (\zeta_2 + \zeta_3 + \zeta_4)(q^2 + m^2) + 2(\zeta_3 - \zeta_4)[(k^2 q^2)^{\frac{1}{2}} \cos\theta + m^2] + M^2]^e} + (\cos\theta \leftrightarrow -\cos\theta), \tag{15}$$

where crossing symmetry has been used.

We may draw the conclusions that:

(1) For positive k^2 and q^2, K is analytic in the cut cos θ plane. It satisfies a dispersion relation of the form [obtained by a change of variables in (15)]

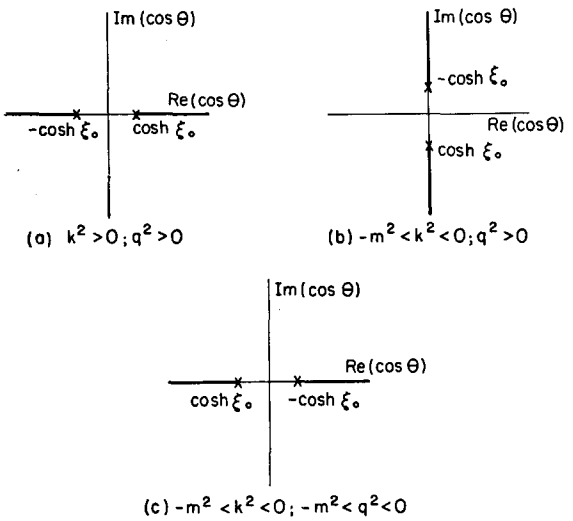


FIG. 4. Singularities in the cos θ plane.

$$K(k^2, q^2, \cos\theta) = \int_{\xi_0(k^2,q^2)}^{\infty} d\xi \frac{\sigma(\xi, k^2, q^2)}{[\cosh\xi + \cos\theta]^\epsilon} + (\cos\theta \leftrightarrow -\cos\theta), \quad (16)$$

with $\xi_0 > 0$.

(2) An analytic continuation from spacelike to timelike k^2 is guaranteed by the representation, provided $|k^2| < m^2$ and $|q^2| < m^2$. This is because the real part of the denominator never vanishes under these conditions. Thus by letting $k^2 \to k^2 e^{i\phi}$, with ϕ running from 0 to π, we can continue back to the timelike region. The cuts in the cos θ plane migrate as shown in Fig. 4. Thus, once we establish the solution for spacelike k^2 and q^2, the integral representation for K gives us a chance of continuing the solution back to physics, namely $k^2 = q^2 = -m^2$.

(3) Because $K(k^2, q^2, \cos\theta)$ exists for $k^2 \geq -m^2$, $q^2 \geq -m^2$, $|\cos\theta| \leq 1$,

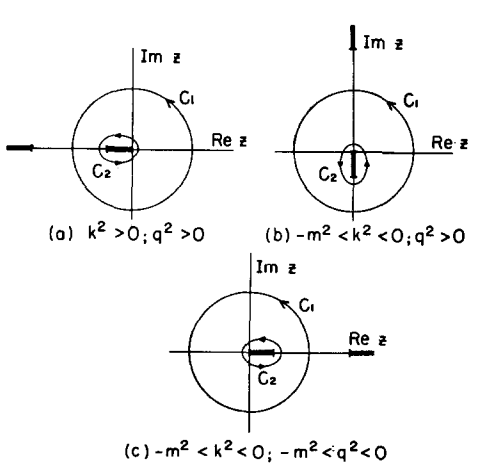


FIG. 5. Distortion of contours in the z plane.

$$\int_{\xi_0}^{\infty} d\xi \frac{\sigma(\xi, k^2, q^2)}{[\cosh\xi - 1]^\epsilon} < M. \quad (17)$$

(4) σ and the weight function ρ are positive definite. This follows from the way in which the Nambu representation[5] is constructed from the original Feynman integral.

(5) From inspection of (15),

$$|\cosh\xi_0| > 1 \quad \text{for} \quad k^2 > -m^2, \quad q^2 > -m^2. \quad (18)$$

We may now construct $K_n(k^2, q^2)$. From the expansion (5), we deduce

$$K_n(k^2, q^2) = \frac{1}{\pi}\int_0^{2\pi} d\theta \sin\theta \sin(n+1)\theta K(k^2, q^2, \cos\theta)$$
$$= \frac{1}{\pi}\,\mathrm{Im}\int_0^{2\pi} d\theta \sin\theta e^{i(n+1)\theta} K(k^2, q^2, \cos\theta)$$
$$= \frac{-\mathrm{Im}}{2\pi}\int_{C_1}\frac{dz}{z}(z - z^{-1})z^{n+1}K\left(k^2, q^2, \frac{z+z^{-1}}{2}\right)$$
$$(k^2 > 0,\ q^2 > 0,\ n = 0, 2, 4\cdots), \quad (19)$$

where the contour C_1 is the unit circle in the z plane. Using the integral representation (16), we may shrink the contour C_1 to the contour C_2 (Fig. 5) and find:

$$K_n(k^2, q^2) = \frac{1}{\pi}\,\mathrm{Im}\int_{\xi_0}^{\infty} d\xi\sigma(\xi, k^2, q^2)\int_{C_2}\frac{dz\, z^{n-1}(1-z^2)(2z)^\epsilon}{(z+e^\xi)^\epsilon(z+e^{-\xi})^\epsilon}$$
$$= \frac{2}{\pi}\sin\pi\epsilon\int_{\xi_0}^{\infty} d\xi\sigma(\xi, k^2, q^2)2^\epsilon e^{-(n+\epsilon)\xi}F(n, \xi), \quad (20)$$

where

$$F(n, \xi) = \int_0^1 dx \frac{x^{n+\epsilon-1}(1 - e^{-2\xi}x^2)}{(1-x)^\epsilon(1 - e^{-2\xi}x)^\epsilon}. \quad (21)$$

It follows that

$$|F(n, \xi)| < e^{\epsilon\xi}/(\epsilon + \mathrm{Re}\, n)(1+\epsilon)2^\epsilon(\sinh\xi)^\epsilon, \quad (22)$$

and consequently, with the help of (22) and (17),

$$|K_n(k^2, q^2)| < \frac{2}{(\epsilon + \mathrm{Re}\, n)} \times \int_{\xi_0}^{\infty} d\xi \frac{|\sigma(\xi, k^2, q^2)|\, e^{-\xi\,\mathrm{Re}\, n}}{(\sinh\xi)^\epsilon} < \frac{2M}{\epsilon + \mathrm{Re}\, n}. \quad (23)$$

These relations (20) and (23) hold for all spacelike k^2 and q^2 and for all n in the right half-plane, and provide us with the desired continuation of K_n into the complex n plane.

We may now check that nothing unpleasant

happens as k^2 and q^2 are made timelike. We find that the cuts in the z plane migrate as shown in Fig. 5(b). The end point of the short cut is now at $z = -ie^{-\chi}$, where

$$\sinh \chi = \cosh \xi.$$

The same kind of analysis can be made as before, with the result that

$$K_n(k^2, q^2) = e^{\frac{1}{2}(i\pi n)} \tilde{K}_n(k^2, q^2),$$

with $|\tilde{K}_n(k^2, q^2)| < 4M/(\epsilon + \text{Re } n)$ for all n in the right half-plane.

As q^2 becomes timelike, the cuts migrate to the positions shown in Fig. 5(c). Another factor of $e^{\frac{1}{2}(i\pi n)}$ appears, which, together with the first such factor, can now be ignored, since $e^{i\pi n} = 1$ for all even n. Thus by an integral representation similar to (20), $K_n(k^2, q^2)$ is constructed for $0 > k^2 > -m^2$, $0 > q^2 > -m^2$; $\text{Re } n \geq 0$ such that

$$|K_n(k^2, q^2)| < 2M/(\epsilon + \text{Re } n)$$

as before.

We may now summarize what we have done. We have found functions $K_n(k^2, q^2)$ and $K_n(k^2, q^2)$ such that for n an even integer, $K_n(k^2, q^2)$ agrees with the partial-wave projections of $K(k, q)$ if k^2 and q^2 are both spacelike or both timelike and $> -m^2$. For k^2 timelike and q^2 spacelike, and n an even integer, $K_n(k^2, q^2)$ agrees with the partial-wave projection of $K(k, q)$ except for a factor $(-1)^{\frac{1}{2}n}$. In addition, the functions K_n and $\tilde{K}_n$ are analytic in the right half n plane and bounded by a constant.

(c) We now can study the full partial-wave amplitude. We first consider even n. Since K_n is bounded by a constant independently of whether k^2 and q^2 are spacelike or timelike (or indeed anywhere in between in the continuation), we may use Hadamard's lemma on the magnitude of the Fredholm determinants to ensure the convergence to the power series in (13) defining A_n. In particular,

$$|m\text{th term}| < \left|\frac{M}{\epsilon + \text{Re } n}\right|^m \frac{m^{\frac{1}{2}n}}{m!} \left[\int_0^\infty \frac{k^2\, dk^2}{(k^2 + m^2)^{2+2\epsilon}}\right]^m,$$

which is sufficient to guarantee convergence for all G and λ.

From this we conclude that, for even integer n, $A_n(-m^2, -m^2)$ is given by the convergent Fredholm expansion (13) and represents the partial-wave projection of the scattering amplitude according to the expansion (5).

Now we observe that $A_n(-m^2, -m^2)$ may be continued into the right half n plane. This is done by first replacing all $K_n(-m^2, k_i^2)$ and $\tilde{K}_n(k_i^2, -m^2)$. Although each such replacement should be accompanied by a factor $(-1)^{\frac{1}{2}n}$, there are always an even number of $\tilde{K}$'s in any term in the expansion of the determinants, and so these factors may fortunately be omitted. There is no longer any obstacle in extending A_n into the right half n plane, since the K_n and $\tilde{K}_n$ have the same (or better) convergence properties that they possess when even integers. Furthermore, because of the factor $(n+1)^{-1}$ in the radial equation, it is clear that as $|n| \to \infty$ (in the right half-plane), $A_n \to K_n < 2M/(\epsilon + \text{Re } n)$.

IV. CONCLUSIONS

As we remarked in the beginning, the essence of the theorem is that, at least for forward scattering, Regge behavior of an elastic amplitude is assured provided the kernel K in the Bethe–Salpeter is "nice enough". We may now define more precisely what a "nice" kernel is:

(1) K should have the general analytic properties of Feynman graphs. There may be some mass restrictions which eliminate cases with anomalous or superanomalous thresholds, but reasonable kinematics such as encountered in $\pi\pi$, πN, and NN scattering appear to possess, in general, sufficiently good analytic properties for the purpose of studying asymptotic behavior via integral transforms.

(2) K must have good asymptotic behavior; the crucial condition appears to be $\text{tr } K < \infty$. In our case,

$$\text{tr } K \sim \int \frac{d^4k\, K(k, k)}{(k^2 + m^2)^{2+2\epsilon}}.$$

This ensures that the Fredholm expansion makes sense.

In order to see the limitations of the proof we have given, it is instructive to see what happens when some of the initial assumptions are relaxed. For instance, if we omit the convergence factor introduced into the propagators, it is possible to obtain a different kind of asymptotic behavior. This has been investigated in detail by Sawyer[6] for the kernel illustrated in Fig. 6. If the divergent

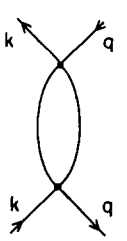


FIG. 6. Sawyer's kernel.

[6] R. F. Sawyer (to be published).

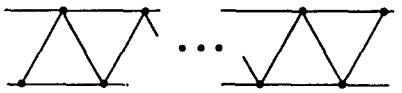


FIG. 7. Truss bridge graphs.

kernel is cut off, Lee and Sawyer[2] found Regge behavior, consistent with the results of our theorem. However, if the kernel is *renormalized*, then asymptotic behavior of the form $s^{\alpha}(\log s)^{-\frac{1}{2}}$ is obtained instead.

The effect of relaxing the assumption that K consist of a *finite* set of graphs has been studied by the author and T. T. Wu.[7] An infinite set of graphs (Fig. 7) for K is asymptotically summed, and the asymptotic behavior of the sum is again $s^{\alpha}(\log s)^{-\frac{1}{2}}$.

Thus there exist counter examples which indicate that the main assumptions which enter the proof cannot be easily relaxed. We are thus forced to conclude that the results of this theorem shed little light on the behavior of the interesting, highly singular, scattering amplitudes of quantum field theory.

V. ACKNOWLEDGMENTS

It is a pleasure to thank N. N. Khuri; his great understanding of the corresponding nonrelativistic problem provided constant guidance in the construction of this theorem. The author thanks Dr. J. R. Oppenheimer and the Institute for Advanced Study, Princeton, New Jersey, for their hospitality, and also the hospitality of the M.I.T. Laboratory for Nuclear Science, Cambridge, Massachusetts, where this work was completed.

[7] J. D. Bjorken and T. T. Wu, Phys. Rev. **130**, 2566 (1963).

APPENDIX. ORTHOGONALITY OF GEGENBAUER POLYNOMIALS

To prove (11), we choose

$$\hat{k}_i = (1, 0, 0, 0),$$

$$\hat{k}_f = (\cos\theta, \sin\theta, 0, 0),$$

$$\hat{k}_n = (\cos\tilde{\theta}, \sin\tilde{\theta}\cos\chi, \sin\tilde{\theta}\sin\chi\cos\varphi, \sin\tilde{\theta}\sin\chi\sin\varphi). \tag{A1}$$

Then

$$\int d\Omega_n = \int_0^{\pi} d\tilde{\theta}\sin^2\tilde{\theta}\int_0^{\pi} d\chi\sin\chi\int_0^{2\pi} d\varphi. \tag{A2}$$

We first consider the case $m > n$. From parity considerations, (11) is zero unless $m - n$ is even, so that we need only consider $m > n + 1$. Then

$$I_{nm} = 2\pi\int_0^{\pi} d\tilde{\theta}\sin\tilde{\theta}\int_0^{\pi} d\chi\sin\chi\, C_n(\cos\tilde{\theta}\cos\theta + \sin\tilde{\theta}\sin\theta\cos\chi)\sin(m+1)\tilde{\theta}. \tag{A3}$$

Since $C_n(z)$ is a polynomial in z, after doing the χ integration, we will be left with a polynomial of degree n in $\cos\tilde{\theta}$ multiplying $\sin\tilde{\theta}\sin(m+1)\tilde{\theta}$; since $m > n + 1$, the result of the integration will be zero.

To prove (11) for $m = n$, we observe that I_{nn} will be a polynomial in $\cos\theta$; hence,

$$I_{nn} = \sum_{m=0}^{2n} A_m C_m(\cos\theta). \tag{A4}$$

By left multiplying by $C_n(\cos\theta)$ and integrating over the angles of $\hat{k}_f$, we find

$$A_m = 0, \qquad m \neq n. \tag{A5}$$

To find a_n, we set $n = m$ and $\hat{k}_f = \hat{k}_i$ in (11); an elementary integration gives the result.

Diffraction by Polygonal Cylinders*

BURT JULES MORSE
Courant Institute of Mathematical Sciences, New York University, New York, New York†

Asymptotic solutions for the two-dimensional reduced wave equation in the exterior of a convex polygonal cylinder with both Dirichlet and Neumann boundary conditions are obtained. The method used is the geometrical theory of diffraction. Both the cases of oblique and grazing incidence are treated, and the far fields in all directions including those along shadow and specular lines are found. A feature of the method is that the solutions may be carried to any desired order. The calculation is illustrated for the case of a rectangular cylinder and, for this geometry, cross sections, surface current densities and radiation patterns are obtained. Graphs of some of these results have been included.

1. INTRODUCTION

IN this paper we treat diffraction by infinite cylinders of convex polygonal cross section using the geometrical theory of diffraction. This amounts to determining asymptotic expansions of solutions of the two-dimensional reduced wave equation in the exterior of a convex polygonal domain. An E or H polarized plane wave is incident with wavelength small compared to the length of any side of the polygon. Either the total field or its normal derivative is required to vanish on the boundary. The time dependence factor $e^{-i\omega t}$ is suppressed.

A survey of the geometrical theory of diffraction and the mathematical basis for its application to polygonal cylinders are contained in articles by Keller.[1,2] Since this theory has already been applied to problems closely related to ours,[3,4] it is relevant to indicate the lacunas in those treatments which we have attempted to fill. Essentially, these concern the determination of shadow-boundary fields and the calculation of higher-order terms in the asymptotic representation of the field. Thus, according to the geometrical theory of diffraction, the diffracted field $u_d(P)$ at a point P is equal to the sum of the fields on all rays through P:

$$u_d(P) = \sum_{\text{rays}} u_j(P). \tag{1.1}$$

Here $u_j(P)$ is the diffracted field on the jth such ray, and if this is an m-fold diffracted ray

$$u_j(P) \sim \frac{e^{iks_j}}{k^{\frac{1}{3}m}} \sum_{n=0}^{\infty} (ik)^{-n} A_{jn}(P), \tag{1.2}$$

where k is the propagation constant, s_j the arclength along the ray and the functions A_{jn} depend on the geometry of the diffracting objects. The gaps in previous work are due to the facts that only the leading term in the series (1.2) was determined and the form of this term was not valid along shadow boundaries. A form valid there is required in problems with grazing incidence and in order to compute the field in directions parallel to sides of cylinders.

For a polygonal cylinder all the diffracted rays are produced by wedges of various vertex angles. Hence the inclusion of higher $A_{jn}(n = 1, 2, \cdots)$ in the expansion (1.2) involves the use of more terms in the asymptotic solution of the wedge-diffraction problem. The calculation of the diffraction coefficients corresponding to these higher-order terms necessitates the solution of the wedge problem for nonplane-wave incidence. However, as is shown in the work of Zitron and Karp,[5] the relevant nonplane waves are expressible in terms of linear combinations of plane waves and their derivatives. Thus the diffraction coefficients are easily found from those for plane waves.

To calculate diffraction coefficients corresponding to shadow-boundary fields, we show that these fields too are expandable in terms of plane waves and their derivatives. Finally, we treat the problem of finding the far fields along shadow boundaries, which are in directions parallel to cylinder sides or which are not. The latter are easily computed by expanding the wedge solution near shadow boundaries. In the former case we employ Green's theorem for distances large compared to the cylinder dimensions. In this

* This research was supported by the Office of Naval Research, under Contract No. Nonr 285(48).

† Present address: The University of New Mexico, Albuquerque, New Mexico.

[1] J. B. Keller, J. Opt. Soc. Am. **52**, 116 (1962).

[2] J. B. Keller in *Electromagnetic Waves*, edited by R. E. Langer (University of Wisconsin Press, Madison, Wisconsin, 1962), pp. 129–137.

[3] S. N. Karp and J. B. Keller, Courant Institute of Mathematical Sciences, New York University, Res. Rept. No. EM-143 (1959).

[4] J. E. Burke and J. B. Keller, Electronic Defense Laboratories, Mt. View, California, Res. Rept. No. EDL-E48 (1960).

[5] N. Zitron and S. N. Karp, Courant Institute of Mathematical Sciences, New York University, Res. Rept. No. EM-126 (1959).

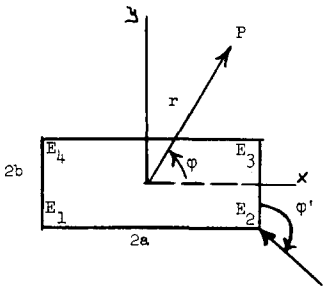


FIG. 1. Notation for diffraction by a rectangular cylinder. The incident angle is φ'. The polar coordinates of point P are (r, φ). The lengths of the sides are 2a and 2b. E_1, E_2, E_3, and E_4 designate the vertices.

connection as well as in many other places we make extensive use of the work of Oberhettinger.[6]

Section 2 contains a collection of results relating to diffraction of a plane wave by hard and soft rectangular cylinders. These include formulas and graphs for total scattering cross section and far fields. They agree with the results of Mei and Van Bladel[7] who treated the same problem by numerical methods. In Sec. 3 will be found the analysis of the rectangular cylinder problem. Finally, in Sec. 4 we summarize our results and show how they apply to a convex polygonal cylinder of arbitrary cross section.

2. SUMMARY OF THE RESULTS FOR RECTANGULAR CYLINDERS

For the case of the soft (i.e., $u=0$ on the boundary) rectangular cylinder at oblique incidence (see Fig. 1),

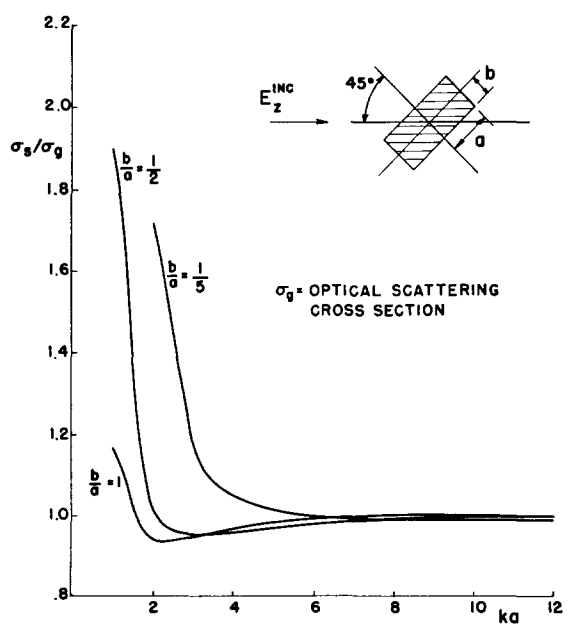


FIG. 2. The ratio of the total to the optical scattering cross section σ_s/σ_g as a function of ka at $\varphi' = \frac{3}{4}\pi$ for various values of b/a in the soft case (i.e., $u = 0$ on the boundary).

we find that the total scattering cross section σ_s is given by

$$\sigma_s = 4(b\sin\varphi' - a\cos\varphi') + \frac{8}{27k\pi^{\frac{1}{2}}}\left[\frac{\cos(kA_- + \frac{1}{4}\pi)}{(ka)^{\frac{3}{2}}} \times E(\tfrac{3}{2}\pi - \varphi')E(\varphi' + \tfrac{1}{2}\pi) + \frac{\cos(kB_+ + \frac{1}{4}\pi)}{(kb)^{\frac{3}{2}}}E(2\pi - \varphi')E(\varphi')\right] + O[k^{-1}(ka)^{-5/2}], \quad \tfrac{1}{2}\pi < \varphi' < \pi, \tag{2.1}$$

where $E(\theta) = (\frac{1}{2}+\cos\frac{2}{3}\theta)^{-2}\sin\frac{2}{3}\theta$, and $2a(1\pm\sin\varphi') = A_\pm$, $2b(1 \pm \cos\varphi') = B_\pm$. Figure 2 shows σ_s/σ_g, where σ_g is the optical scattering cross section, as a function of ka at $\varphi' = \frac{3}{4}\pi$ for various values of b/a. For the hard (i.e., $\partial u/\partial n = 0$ on the boundary) rectangle, the corresponding quantity, σ_H, is

$$\sigma_H = 4(b\sin\varphi' - a\cos\varphi') + \frac{4}{3k\pi^{\frac{1}{2}}}\left[\frac{\cos(kA_- - \frac{1}{4}\pi)}{(ka)^{\frac{1}{2}}}F(\tfrac{3}{2}\pi - \varphi')F(\varphi' + \tfrac{1}{2}\pi) + \frac{\cos(kB_+ - \frac{1}{4}\pi)}{(kb)^{\frac{1}{2}}}F(\varphi')F(2\pi - \varphi')\right] + O[k^{-1}(ka)^{-1}], \quad \tfrac{1}{2}\pi < \varphi' < \pi, \tag{2.2}$$

where

$$F(\theta) = (\tfrac{1}{2} + \cos\tfrac{2}{3}\theta)^{-1}.$$

Additional terms of (2.2) are included in (3.29). Figure 3 corresponds to Fig. 2 with σ_H replacing σ_s.

At grazing incidence, i.e., $\varphi' = \pi$, (2.1) and (2.2) are not valid. They are replaced by

$$\sigma_s = 4a + (4/k\pi^{\frac{1}{2}})(2kb)^{\frac{1}{2}} - (19/18k\pi^{\frac{1}{2}})(2kb)^{-\frac{1}{2}} + O[k^{-1}(ka)^{-\frac{3}{2}}], \tag{2.3}$$

$$\sigma_H = 4a + \frac{2^{\frac{3}{2}} + 4\sin(4kb + \frac{1}{4}\pi)}{27k\pi^{\frac{1}{2}}(kb)^{\frac{1}{2}}} - \frac{8\sin(2ka + \frac{1}{4}\pi)}{3k\pi^{\frac{1}{2}}(ka)^{\frac{1}{2}}} + O[k^{-1}(ka)^{-1}]. \tag{2.4}$$

Figure 4 shows $\sigma_s/4a$ as a function of ka for various values of b/a where σ_s is given by (2.3). Figures 5 and 6 correspond to Fig. 4 with σ_H given by (2.4) replacing σ_s.

We have also calculated the scattered far fields for the soft rectangle at oblique incidence, and the hard rectangle at grazing incidence. These results are illustrated in Figs. 7 and 8 in terms of the gain, $G(\varphi)$. Using the polar coordinates (r, φ) as shown in Fig. 1, $G(\varphi)$ is defined as the ratio of the power

[6] F. Oberhettinger, J. Math. and Phys. 34, 245 (1955).
[7] K. Mei and J. Van Bladel (to be published).

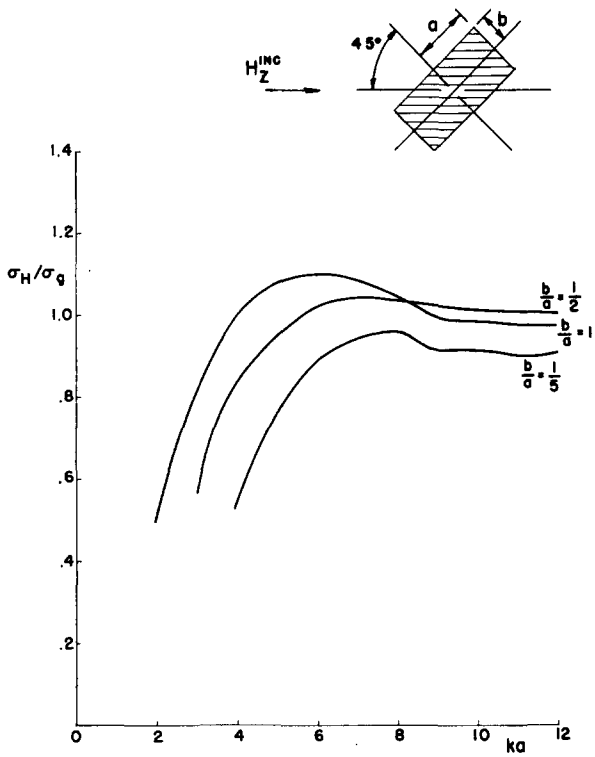


FIG. 3. The same data as in Fig. 2 in the hard case (i.e., $\partial u/\partial n = 0$ on the boundary).

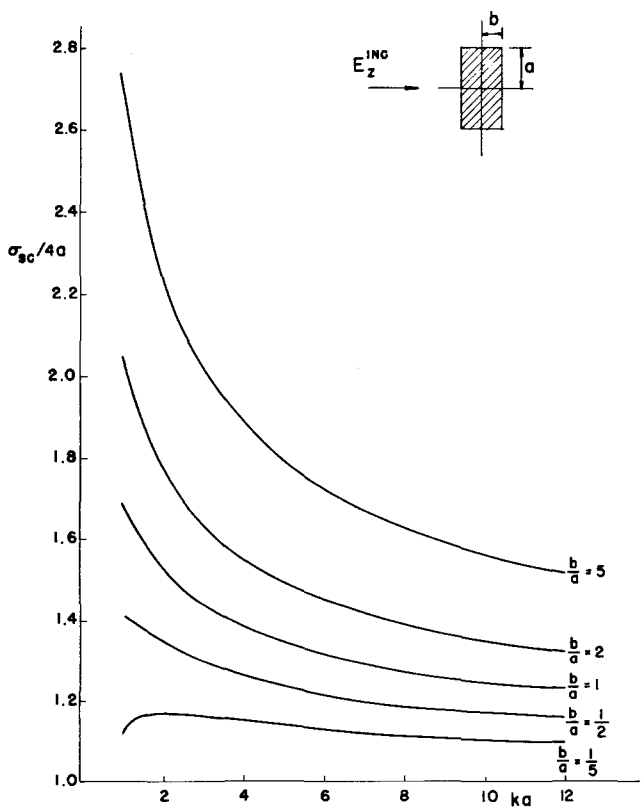


FIG. 4. The ratio of total to optical scattering cross section $\sigma_s/4a$ at grazing incidence, i.e., $\varphi' = \pi$, in the case of the soft cylinder.

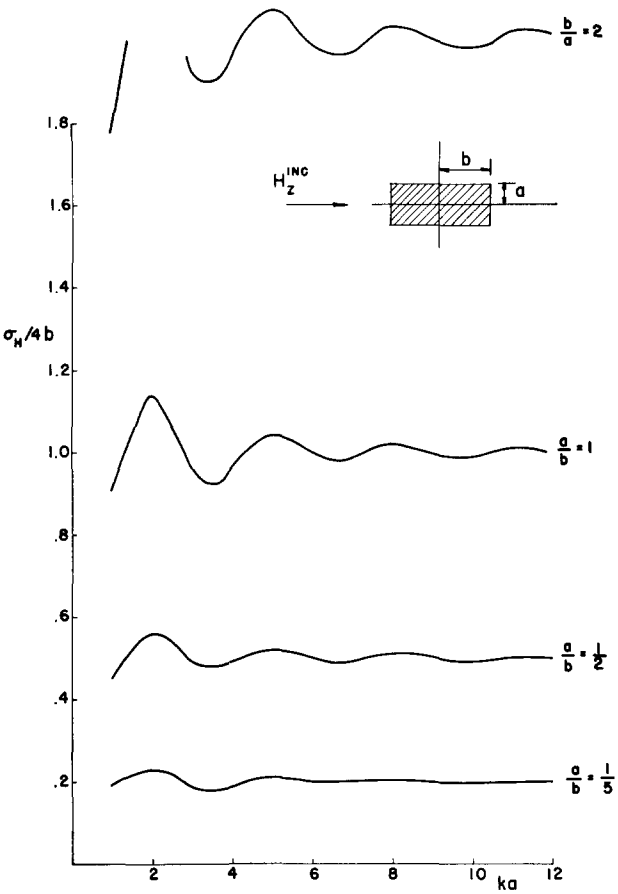


FIG. 5. The ratio of total to optical scattering cross section $\sigma_H/4a$ at grazing incidence, in the case of the hard cylinder.

scattered in direction φ to the average power scattered in all directions. If the scattered far field u_s be written in the form

$$u_s \sim [e^{ikr}/(kr)^{\frac{1}{2}}]f(\varphi) + \cdots,$$

then it may be shown that $G(\varphi)$ for the soft cylinder becomes

$$G(\varphi) = (2\pi/k\sigma_s)\,|f(\varphi)|^2. \tag{2.5}$$

The same formula also holds for the hard cylinder with σ_H in place of σ_s.

If J_z and $J_{\tan}$, respectively, denote the axial and tangential current densities on the surface of the cylinder, then from Maxwell's equations we have in the soft case

$$\partial u/\partial n = -i\omega\mu J_z, \tag{2.6}$$

and in the hard case

$$u = J_{\tan}.$$

Here u and $\partial u/\partial n$ are, respectively, the total H-polarized field and normal derivative of the E-polarized field, both evaluated along the sides of

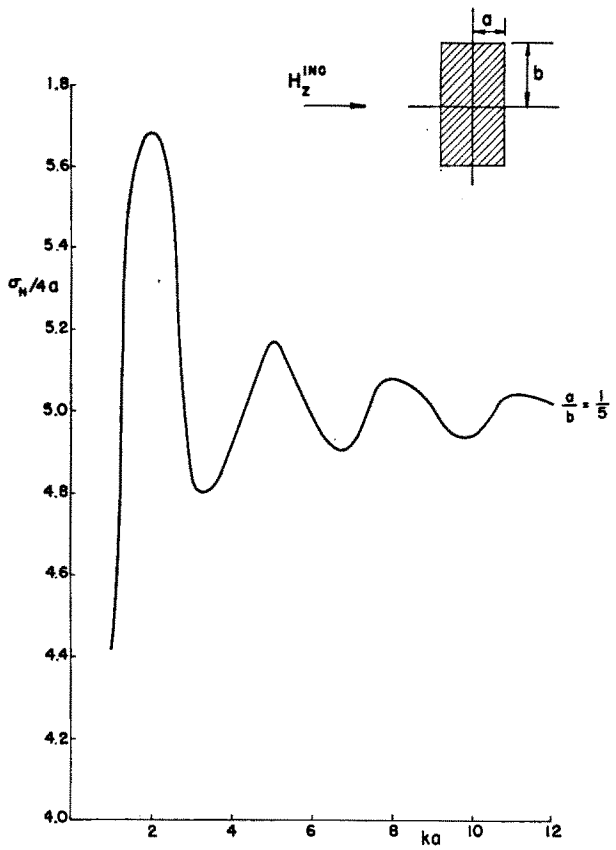


FIG. 6. The same situation as in Fig. 5, but with $a/b = \frac{1}{5}$.

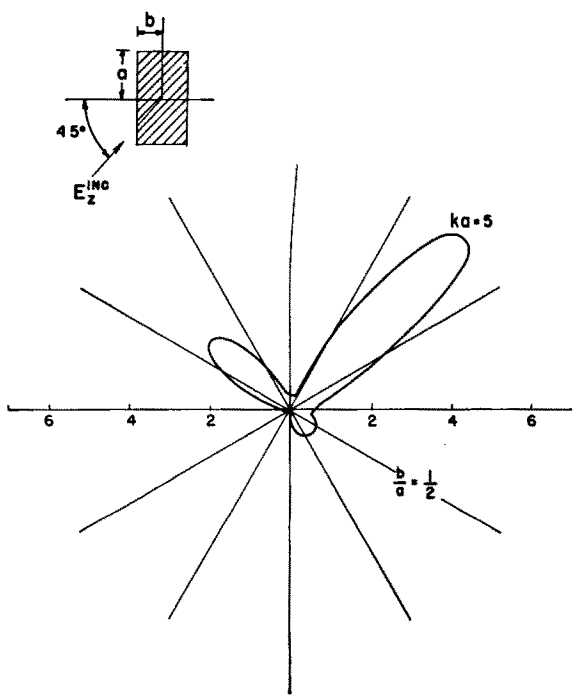


FIG. 7. The gain $G(\varphi)$ for a soft rectangle at oblique incidence.

the polygon. These quantities are readily obtainable by our method. For example, some components of the quantity $\partial u/\partial n$ for grazing incidence will be found in (3.51)–(3.53).

3. ANALYSIS OF THE RECTANGULAR CYLINDER PROBLEM

A. The Field on a Wedge-Diffracted Ray

The exact solution for a plane wave diffracted by a semi-infinite wedge has been examined by Oberhettinger.[6] If the incident wave is $u_i = \exp[-ik\rho\cos(\varphi - \varphi')]$, he finds that the diffracted field can be written as

$$u_d(\rho, \varphi) = I(\pi - \varphi + \varphi', \alpha) + I(\pi + \varphi - \varphi', \alpha) \pm I(\pi - \varphi - \varphi', \alpha) \pm I(\pi + \varphi + \varphi', \alpha). \quad (3.1)$$

Here (ρ, φ) are the polar coordinates of the field point, the wedge faces are at $\varphi = 0$ and $\varphi = \alpha > \pi$ (i.e., α is the exterior angle of the wedge), φ' is the direction of the incident wave; the upper signs apply to a hard wedge, the lower to a soft one. It is easy to show that the function I, which also depends on $k\rho$, is a wavefunction. It has the properties

$$I(\delta + 2n\alpha, \alpha) = I(\delta, \alpha), \quad |n| = 0, 1, 2 \cdots, \qquad I(\delta, \alpha) = -I(-\delta, \alpha), \quad (3.2)$$

and for $|\delta| \leq \alpha$, it is given by

$$I(\delta, \alpha) = -\frac{\operatorname{sgn}\delta}{\pi^{\frac{1}{2}}} \exp[i(k\rho\cos\delta - \tfrac{1}{4}\pi)] \times S\{[k\rho(1 - \cos\delta)]^{\frac{1}{2}}\} - \frac{e^{i(k\rho + \frac{1}{4}\pi)}}{(k\rho)^{\frac{1}{2}}} \sum_{n=0}^{\infty} i^n \Gamma(n + \tfrac{1}{2}) \times \left[A_n(\delta, \alpha) - \frac{(-)^n 2^{-n-\frac{1}{2}}}{2\pi}(\sin\tfrac{1}{2}\delta)^{-2n-1}\right](k\rho)^{-n}. \quad (3.3)$$

Here $S(z)$ is the Fresnel integral

$$S(z) = \int_z^{\infty} e^{it^2}\, dt, \quad (3.4)$$

and the first two of the coefficients $A_n(\delta, \alpha)$ are

$$A_0(\delta, \alpha) = (2^{-\frac{1}{2}}/2\alpha)\cot(\pi\delta/2\alpha)$$
$$A_1(\delta, \alpha) = -\frac{2^{-\frac{1}{2}}}{8\alpha}\cot\left(\frac{\pi\delta}{2\alpha}\right) \times \left[1 + \frac{2\pi^2}{\alpha^2}\sin\left(\frac{\pi\delta}{2a}\right)^{-2}\right]. \quad (3.5)$$

If δ is not near zero and $k\rho \gg 1$, the argument of the Fresnel integral in (3.3) is large and the integral can be expanded asymptotically. Then (3.1)

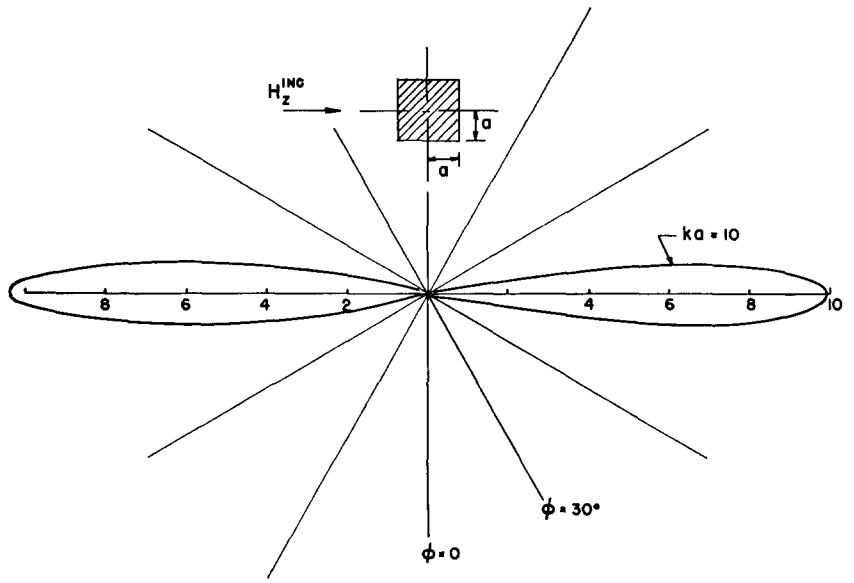


FIG. 8. The gain for a hard square at grazing incidence.

yields

$$u_d(\rho, \varphi) = (e^{ik\rho}/\rho^{\frac{1}{2}})[D(\varphi', \varphi, n) + O(\rho^{-1})], \tag{3.6}$$

where

$$D(\varphi', \varphi, n) = \frac{e^{i\frac{1}{4}\pi} \sin(\pi/n)}{n(2\pi k)^{\frac{1}{2}}} \times \left[\left(\cos\frac{\pi}{n} - \cos\frac{\varphi - \varphi'}{n}\right)^{-1} \pm \left(\cos\frac{\pi}{n} - \cos\frac{\varphi + \varphi'}{n}\right)^{-1}\right], \tag{3.7}$$

with $n = \alpha/\pi$; the upper sign applies to the hard wedge, the lower to the soft.

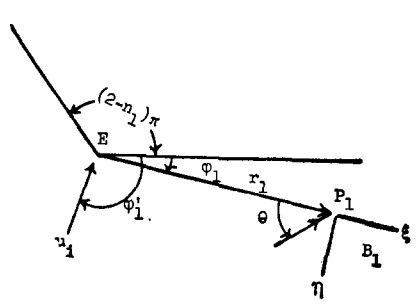


FIG. 9. Illustration of the terminology of formula (3.8) which gives the far field near a ray diffracted by a semi-infinite wedge.

From (3.6), applying a formula due to Zitron and Karp,[5] we obtain the field on and near the ray determined by the point $P_1 = P_1(r_1, \varphi_1)$. More precisely, we obtain the asymptotic expansion of the diffracted field, $u_d(B_1; E_1)$, at any point $B_1 = B_1(\xi, \eta)$ in the neighborhood of the point P_1 (see Fig. 9),

$$u_d(B_1; E_1) = (e^{ikr_1}/r_1^{\frac{1}{2}})\{D(\varphi_1', \varphi_1, n_1)p(0) + [(2ik)^{-1}p''(0)D(\varphi_1', \varphi_1, n_1) - (ik)^{-1}p'(0)D_\varphi(\varphi_1', \varphi_1, n_1) + p(0)D^{(1)}(\varphi_1', \varphi_1, n_1)]r_1^{-1} + O[(kr_1)^{-2}]\}. \tag{3.8}$$

Here

$$D^{(1)}(\varphi', \varphi, n) = (2ik)^{-1}[\tfrac{1}{4}D(\varphi', \varphi, n) + D_{\varphi\varphi}(\varphi', \varphi, n)],$$

and

$$p(\theta) = \exp[ik(\xi \cos\theta - \eta \sin\theta)]. \tag{3.9}$$

Since (3.8), which may be continued to any desired order, is linear in $p(\theta)$ and its derivatives, the fields resulting from successive interactions are readily derived. Indeed, if the field given by (3.8) strikes a second wedge, with edge at E_2, the resulting diffracted field $u_d(B_2; E_1, E_2)$ is given by

$$u_d(B_2; E_1, E_2) = \frac{e^{ikr_{12}}}{r_{12}^{\frac{1}{2}}}\Big\{D(\varphi_1', \varphi_{12}, n_1)u_d(B_2; E_2) + \frac{(2ik)^{-1}u_d''(B_2; E_2)D(\varphi_1', \varphi_{12}, n_1) - (ik)^{-1}u_d'(B_2; E_2)D_\varphi(\varphi_1', \varphi_{12}, n_1)}{r_{12}} + \frac{u_d(B_2; E_2)D^{(1)}(\varphi_1', \varphi_{12}, n_1)}{r_{12}} + O[(kr_{12})^{-2}]\Big\}. \tag{3.10}$$

In this equation, (r_{12}, φ_{12}) are the polar coordinates of E_2 with respect to E_1, $u_d(B_2; E_2)$ is given by (3.8) with the subscripts 2 in place of 1, and the primes on this function indicate derivatives with respect to plus or minus φ_2' depending, respectively, on whether or not θ and φ_2' have the same orientation.

Equation (3.8) is premised on oblique incidence. Whenever the incident field is along a side (i.e., when $\varphi_i' = 0$), the value of the field given by (3.8) must be divided by two. This normalization is due to the coalescing of the reflected and incident fields at this angle of incidence.

In deriving (3.8) we assumed that the angle δ in (3.3) was not small. As a result, (3.8) becomes infinite along shadow or specular boundaries. When δ is small, the argument of the Fresnel integral in (3.3) is also small, and the integral can be expanded in a Taylor series:

$$S(x) = \tfrac{1}{2}\pi^{\frac{1}{2}}e^{i\frac{1}{4}\pi} - x - \tfrac{1}{3}(ix^3) + \cdots . \tag{3.11}$$

Then expanding (3.3) in terms of the variables (ξ, η), where $\rho = [(r+\xi)^2 + \eta^2]^{\frac{1}{2}}$ and $\delta = \tan^{-1}[\eta/(r+\xi)]$, yields

$$I(\delta, \alpha) = \operatorname{sgn}\delta e^{ik(r+\xi)}\left\{-\frac{1}{2} + \frac{e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}}\left[\frac{k\,|\eta|}{(2kr)^{\frac{1}{2}}} + \frac{-3\sqrt{2}\,k^2\xi\,|\eta| + i\sqrt{2}\,k^3\,|\eta|^3}{12(kr)^{\frac{3}{2}}}\right]\right\}$$
$$+ e^{ik(r+\xi)}\frac{e^{i\frac{1}{4}\pi}}{2^{\frac{1}{2}}}\left(\frac{\pi^{\frac{1}{2}}}{6\alpha^2} + \frac{1}{12\pi}\right)$$
$$\times \frac{k\eta}{(kr)^{\frac{1}{2}}} + O[(kr)^{-5/2}]. \tag{3.12}$$

Hence if $\delta > 0$, we obtain, using (3.9),

$$I(\delta, \alpha) = -\tfrac{1}{2}e^{ikr}p(0) + \frac{e^{-i\frac{1}{4}\pi}e^{ikr}}{\pi^{\frac{1}{2}}}\left\{\frac{ip'(0)}{(2kr)^{\frac{1}{2}}} + \frac{\sqrt{2}\,[p'(0) + p'''(0)]}{12(kr)^{\frac{3}{2}}}\right\}$$
$$+ e^{ikr}\frac{e^{i\frac{1}{4}\pi}}{2^{\frac{1}{2}}}\left(\frac{\pi^{\frac{1}{2}}}{6\alpha^2} + \frac{1}{12\pi^{\frac{1}{2}}}\right)\frac{ip'(0)}{(kr)^{\frac{1}{2}}} + O[(kr)^{-5/2}]. \tag{3.13}$$

In addition to (3.12) and (3.13), we shall need the following expansion in order to treat shadow boundary fields:

$$I(\theta \pm \delta) = -\frac{e^{i\frac{1}{4}\pi}e^{ikr}}{3(2\pi)^{\frac{1}{2}}(kr)^{\frac{1}{2}}}\left\{p(0)\cot\tfrac{1}{3}\theta + \frac{\mp 8ip'(0)\csc^2(\frac{1}{3}\theta) - 3ip(0)\cot(\frac{1}{3}\theta)[1 + \frac{8}{9}\csc^2(\frac{1}{3}\theta)] - 12ip''(0)\cot(\frac{1}{3}\theta)}{24(kr)} + O[(kr)^{-2}]\right\}. \tag{3.14}$$

Here θ is an arbitrary but fixed large angle, and δ a small angle defined as before by the relation $\delta = \tan^{-1}[\eta/(r+\xi)]$. Equation (3.14) follows from (3.3) after setting α equal to $\frac{1}{2}(3\pi)$ and proceeding as in the derivation of (3.12) and (3.13). In (3.14) as well as the sequel we use the notation $I(\omega) \equiv I(\omega, \frac{3}{2}\pi)$.

B. Scattering Cross Sections at Oblique Incidence

We consider a plane wave

$$u_i(P) = \exp[-ikr\sin(\varphi + \varphi')], \tag{3.15}$$

incident on a rectangular cylinder, the vertices of whose cross sections are E_1, E_2, E_3, and E_4. The polar coordinates (r, φ) of the field point $P = P(r, \varphi)$ have their origin at the center of the rectangle (see Fig. 1). The length of side $\overline{E_1E_2}$ is $2a$ and that of side $\overline{E_2E_3}$, $2b$. Equation (3.15) implies that the direction of propagation of the incident wave makes an angle φ' with side $\overline{E_2E_3}$. By oblique incidence, we mean that $\frac{1}{2}\pi < \varphi' < \pi$.

We begin by applying (3.8) and the method illustrated by (3.10) to calculate the multiply-diffracted fields at the vertices. $u_n(E)$, $n = 1, 2, \cdots$, indicates the n-tuply diffracted field at E. The notations $h(0)$ and $v(0)$ will be used, in place of $p(0)$, given in (3.9), to distinguish between plane waves striking the same vertex on the "horizontal" or "vertical" side respectively. Thus $u_1(E_1) = u_d(E_1; E_2)$. $u_2(E_1) = u_d(E_1; E_1, E_4) + u_d(E_1; E_1, E_2) + u_d(E_1; E_3, E_4) + u_d(E_1; E_3, E_2)$ and so on.

The functions $u_n(E_i)$ with $n \geq 4$ consist entirely of terms of order higher than $(ka)^{-\frac{1}{2}}$ and hence will be disregarded. The incident vertex fields $u_i(E_j)$ $j = 1, 2, 3$ are found from (3.15).

The resulting vertex fields apply to both hard and soft rectangular cylinders. However they simplify when applied to each of these cases separately. From (3.7) it may be seen that the function $D(\varphi', \varphi, n)$ is analytic and odd in φ for the soft case, but even in φ for the hard. Thus $D(\varphi', 0, n)$, $D^{(1)}(\varphi', 0, n)$ and all even-order derivatives $D^{(2j)}_{\varphi\cdots\varphi}(\varphi', 0, n)$ vanish in the soft case, whereas all odd-order derivatives $D^{(2j+1)}_{\varphi\cdots\varphi}(\varphi', 0, n)$ vanish in the hard case.

The Soft Cylinder

For the soft rectangular cylinder, the vertex fields reduce to:

$$u_1(E_1) = -u_i(E_2)L(a, a) \times D_\varphi(\tfrac{3}{2}\pi - \varphi', 0, \tfrac{3}{2})h'(0) + O(ka)^{-5/2}, \quad (3.16)$$

$$u_1(E_2) = -u_i(E_1)L(a, a) \times D_\varphi(\varphi' - \tfrac{1}{2}\pi, 0, \tfrac{3}{2})h'(0) - u_i(E_3)L(b, b) \times D_\varphi(\pi - \varphi', 0, \tfrac{3}{2})v'(0) + O(ka)^{-5/2}, \quad (3.17)$$

$$u_1(E_3) = -u_i(E_2)L(a, b) \times D_\varphi(\varphi', 0, \tfrac{3}{2})v'(0) + O(ka)^{-5/2}, \quad (3.18)$$

$$u_1(E_4) = -u_i(E_3)L(a, a) \times D_\varphi(\varphi' + \tfrac{1}{2}\pi, 0, \tfrac{3}{2})h'(0) - u_i(E_1)L(b, b)D_\varphi(2\pi - \varphi', 0, \tfrac{3}{2})v'(0) + O(ka)^{-5/2}, \quad (3.19)$$

where

$$L(u, v) = e^{ik2u}/ik(2v)^{\frac{1}{2}},$$

and up to the order considered,

$$u_n(E_i) = 0 \qquad (i = 1, \cdots, 4), \qquad n > 1.$$

In order to find the total scattering cross section we compute the far field in the forward direction i.e., referring to Fig. 1, in the direction $\varphi = \frac{3}{2}\pi - \varphi'$. Rays from vertex E_2, of course, do not go in this direction but those from the other three do. We consider the contributions from E_1.

The total field, $u(E_1)$, at vertex E_1 is

$$u(E_1) = u_i(E_1) + u_1(E_1) + O[(ka)^{-5/2}].$$

In order to find the forward response of E_1 to the excitation $u_i(E_1)$ we construct the configuration shown in Fig. 10. Here the line $\overline{E_1 0_1}$ is in the forward direction and of length r_1, where r_1 is assumed large with respect to a or b. $P_1 = P_1(\xi_1, \eta_1)$ is any point whose distance from the point 0_1 is small compared to r_1. The angle $\varphi_1' = 2\pi - \varphi'$ is the angle between the direction of propagation of the incident field, $u_i(E_1)$, and the side $\overline{E_1E_4}$. $\varphi_1 = \pi - \varphi'$ is the angle between $\overline{E_1 0_1}$ and this same side. δ_1, the angle between $\overline{E_1P_1}$ and the shadow line $\overline{E_1 0_1}$ is, of course, small.

Since $u_d(P_1; E_1)$ is the diffracted field at P_1 caused by the field $u_i(E_1)$, we have from (3.1) and (3.2)

$$u_d(P_1; E_1) = u_i(E_1)[I(\delta_1) - I(\pi + \delta_1) + I(2\pi - 2\varphi' + \delta_1) - I(\pi - 2\varphi' + \delta_1)].$$

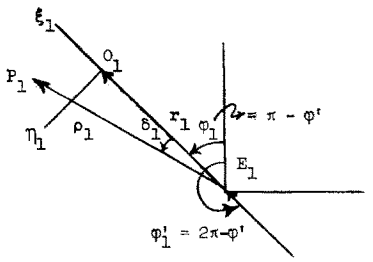


FIG. 10. Configuration for finding the far field in the neighborhood of a shadow line at oblique incidence.

From (3.12) we obtain

$$I(\delta_1) = \operatorname{sgn} \delta_1 e^{ik(r_1+\xi_1)} \times \left[-\frac{1}{2} + \frac{e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}} \frac{k\,|\eta_1|}{(2kr_1)^{\frac{1}{2}}} + O(kr_1)^{-\frac{3}{2}}\right], \quad (3.20)$$

and from (3.14) and (3.9),

$$I(\theta + \delta_1) = -[e^{i\frac{1}{4}\pi}e^{ik(r_1+\xi_1)}/3(2\pi)^{\frac{1}{2}}(kr_1)^{\frac{1}{2}}] \times \cot \tfrac{1}{3}\theta + O(kr_1)^{-\frac{3}{2}}. \quad (3.21)$$

Hence, using these expressions in the above,

$$u_d(P_1; E_1) = u_i(E_1)e^{ik(r_1+\xi_1)} \times \left\{\operatorname{sgn} \delta_1\left[-\frac{1}{2} + \frac{e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}} \frac{k\,|\eta_1|}{(2kr_1)^{\frac{1}{2}}}\right] + \frac{e^{i\frac{1}{4}\pi}}{3\pi^{\frac{1}{2}}(2kr_1)^{\frac{1}{2}}} \times \left[\frac{\sqrt{3}}{3} + \frac{2\sqrt{3}}{2\cos\frac{4}{3}\varphi' + 1}\right] + O(kr_1)^{-\frac{3}{2}}\right\}. \quad (3.22)$$

It remains to determine the field $u_d(P_1; E_2, E_1)$ using (3.16). Recalling that $h'(0)$ in (3.16) represents the derivative of a horizontal plane wave at vertex E_1, we obtain, after applying (3.10) and dividing by two [since $h'(0)$ is along a side],

$$u_d(P_1; E_2, E_1) = -u_i(E_2) \times [e^{ik(r_1+\xi_1)}/2r_1^{\frac{1}{2}}]D_\varphi(\tfrac{3}{2}\pi - \varphi', 0, \tfrac{3}{2}) \times D_{\varphi'}(0, \varphi' + \tfrac{1}{2}\pi, \tfrac{3}{2})[L(a, a) + O(2ka)^{-5/2}] + O(kr_1)^{-\frac{3}{2}}. \quad (3.23)$$

Similarly, the forward responses from E_3 and E_4 are also calculated.

The total diffracted field within the shadow boundary $u_d(P)$ is now found by adding the contributions from E_1, E_3, E_4. Here P has the coordinates $(r, \frac{3}{2}\pi - \varphi')$ relative to the system shown in Fig. 1. The geometry of the relevant coordinate transformations is shown in Fig. 11. Hence, defining the scattered field at P, $u_s(P)$, by

$$u_s(P) = u_d(P) - u_i(P) \quad \text{where} \quad u_i(P) = e^{ikr}$$

and

$$u_d(P) = u_d(P; E_1) + u_d(P; E_2, E_1) + u_d(P; E_3) + u_d(P; E_2, E_3) + u_d(P_4; E_1, E_4) + u_d(P_4; E_3, E_4) + O(ka)^{-5/2},$$

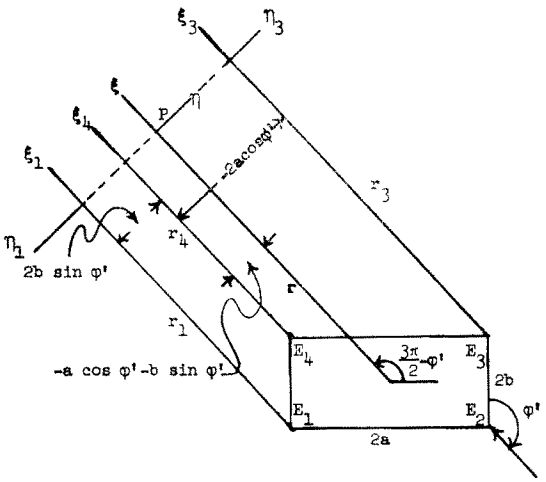


FIG. 11. Coordinate relationships used in calculating the total diffracted field in the shadow of a rectangular cylinder.

we obtain

$$u_s(P) = \frac{e^{ikr}}{r^{\frac{1}{2}}}\left\{-\frac{e^{-i\frac{1}{4}\pi}}{(2k\pi)^{\frac{1}{2}}}k(2b\sin\varphi' - 2a\cos\varphi')\right.$$
$$+\frac{2\sqrt{3}\,e^{i\frac{1}{4}\pi}}{3(2k\pi)^{\frac{1}{2}}}\left(\frac{1}{3}+\frac{2}{2\cos\frac{4}{3}\varphi'+1}\right)$$
$$-\frac{e^{ikA-}}{2ik(2a)^{\frac{1}{2}}}[D_\varphi(\tfrac{3}{2}\pi-\varphi',0,\tfrac{3}{2})D_{\varphi'}(0,\varphi'+\tfrac{1}{2}\pi,\tfrac{3}{2})$$
$$+D_\varphi(\varphi'+\tfrac{1}{2}\pi,0,\tfrac{3}{2})D_{\varphi'}(0,\tfrac{3}{2}\pi-\varphi',\tfrac{3}{2})[-\frac{e^{ikB+}}{2ik(2b)^{\frac{1}{2}}}$$
$$\times[D_\varphi(2\pi-\varphi',0,\tfrac{3}{2})D_{\varphi'}(0,\varphi',\tfrac{3}{2})$$
$$+D_\varphi(\varphi',0,\tfrac{3}{2})D_{\varphi'}(0,2\pi-\varphi',\tfrac{3}{2})]$$
$$\left.+O[k^{-\frac{1}{2}}(ka)^{-5/2}]\right\}+O(kr)^{-\frac{1}{2}}. \qquad (3.24)$$

Here, applying (3.7),

$$D_\varphi(\theta,0,\tfrac{3}{2}) = D_{\varphi'}(0,\theta,\tfrac{3}{2})$$
$$= \frac{4e^{i\frac{1}{4}\pi}\sqrt{3}}{9(2\pi k)^{\frac{1}{2}}}(\tfrac{1}{2}+\cos\tfrac{2}{3}\theta)^{-2}\sin\tfrac{2}{3}\theta. \qquad (3.25)$$

The total scattering cross section σ_s of the soft cylinder can now be found from (3.24) by using the cross-section theorem. Hence defining $E(\theta)$ by the equation

$$E(\theta) = (\tfrac{1}{2}+\cos\tfrac{2}{3}\theta)^{-2}\sin\tfrac{2}{3}\theta,$$

we have from (3.24)

$$\sigma_s = 4(b\sin\varphi' - a\cos\varphi')$$
$$+\frac{8}{27k\pi^{\frac{1}{2}}}\left[\frac{\cos(kA_- + \frac{1}{4}\pi)}{(ka)^{\frac{1}{2}}}\right.$$
$$\times E(\tfrac{3}{2}\pi-\varphi')E(\varphi'+\tfrac{1}{2}\pi)$$
$$\left.+\frac{\cos(kB_+ + \frac{1}{4}\pi)}{(kb)^{\frac{1}{2}}}E(2\pi-\varphi')E(\varphi')\right]$$
$$+O[k^{-1}(ka)^{-5/2}]. \qquad (3.26)$$

It is to be remembered that, by our premise of oblique incidence, (3.24) and (3.26) are only valid in the range $\frac{1}{2}\pi < \varphi' < \pi$.

The Hard Cylinder

In this case, as we have seen above, $D_\varphi(\varphi',0,n) = D_{\varphi'}(0,\varphi,n) = 0$. Since the procedure is exactly the same as in the soft case, we shall not repeat the details but only give the results. Thus, corresponding to (3.24),

$$u_s(P) = \frac{e^{ikr}}{r^{\frac{1}{2}}}\left\{-\frac{e^{-i\frac{1}{4}\pi}}{(2k\pi)^{\frac{1}{2}}}k(2b\sin\varphi' - 2a\cos\varphi')\right.$$
$$+\frac{2\sqrt{3}\,e^{i\frac{1}{4}\pi}}{3(2k\pi)^{\frac{1}{2}}}\left(\frac{1}{3}-\frac{2}{2\cos\frac{4}{3}\varphi'+1}\right)+\frac{e^{ikA-}}{(2a)^{\frac{1}{2}}}$$
$$\times D(\tfrac{3}{2}\pi-\varphi',0,\tfrac{3}{2})D(0,\varphi'+\tfrac{1}{2}\pi,\tfrac{3}{2})$$
$$+\frac{e^{ikB+}}{(2b)^{\frac{1}{2}}}D(\varphi',0,\tfrac{3}{2})D(0,2\pi-\varphi',\tfrac{3}{2})$$
$$+\frac{e^{4aik}}{4a}D(\varphi'-\tfrac{1}{2}\pi,0,\tfrac{3}{2})D(0,0,\tfrac{3}{2})D(0,\varphi'+\tfrac{1}{2}\pi,\tfrac{3}{2})$$
$$+\frac{e^{ik(A-+B-)}}{4(ab)^{\frac{1}{2}}}D(\pi-\varphi',0,\tfrac{3}{2})$$
$$\times D(\tfrac{3}{2}\pi,0,\tfrac{3}{2})D(0,\varphi'+\tfrac{1}{2}\pi,\tfrac{3}{2})$$
$$+\frac{e^{ik(A-+B+)}}{4(ab)^{\frac{1}{2}}}D(\tfrac{3}{2}\pi-\varphi',0,\tfrac{3}{2})$$
$$\times D(\tfrac{3}{2}\pi,0,\tfrac{3}{2})D(0,\varphi',\tfrac{3}{2})$$
$$+\frac{e^{ik(A++B+)}}{4(ab)^{\frac{1}{2}}}D(\varphi'-\tfrac{1}{2}\pi,0,\tfrac{3}{2})$$
$$\times D(\tfrac{3}{2}\pi,0,\tfrac{3}{2})D(0,2\pi-\varphi',\tfrac{3}{2})$$
$$+\frac{e^{4bik}}{4b}D(2\pi-\varphi',0,\tfrac{3}{2})D(0,0,\tfrac{3}{2})D(0,\pi-\varphi',\tfrac{3}{2})$$
$$\left.+O[k^{-\frac{1}{2}}(ka)^{-\frac{1}{2}}]\right\}+O(kr)^{-\frac{1}{2}}. \qquad (3.27)$$

The function D and its derivatives are determined from (3.7). Hence, defining $F(\theta)$ as

$$F(\theta) = (\tfrac{1}{2}+\cos\tfrac{2}{3}\theta)^{-1}, \qquad (3.28)$$

we find from (3.27) and the cross-section theorem,

$$\sigma_H = 4(b\sin\varphi' - a\cos\varphi')$$
$$+\frac{4}{3k\pi^{\frac{1}{2}}}\left[\frac{\cos(kA_- - \frac{1}{4}\pi)}{(ka)^{\frac{1}{2}}}F(\tfrac{3}{2}\pi-\varphi')F(\varphi'+\tfrac{1}{2}\pi)\right.$$
$$\left.+\frac{\cos(kB_+ - \frac{1}{4}\pi)}{(kb)^{\frac{1}{2}}}F(\varphi')F(2\pi-\varphi')\right]$$
$$-\frac{4\sqrt{3}}{27k\pi}\frac{\cos 4ka}{ka}F(\varphi'-\tfrac{1}{2}\pi)F(\varphi'+\tfrac{1}{2}\pi)$$

$$+\frac{4\sqrt{3}}{9k\pi}\frac{\cos(kA_- + kB_-)}{k(ab)^{\frac{1}{2}}}F(\pi-\varphi')F(\varphi'+\tfrac{1}{2}\pi)$$

$$+\frac{4\sqrt{3}}{9k\pi}\frac{\cos(kA_- + kB_+)}{k(ab)^{\frac{1}{2}}}F(\tfrac{3}{2}\pi-\varphi')F(\varphi')$$

$$+\frac{4\sqrt{3}}{9k\pi}\frac{\cos(kA_+ + kB_+)}{k(ab)^{\frac{1}{2}}}F(\varphi'-\tfrac{1}{2}\pi)F(2\pi-\varphi')$$

$$-\frac{4\sqrt{3}}{27k\pi}\frac{\cos 4kb}{kb}F(2\pi-\varphi')F(\pi-\varphi')$$

$$+O[k^{-1}(ka)^{-\frac{1}{2}}]. \tag{3.29}$$

Since they apply to the cases of oblique incidence, (3.27) and (3.29) are only valid for $\frac{1}{2}\pi < \varphi' < \pi$.

C. Asymptotic Expansion of Green's Theorem Integrals

As we have seen above, the interacting vertex fields are linear expressions which involve plane waves and their derivatives traveling along the sides of the rectangle. Hence directions parallel to a side are along shadow lines with respect to these vertex fields. The determination of the far fields along these shadow lines by the methods used heretofore consists of forming appropriate linear combinations of angular derivatives of (3.1). When this process is carried out, it is found that the resulting far-field formulas contain positive powers of kr and thus are not in the desired form.

In order to see the cause of this heuristically, we observe that the field changes rapidly across a shadow line. Hence the angular derivative of the field evaluated at a shadow line is a large quantity rather than one which becomes asymptotically small with increasing distance. Analytically this is demonstrated by differentiating (3.3) with respect to δ and then letting δ approach zero.

Thus to generate the far fields in these cases, a different procedure is required and we shall use Green's theorem,

$$u_s(r,\varphi) = u(r,\varphi) - u_i(r,\varphi)$$

$$= \frac{i}{4}\oint\left[u\frac{\partial}{\partial n}H_0^1(kR) - H_0^1(kR)\frac{\partial u}{\partial n}\right]ds. \tag{3.30}$$

Here $u_s(r,\varphi)$ is the scattered field at the point (r,φ), and hence is the total field $u(r,\varphi)$ minus the incident field $u_i(r,\varphi)$. In the integral, which is taken around the sides of the rectangle, $u = u(s)$ is the total field at the surface of the rectangle, n the outer normal, and $R = R(s, P)$ the distance from a point s on the side to the field point, P. We observe that, depending on whether the soft or hard case is being considered, u or $\partial u/\partial n$ will vanish.

To obtain the asymptotic representation of Green's theorem for r large with respect to the lengths $2a$ and $2b$, we use the following expression for the Hankel function:

$$H_0^1(kR) = (2/\pi kR)^{\frac{1}{2}}e^{i(kR-\frac{1}{4}\pi)} + O(kR)^{-\frac{3}{2}}. \tag{3.31}$$

Applying (3.31) after substituting appropriate formulas for R in terms of r we obtain

$$u_s(r,\varphi) = \frac{e^{i(kr-\frac{1}{4}\pi)}}{2(2\pi kr)^{\frac{1}{2}}}\sum_{j=1}^{4}\exp[-ik(x_j\cos\varphi + y_j\sin\varphi)]$$

$$\times\int_0^{2a_j}\left\{-ku\sin[\varphi+(1-j)\tfrac{1}{2}\pi] - i\frac{\partial u}{\partial n}\right\}$$

$$\times\exp\{-ik\rho_j\cos[\varphi+(1-j)\tfrac{1}{2}\pi]\}\,d\rho_j$$

$$+O(kr)^{-\frac{3}{2}}, \tag{3.32}$$

where ρ_j, $j = 1, \cdots, 4$ is the distance measured from vertex E_j toward E_{j+1}, a_j is the length of this side, and (x_j, y_j) are the coordinates of E_j (see Fig. 1).

Examining (3.32) it is observed that the integrals to be evaluated are of the form

$$\int_0^{2d}\left(u \text{ or } \frac{\partial u}{\partial n}\right)\exp(ik\rho\sin\tau)\,d\rho,$$

with suitable values for τ and d. As will be seen below, u and $\partial u/\partial n$ are expressible as linear combinations of the functions

$$I(\delta, \tfrac{3}{2}\pi) \quad \text{and} \quad \partial^j I(\delta, \tfrac{3}{2}\pi)/\rho\,\partial\delta^j, \qquad j = 1, 2\cdots.$$

Hence the integrals arising in the application of Green's theorem reduce to the two types

$$\int_0^{2d} I(\delta,\alpha)\exp(ik\rho\sin\tau)\,d\rho, \tag{3.33a}$$

and

$$\int_0^{2d}\frac{\partial^j I(\delta,\alpha)}{\partial\delta^j}\exp(ik\rho\sin\tau)\frac{d\rho}{\rho}. \tag{3.33b}$$

We require asymptotic representations of these integrals for $\alpha = \frac{3}{2}\pi$ with respect to the large quantity kd. To obtain these representations it is not sufficient to use the series given in Eq. (3.3) since the resulting expressions would be divergent. Therefore we shall employ the integral form of the function $I(\delta,\alpha)$ which, according to Oberhettinger,[6] is

$$I(\delta,\alpha) = -\frac{1}{2\alpha}\sin\left(\frac{\pi\delta}{\alpha}\right)$$

$$\times\int_0^\infty\frac{\exp(ik\rho\cosh x)\,dx}{\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)}. \tag{3.34}$$

The form of this function forces us to consider several cases. We consider type (3.33a) first and return to (3.33b) later.

Case a1: $\delta \neq 0$, $\tau \neq -\frac{1}{2}\pi$.

Substituting (3.34) into (3.33a), we may interchange order of integration by Fubini's theorem. Thus letting $\beta = -ik$,

$$\int_0^{2d} I(\delta, \alpha) \exp(-\beta\rho \sin\tau)\, d\rho = \frac{1}{2\alpha\beta} \sin\frac{\pi\delta}{\alpha} \left\{ \exp(-2d\beta \sin\tau) \times \int_0^\infty \frac{\exp(-2d\beta \cosh x)\, dx}{(\cosh x + \sin\tau)[\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)]} - \int_0^\infty \frac{dx}{(\cosh x + \sin\tau)[\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)]} \right\}. \tag{3.35}$$

The second integral on the right side of this equation has been evaluated in the appendix for α, a rational mutiple of π. The remaining integral on the right can be expanded by the use of Watson's lemma. Since similar integrals have been so expanded by Oberhettinger,[6] we omit the details. For $\alpha = \frac{3}{2}\pi$, our result is

$$\int_0^{2d} I(\delta, \tfrac{3}{2}\pi) \exp(ik\rho \sin\tau)\, d\rho \sim \frac{i}{3\pi k} \sin\tfrac{2}{3}\delta \times \left\{ -\sum_{j=0}^{2} [2\tau + \pi(4j-5)]R_j - \sum_{h=0}^{1} [\delta + 3\pi \operatorname{sgn}\delta(h-1)]T_h \right\} + \exp\{i[2kd(1+\sin\tau) + \tfrac{1}{4}\pi]\} \times \sum_{n=0}^{\infty} i^n B_n(\delta, \tfrac{3}{2}\pi, \tau)\Gamma(n+\tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}, \tag{3.36}$$

where R_j and T_h are given in the appendix ($m = 3$, $n = 2$), and the first two coefficients are

$$B_0(\delta, \alpha, \tau) = [2^{-\frac{1}{2}}/2\alpha\beta(1+\sin\tau)] \cot(\pi\delta/2\alpha)$$

$$B_1(\delta, \alpha, \tau) = -\frac{2^{-\frac{1}{2}} \cot(\pi\delta/2\alpha)}{2\alpha\beta(1+\sin\tau)} \times \left\{ \frac{1}{4} + \frac{1}{1+\sin\tau} + \frac{\pi^2}{2\alpha^2} \csc^2\left(\frac{\pi\delta}{2\alpha}\right) \right\}. \tag{3.37}$$

Case a2: $\delta \neq 0$, $\tau = -\frac{1}{2}\pi$.

When $\tau = -\frac{1}{2}\pi$, poles occur in the integrands on the right side of (3.35) at the lower limit of integration. To circumvent this we add and subtract a known integral with the same singularity. Thus from (3.35) with $\tau = -\frac{1}{2}\pi$,

$$\int_0^{2d} I(\delta, \alpha) e^{\beta\rho}\, d\rho = \frac{1}{2\alpha\beta} \sin\frac{\pi\delta}{\alpha} \left(e^{2\beta d} \int_0^\infty e^{-2\beta d \cosh x} \times \left\{ \frac{1}{(\cosh x - 1)[\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)]} - \frac{[1-\cos(\pi\delta/\alpha)]^{-1} \cosh\frac{1}{2}x}{\cosh x - 1} \right\} dx + \lim_{\tau \to -\frac{1}{2}\pi^+} \left\{ e^{-2\beta d \sin\tau} \left(1 - \cos\frac{\pi\delta}{\alpha}\right)^{-1} \times \int_0^\infty \frac{\cosh\frac{1}{2}x e^{-2\beta d \cosh x}\, dx}{\cosh x + \sin\tau} - \int_0^\infty \frac{dx}{(\cosh x + \sin\tau)[\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)]} \right\} \right). \tag{3.38}$$

The limit in (3.38) exists and may be evaluated. Furthermore, the first integrand on the right-hand side of this equation is regular at $x = 0$. Thus the asymptotic expansion of this integral can be found by Watson's lemma. Again, since similar integrals have been so treated by Oberhettinger,[6] we only state the result:

$$\int_0^{2d} I(\delta, \tfrac{3}{2}\pi) e^{-ik\rho}\, d\rho \sim T(\delta) + e^{i\frac{1}{4}\pi} \sum_{n=0}^{\infty} i^n C_n(\delta, \tfrac{3}{2}\pi)\Gamma(n+\tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}. \tag{3.39}$$

Here

$$T(\delta) = \frac{i}{3\pi k} \sin\tfrac{2}{3}\delta \left[\frac{1}{3(1-\cos\frac{2}{3}\delta)} + \frac{2\pi\sqrt{3}}{9(\frac{1}{2}+\cos\frac{2}{3}\delta)^2} - \frac{2}{3(\frac{1}{2}+\cos\frac{2}{3}\delta)} + \frac{-\delta}{2(\cos\delta+1)\sin\frac{2}{3}\delta} + \frac{\delta - 3\pi\operatorname{sgn}\delta}{2(\cos\delta-1)\sin\frac{2}{3}\delta} - \frac{(2\pi)^{\frac{1}{2}} e^{-i\frac{1}{4}\pi}(2kd)^{\frac{1}{2}}}{1-\cos\frac{2}{3}\delta} \right],$$

and the first two coefficients are

$$C_0(\delta, \alpha) = -\frac{i2^{-\frac{1}{2}}}{8\alpha k} \cot\frac{\pi\delta}{2\alpha} \times \left[1 + \frac{2\pi^2}{\alpha^2} \csc^2\left(\frac{\pi\delta}{2\alpha}\right) \right],$$

$$C_1(\delta, \alpha) = \frac{i2^{-\frac{1}{2}}}{8\alpha k} \cot\frac{\pi\delta}{2\alpha} \left[\frac{3}{8} + \left(\frac{5\pi^2}{6\alpha^2} - \frac{\pi^4}{3\alpha^4}\right) \times \csc^2\left(\frac{\pi\delta}{2\alpha}\right) + \frac{\pi^4}{\alpha^4} \csc^4\left(\frac{\pi\delta}{2\alpha}\right) \right]; \tag{3.40}$$

in fact,

$$A_{n+1}(\delta, \alpha) = -ikC_n(\delta, \alpha), \tag{3.41}$$

where the coefficients $A_{n+1}(\delta, \alpha)$ have been given in (3.5).

Case a3: $\tau = -\frac{1}{2}\pi$, δ in neighborhood of zero.

As δ approaches zero, (3.36) and (3.39) become infinite. This is due to the occurrence of a pole in (3.34) at $x = 0$. The limit still exists however, since the effect of the pole is balanced by the vanishing of the factor, $\sin \pi\delta/\alpha$, multiplying the integral.

In this case, applying a formula of Oberhettinger[6] yields

$$\begin{aligned}\int_0^{2d} I(\delta, \tfrac{3}{2}\pi)e^{-ik\rho}\, d\rho &= -\frac{\operatorname{sgn}\delta}{\pi^{\frac{1}{2}}} e^{-i\frac{1}{4}\pi} \\ &\times \int_0^{2d} e^{ik\rho(\cos\delta - 1)} S[\sqrt{k\rho(1 - \cos\delta)}]\, d\rho \\ &- \int_0^{2d} I(2\pi - \delta, 2\pi)e^{-ik\rho}\, d\rho \\ &+ \int_0^{2d} [I(\delta, \tfrac{3}{2}\pi) - I(\delta, 2\pi)]e^{-ik\rho}\, d\rho. \end{aligned} \tag{3.42}$$

As he shows neither of the terms $I(2\pi - \delta, 2\pi)$ or $I(\delta, \frac{3}{2}\pi) - I(\delta, 2\pi)$ are singular at $\delta = 0$, hence the second and third integrals on the right hand side of (3.42) can be expanded by the same method as was employed in Case a2. The remaining integral may be integrated by parts. Collecting terms, our result is

$$\begin{aligned}\int_0^{2d} I(\delta, \tfrac{3}{2}\pi)e^{-ik\rho}\, d\rho &\sim -\frac{\operatorname{sgn}\delta}{k\pi^{\frac{1}{2}}} e^{i\frac{1}{4}\pi} \\ &\times \left(\frac{S\{[2kd(1 - \cos\delta)]^{\frac{1}{2}}\}e^{2ikd(\cos\delta - 1)}}{1 - \cos\delta}\right) \\ &+ T(\delta) + e^{i\frac{1}{4}\pi} \sum_{n=0}^{\infty} i^n[C_n(\delta, \tfrac{3}{2}\pi) - C_n(\delta, 2\pi) \\ &- C_n(2\pi - \delta, 2\pi)]\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}. \end{aligned} \tag{3.43}$$

In the limit as δ goes to zero it may be shown that (3.43) becomes $-d \operatorname{sgn}\delta$. For δ not near zero, (3.43) may be transformed into Eq. (3.39).

This concludes our treatment of integrals of type (3.33a). It remains to consider those of type (3.33b). Of this type we shall only require the cases $j = 1$ and 2. The case $j = 2$ will be taken up first.

Since it may be shown that $I(\delta, \alpha)$ is a wave-function, we have from the reduced wave equation

$$\frac{1}{\rho}\frac{\partial^2 I}{\partial\delta^2} = -\rho k^2 I - \frac{\partial}{\partial\rho}\left(\rho\frac{\partial I}{\partial\rho}\right);$$

therefore,

$$\begin{aligned}\int_0^{2d} \frac{\partial^2 I(\delta, \alpha)}{\rho\, \partial\delta^2} e^{ik\rho\sin\tau}\, d\rho \\ = -\int_0^{2d} \left[\rho k^2 I + \frac{\partial}{\partial\rho}\left(\rho\frac{\partial I}{\partial\rho}\right)\right] e^{ik\rho\sin\tau}\, d\rho.\end{aligned}$$

Integrating by parts twice we therefore obtain

$$\begin{aligned}\int_0^{2d} \frac{\partial^2}{\partial\delta^2} I(\delta, \alpha)e^{ik\rho\sin\tau}\frac{d\rho}{\rho} &= -\rho\frac{\partial I}{\partial\rho} e^{ik\rho\sin\tau}\Big|_{\rho=0}^{2d} \\ &+ ik\rho\sin\tau I e^{ik\rho\sin\tau}\Big|_{\rho=0}^{2d} \\ &+ ik\left[\cos\tau\frac{\partial}{\partial\tau} - \sin\tau\right]\int_0^{2d} I e^{ik\rho\sin\tau}\, d\rho. \end{aligned} \tag{3.44}$$

Hence the asymptotic representations of the integral on the left side of Eq. (3.44) may be obtained from those previously found in Cases a1, a2, and a3 together with the expansion of $I(\delta, \alpha)$ given in Eq. (3.3). We omit the details of these calculations and only give the results.

Case b1: $\delta \neq 0$, $\tau \neq -\frac{1}{2}\pi$.

$$\begin{aligned}\int_0^{2d} \frac{\partial^2}{\partial\delta^2} I(\delta, \tfrac{3}{2}\pi)e^{ik\rho\sin\tau}\frac{d\rho}{\rho} &\sim \frac{1}{6\pi}\sin\tfrac{2}{3}\delta \\ &\times \left\{\frac{-2}{3[\cos(\frac{2}{3}\tau + \frac{1}{3}\pi) - \cos\frac{2}{3}\delta]}\right. \\ &+ \frac{2}{3[\cos\frac{2}{3}\tau + \cos\frac{2}{3}\delta]} - \frac{2}{3[\cos(\frac{2}{3}\tau - \frac{1}{3}\pi) - \cos\frac{2}{3}\delta]} \\ &+ \frac{\delta(1 - \cos\delta\sin\tau)}{\sin\frac{2}{3}\delta(\cos\delta - \sin\tau)^2} \\ &+ \frac{(\delta - 3\pi\operatorname{sgn}\delta)(1 + \cos\delta\sin\tau)}{\sin\frac{2}{3}\delta(\cos\delta + \sin\tau)^2} \\ &\left. - \frac{2(2\tau - \pi)\sin(\frac{2}{3}\tau - \frac{1}{3}\pi)}{9[\cos(\frac{2}{3}\tau - \frac{1}{3}\pi) - \cos\frac{2}{3}\delta]^2}\right\} \\ &+ \frac{2(2\tau + 3\pi)\sin\frac{2}{3}\tau}{9[\cos\frac{2}{3}\tau + \cos\frac{2}{3}\delta]^2} \\ &- \frac{2(2\tau - 5\pi)\sin(\frac{2}{3}\tau + \frac{1}{3}\pi)}{9[\cos(\frac{2}{3}\tau + \frac{1}{3}\pi) - \cos\frac{2}{3}\delta]^2} + e^{i[2kd(1+\sin\tau)+\frac{1}{4}\pi]} \\ &\times \left\{(1 - \sin\tau)\sum_{n=0}^{\infty} i^{n+1}A_n(\delta, \tfrac{3}{2}\pi)\Gamma(n + \tfrac{1}{2})(2kd)^{-n+\frac{1}{2}}\right. \\ &- \sum_{n=0}^{\infty} i^n(n + \tfrac{1}{2})A_n(\delta, \tfrac{3}{2}\pi)\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}} \\ &+ ik\cos\tau\sum_{n=0}^{\infty} i^n\frac{\partial}{\partial\tau}[B_n(\delta, \tfrac{3}{2}\pi, \tau)]\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}} \\ &\left. - (2k^2d\cos^2\tau + ik\sin\tau)\sum_{n=0}^{\infty} i^n \right. \\ &\left. \times B_n(\delta, \tfrac{3}{2}\pi, \tau)\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}\right\}. \end{aligned} \tag{3.45}$$

Case b2: $\delta \neq 0$, $\tau = -\frac{1}{2}\pi$.

$$\int_0^{2d} \frac{\partial^2}{\partial \delta^2} I(\delta, \tfrac{3}{2}\pi) e^{-ik\rho} \frac{d\rho}{\rho}$$
$$\sim ikT(\delta) - 2/(3\pi^{\frac{1}{2}}) e^{-i\frac{1}{4}\pi} (kd)^{\frac{1}{2}} \cot \tfrac{1}{3}\delta$$
$$+ e^{i\frac{1}{4}\pi}\Bigg[-2 \sum_{n=0}^{\infty} i^n (n+1) A_{n+1}(\delta, \tfrac{3}{2}\pi)$$
$$\times \Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}} - \sum_{n=0}^{\infty} i^n (n + \tfrac{1}{2})$$
$$\times A_n(\delta, \tfrac{3}{2}\pi)\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}\Bigg]. \qquad (3.46)$$

Case b3: $\tau = -\frac{1}{2}\pi$, δ in the neighborhood of zero.

$$\int_0^{2d} \frac{\partial^2}{\partial \delta^2} I\left(\delta, \frac{3\pi}{2}\right) e^{-ik\rho} \frac{d\rho}{\rho} \sim \frac{\operatorname{sgn} \delta}{\pi^{\frac{1}{2}}} e^{-i\frac{1}{4}\pi}$$
$$\times \left(\frac{(2kd)^{\frac{1}{2}}(1 + \cos \delta)}{2(1 - \cos \delta)^{\frac{1}{2}}} + S\{[2kd(1 - \cos \delta)]^{\frac{1}{2}}\}\right.$$
$$\left. \times e^{2ikd(\cos \delta - 1)}\left[2ikd(1 + \cos \delta) + \frac{1}{1 - \cos \delta}\right]\right)$$
$$+ ikT(\delta) - 2/(3\pi^{\frac{1}{2}}) e^{-i\frac{1}{4}\pi} (kd)^{\frac{1}{2}} \cot \tfrac{1}{3}\delta$$
$$+ e^{i\frac{1}{4}\pi}\Bigg\{-2 \sum_{n=0}^{\infty} i^n [A_{n+1}(\delta, \tfrac{3}{2}\pi) - A_{n+1}(\delta, 2\pi)$$
$$- A_{n+1}(2\pi - \delta, 2\pi)](n + 1)\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}$$
$$- \sum_{n=0}^{\infty} i^n [A_n(\delta, \tfrac{3}{2}\pi) - A_n(\delta, 2\pi) - A_n(2\pi - \delta, 2\pi)]$$
$$\times (n + \tfrac{1}{2})\Gamma(n + \tfrac{1}{2})(2kd)^{-n-\frac{1}{2}}\Bigg\}. \qquad (3.47)$$

We also require the expansions of integrals of the type given in (3.33b) with $j = 1$. It may be shown that

$$\lim_{k\rho \to 0} \frac{\partial}{\partial \delta} I(\delta, \alpha) = (2\alpha)^{-1}.$$

Hence these integrals are divergent. But we shall always be concerned with differences of the form

$$\int_0^{2d} \frac{\partial}{\partial \delta} [I(\delta_1, \alpha) - I(\delta_2, \alpha)] e^{ik\rho \sin \tau} \frac{d\rho}{\rho},$$

and it is easily seen that these do converge.

The asymptotic expansions of these integrals ($j = 1$) may be found by integrating the corresponding expansions for $j = 2$ in Cases b1, b2, and b3 with respect to δ. They are calculated in Ref. 8.

[8] B. J. Morse, Ph.D. thesis, New York University, June 1963.

D. Scattering Cross Sections at Grazing Incidence

When the direction of the incident wave is parallel to a side of the cylinder, i.e., $\varphi' = \pi$ in Fig. 1, we define the incidence as grazing. In this situation the two trailing edges E_3 and E_4 lie in the shadow of the leading edges E_1 and E_2. Furthermore, the direction of the forward far field, $\varphi = \frac{1}{2}\pi$ in Fig. 1, is also parallel to a side. Both these circumstances require that the precedure used for oblique incidence be modified. These details are illustrated below in finding the scattering cross sections for hard and soft rectangular cylinders. The latter will be considered first.

The Soft Cylinder

We first calculate the interacting fields at the vertices of the cylinder. As in our previous computations, only terms up to order $(ka)^{\frac{1}{2}}$ will be retained. As shown in Fig. 12, $P(\xi, \eta)$ is any point in the neighborhood of E_3 such that $\eta > 0$. Then the diffracted field at P due to the incident field at E_2, $u_d(P; E_2)$ is found from (3.2), (3.3), (3.13) and (3.14). We define $u_1(E_3) = u_i(P) + u_d(P; E_2)$. Hence

$$u_1(E_3) = u_1(E_4) = \frac{e^{ikb} e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}}\left[\frac{iv'(0)}{(kb)^{\frac{1}{2}}}\right.$$
$$\left. + \frac{17v'(0) + 12v'''(0)}{144(kb)^{\frac{3}{2}}}\right] + O(kb)^{-5/2}, \qquad (3.48)$$

where $u_1(E_3) = u_1(E_4)$ by symmetry.

Consistent with our definition for $u_1(E_3)$, the only singly diffracted field reaching E_1 comes from E_2. Hence from (3.8) we obtain

$$u_1(E_1) = u_1(E_2) = -e^{-ikb} L(a, a)$$
$$\times [D_\varphi(\tfrac{1}{2}\pi, 0, \tfrac{3}{2})h'(0)] + O(ka)^{-5/2}. \qquad (3.49)$$

It is easily shown that

$$u_n(E_i) = 0 + O(ka)^{-2}, \qquad n \geq 2.$$

Now, having found the vertex fields, we proceed to obtain the forward far field. For reasons previously explained, Green's theorem will be employed in this task. Letting u_{ij} be the total surface field along side $\overline{E_iE_j}$ of the cylinder and observing that $u_{ij} = 0$ in the soft case, we have from (50)

$$u_s\left(r, \frac{\pi}{2}\right) = -\frac{e^{i(kr + \frac{1}{4}\pi)}}{2(2\pi kr)^{\frac{1}{2}}}\left\{e^{ikb}\left[\int_0^{2b} \frac{\partial u_{23}}{\partial n} e^{-ik\rho_2}\, d\rho_2\right.\right.$$
$$\left. + \int_0^{2b} \frac{\partial u_{41}}{\partial n} e^{-ik\rho_1}\, d\rho_1\right] + e^{ikb} \int_0^{2a} \frac{\partial u_{12}}{\partial n}\, d\rho_2$$
$$\left. + e^{-ikb} \int_0^{2a} \frac{\partial u_{34}}{\partial n}\, d\rho_3\right\} + O(kr)^{-\frac{3}{2}}. \qquad (3.50)$$

As seen from (2.6) the fields $\partial u_{ij}/\partial n$ are proportional to the axial surface currents. Our first step will be to calculate these quantities with the help of the vertex fields found above. The notation

$$[\partial u_{ij}/\partial n \mid u_h(E_m)]$$

denotes the component of $\partial u_{ij}/\partial n$ contributed by the vertex field $u_h(E_m)$.

Determination of $\partial u_{34}/\partial n$: Referring to Fig. 13, we have for the diffracted field from E_3 by (3.1) and (3.2)

$$u_d(\rho, \varphi) = I(\pi - \varphi - \varphi') + I(-2\pi + \varphi - \varphi') - I(\pi - \varphi - \varphi') - I(-2\pi + \varphi + \varphi').$$

But since n is the outer normal, along the side $\overline{E_3E_4}$,

$$\frac{\partial^2 u_d}{\partial n\, \partial\varphi'} = \frac{\partial^2 I(-\frac{1}{2}\pi + \varphi')}{\rho_3\, \partial\varphi\, \partial\varphi'} - \frac{\partial^2 I(-\frac{1}{2}\pi - \varphi')}{\rho_3\, \partial\varphi\, \partial\varphi'} - \frac{\partial^2 I(-\frac{1}{2}\pi - \varphi')}{\rho_3\, \partial\varphi\, \partial\varphi'} + \frac{\partial^2 I(-\frac{1}{2}\pi + \varphi')}{\rho_3\, \partial\varphi\, \partial\varphi'}.$$

Hence evaluating for $\varphi' = 0$ and remembering that in this case we must divide by two,

$$\partial^2 u_d(\rho, \tfrac{3}{2}\pi)/\partial n\, \partial\varphi' \mid_{\varphi'=0} = 2[\partial^2 I(-\tfrac{1}{2}\pi)/\rho_3\, \partial\delta^2].$$

Similarly we obtain

$$\partial^4 u_d(\rho, \tfrac{3}{2}\pi)/\partial n\, \partial^3\varphi' \mid_{\varphi'=0} = 2[\partial^4 I(-\tfrac{1}{2}\pi)/\rho_3\, \partial\delta^4].$$

Thus from (3.48)

$$\left[\frac{\partial u_{34}}{\partial n} \mid u_1(E_3)\right] = \frac{e^{ikb}e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}}\left\{\frac{2i}{(kb)^{\frac{1}{2}}}\frac{\partial^2 I(-\frac{1}{2}\pi)}{\rho_3\, \partial\delta^2} + \frac{17[\partial^2 I(-\frac{1}{2}\pi)/\rho_3\, \partial\delta^2] + 12[\partial^4 I(-\frac{1}{2}\pi)/\rho_3\, \partial\delta^4]}{72(kb)^{\frac{1}{2}}}\right\} + O(kb)^{-5/2}. \tag{3.51}$$

By symmetry (except that ρ_4 replaces ρ_3),

$$[\partial u_{34}/\partial n \mid u_1(E_4)] = [\partial u_{34}/\partial n \mid u_1(E_3)]. \tag{3.52}$$

The effect of $u_i(E_3)$ has been accounted for in our definition of $u_1(E_3)$, that is, (3.48). Similarly the components of $\partial u_{23}/\partial n$, $\partial u_{12}/\partial n$, and $\partial u_{14}/\partial n$ may be found. We omit these except to note that

$$\left[\frac{\partial u_{23}}{\partial n} \mid u_i(E_2)\right] = e^{-ikb}\left[\frac{\partial I(0^+)}{\rho_2\, \partial\delta} + \frac{\partial I(0^-)}{\rho_2\, \partial\delta} - 2\frac{\partial I(-\pi)}{\rho_2\, \partial\delta}\right], \tag{3.53}$$

where 0^+ and 0^- indicate the limits of δ going to zero through positive and negative real values, respectively.

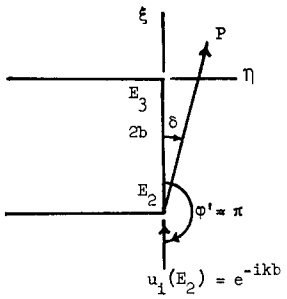


FIG. 12. Configuration for calculating the vertex fields at grazing incidence.

Substituting these values into (3.50), we obtain

$$u_s(r, \tfrac{1}{2}\pi) = -\frac{e^{i(kr+\frac{1}{4}\pi)}}{2(2\pi kr)^{\frac{1}{2}}}\left\{-2ik\int_0^{2a} d\rho + 4\int_0^{2a}\left[\frac{\partial I(\frac{1}{2}\pi)}{\partial\delta} - \frac{\partial I(\frac{3}{2}\pi)}{\partial\delta}\right]\frac{d\rho_2}{\rho_2} + 2\int_0^{2b}\left[\frac{\partial I(0^+)}{\partial\delta} + \frac{\partial I(0^-)}{\partial\delta} - 2\frac{\partial I(-\pi)}{\partial\delta}\right]e^{-ik\rho_4}\frac{d\rho_2}{\rho_2} - 4\frac{e^{i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}(kb)^{\frac{1}{2}}}\int_0^{2b}\frac{\partial^2 I(\pi)}{\partial\delta^2}e^{ik\rho_3}\frac{d\rho_3}{\rho_3} + \frac{4e^{i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}(kb)^{\frac{1}{2}}}\int_0^{2a}\frac{\partial^2 I(-\frac{1}{2}\pi)}{\partial\delta^2}\frac{d\rho_3}{\rho_3} + O(kd)^{-\frac{1}{2}}\right\} + O(kr)^{-\frac{1}{2}}. \tag{3.54}$$

The expansions required for evaluating these integrals have been found in Sec. 3C. Thus, for instance,

$$\int_0^{2b}\left[\frac{\partial I(0^+)}{\partial\delta} + \frac{\partial I(0^-)}{\partial\delta} - 2\frac{\partial I(-\pi)}{\partial\delta}\right]e^{-ik\rho}\frac{d\rho}{\rho} = \lim_{\delta\to 0^+}\int_0^{2b}\frac{\partial I(\delta)}{\partial\delta}e^{-ik\rho}\frac{d\rho}{\rho} + \lim_{\delta\to 0^-}\int_0^{2b}\frac{\partial I(\delta)}{\partial\delta}e^{-ik\rho}\frac{d\rho}{\rho} - 2\lim_{\delta\to -\pi}\int_0^{2b}\frac{\partial I(\delta)}{\partial\delta}e^{-ik\rho}\frac{d\rho}{\rho} \sim -\frac{2\sqrt{3}}{9} + \frac{e^{-i\frac{1}{4}\pi}}{(2\pi)^{\frac{1}{2}}}4(2kb)^{\frac{1}{2}} + \frac{e^{i\frac{1}{4}\pi}}{(2\pi)^{\frac{1}{2}}}\frac{5}{18}(2kb)^{-\frac{1}{2}} + O(kb)^{-\frac{3}{2}}. \tag{3.55}$$

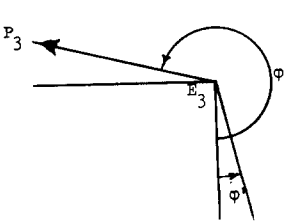


FIG. 13. Angular relationships used in the determination of the normal derivative of the surface field along side $\overline{E_3E_4}$.

Substituting these evaluations into (3.54) yields the scattered forward far field

$$u_s(r, \tfrac{1}{2}\pi) = -\frac{e^{i(kr-\frac{1}{4}\pi)}}{(2\pi kr)^{\frac{1}{2}}}\left[2ka + \frac{4e^{i\frac{1}{4}\pi}}{(2\pi)^{\frac{1}{2}}}(2kb)^{\frac{1}{2}} - \frac{19e^{-i\frac{1}{4}\pi}}{18(2\pi)^{\frac{1}{2}}}(2kb)^{-\frac{1}{2}} + O(ka)^{-\frac{3}{2}}\right] + O(kr)^{-\frac{3}{2}}. \quad (3.56)$$

Hence applying the cross-section theorem,

$$\sigma_s = 4a + (4/k\pi^{\frac{1}{2}})(2kb)^{\frac{1}{2}} - (19/18k\pi^{\frac{1}{2}})(2kb)^{-\frac{1}{2}} + O[k^{-1}(ka)^{-\frac{3}{2}}]. \quad (3.57)$$

The Hard Cylinder

Proceeding as in the soft case, the interacting vertex fields are obtained first. We shall give our results to $O(ka)^{-\frac{1}{2}}$, although, of course, with additional computation it is possible to obtain any order desired. Since the computational details are similar to the preceding, they are omitted.

Since $\partial u_{ij}/\partial n = 0$ in the present case, (3.32) yields

$$u_s\left(r, \frac{\pi}{2}\right) = \tfrac{1}{2}k\frac{e^{i(kr-\frac{1}{4}\pi)}}{(2\pi kr)^{\frac{1}{2}}}\left(e^{-ikb}\int_0^{2a} u_{34}\, d\rho_3 - e^{ikb}\int_0^{2a} u_{12}\, d\rho_2\right) + O(kr)^{-\frac{3}{2}}. \quad (3.58)$$

As indicated in (2.7), the quantities u_{ij} are equal to the tangential surface currents. They are calculated using the vertex fields as in the soft case.

Substituting the values of u_{ij} into (3.58) we obtain

$$u_s(r, \tfrac{1}{2}\pi) = \tfrac{1}{2}k\frac{e^{i(kr-\frac{1}{4}\pi)}}{(2\pi kr)^{\frac{1}{2}}}\left[-4a + 8\int_0^{2a} I(-\tfrac{1}{2}\pi)\, d\rho - 4\int_0^{2a} I(\tfrac{3}{2}\pi)\, d\rho + \frac{4e^{i\frac{1}{4}\pi}}{3\sqrt{3\pi}}\frac{(1+e^{4ikb})}{(kb)^{\frac{1}{2}}} \times \int_0^{2a} I(-\tfrac{1}{2}\pi)\, d\rho + \frac{8e^{i\frac{1}{4}\pi}}{(3\pi)^{\frac{1}{2}}}\frac{e^{2ika}}{(ka)^{\frac{1}{2}}} \times \int_0^{2a} I(\pi)\, d\rho + O(k^2a)^{-1}\right] + O(kr)^{-\frac{3}{2}}. \quad (3.59)$$

The expansions of the integrals in Eq. (3.59) are obtained from (3.36). Placing these evaluations in (3.59), and collecting terms,

$$u_s(r, \tfrac{1}{2}\pi) = \frac{e^{i(kr-\frac{1}{4}\pi)}}{(2\pi kr)^{\frac{1}{2}}}\left[-2ka + \frac{4i\sqrt{3}}{9} + \frac{4e^{-i\frac{1}{4}\pi}e^{2ika}}{3\pi^{\frac{1}{2}}(ka)^{\frac{1}{2}}} - \frac{2e^{-i\frac{1}{4}\pi}(1+e^{4ikb})}{27\pi^{\frac{1}{2}}(kb)^{\frac{1}{2}}} + O(ka)^{-1}\right] + O(kr)^{-\frac{3}{2}}. \quad (3.60)$$

Hence applying the cross-section theorem,

$$\sigma_H = 4a + \frac{2\sqrt{2} + 4\sin(4kb + \frac{1}{4}\pi)}{27k\pi^{\frac{1}{2}}(kb)^{\frac{1}{2}}} - \frac{8\sin(2ka + \frac{1}{4}\pi)}{3k\pi^{\frac{1}{2}}(ka)^{\frac{1}{2}}} + O(k^{-1}[ka]^{-1}). \quad (3.61)$$

E. Far-Field Radiation Patterns

The calculation of the far-field radiation pattern is a straightforward application of the geometrical theory. Equation (3.8) alone is sufficient in all but certain exceptional directions. They correspond to directions parallel to a side of the cylinder ($\varphi = 0$, $\frac{1}{2}\pi$, π, $\frac{3}{2}\pi$), along specular lines ($\varphi = \varphi' - \frac{1}{2}\pi$, $\varphi' + \frac{1}{2}\pi$), or along the shadow line ($\varphi = \frac{3}{2}\pi - \varphi'$).

The problems arising in these cases have already been met in our calculation of the scattering cross sections. Indeed, the far field along a shadow line has been found in (3.24) and (3.27). Thus the same methods which have been previously applied again suffice. The patterns for the soft cylinder at oblique incidence and for the hard cylinder at grazing incidence are shown in Figs. 7 and 8. The complete set of formulas from which these figures are calculated are contained in Ref. 8. Those contained herein are (3.24) and (3.60).

4. DIFFRACTION BY CONVEX POLYGONAL CYLINDERS

Although the foregoing calculations have been carried out for a rectangular cylinder, nothing essentially new is needed to apply them to any convex polygonal cylinder whose sides are large with respect to the incident wavelength. This section will serve to recapitulate the method as well as to provide an outline of its application to these more general cylinders.

Figure 14 illustrates our terminology. It is assumed that we are dealing with a convex cylinder of s sides, $s \geq 2$, with lengths $2a_i (ka_i \gg 1)$ and exterior angles α_i, $i = 1, \cdots, s$. The coordinates of vertex E_i measured with respect to suitable axes are (x_i, y_i). The origin of these axes is arbitrary but we assume the x axis is parallel to side $\overline{E_sE_1}$. ρ_i is the distance from E_i to any point on the side $\overline{E_iE_{i+1}}$. The angle θ_j defined by

$$\theta_j = \sum_{i=1}^{j} \alpha_i - j\pi, \qquad j = 1, \cdots, s, \quad (4.1)$$

is the inclination of the side $\overline{E_jE_{j+1}}$ to the x axis. The steps below are listed in the order in which the calculation proceeds.

A. Computation of Vertex Fields

(1) Oblique incidence, i.e., direction of u_i not parallel to any side: In this case the vertex fields are calculated by application of (3.8). The only

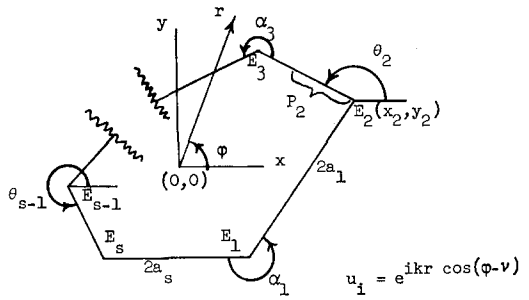


FIG. 14. Notation used in diffraction by a polygonal cylinder of arbitrary cross section.

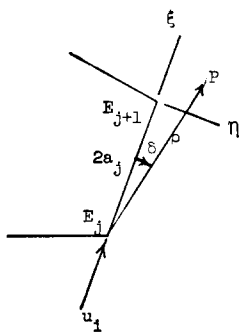


FIG. 15. Configuration for finding vertex field along a grazing side of a general polygonal cylinder.

difference is that n, as given in (3.7), now varies from vertex to vertex:

$$n_i = \alpha_i/\pi, \qquad i = 1, \cdots, s. \tag{4.2}$$

(2) Grazing incidence, i.e., direction of u_i parallel to one or more sides: Here the procedure is the same as in the oblique case except when calculating the response to u_i at a vertex which is in the shadow of the excited vertex. Since (3.8) is not valid along shadow lines, it cannot be used for this purpose. Let $\overline{E_iE_{i+1}}$ be such a pair of vertices and assume that E_{i+1} is in the shadow of E_i. If P is any point in the geometrically lit region near E_{i+1} we observed from (3.1) and Fig. 15,

$$u_d(P; E_i) = u_i(E_i)[I(\delta, \alpha_i) \mp I(\delta, \alpha_i) + I(2\pi - \delta, \alpha_i) \pm I(2\pi + \delta, \alpha_i)]$$

where the upper signs refer to the hard cylinder, and the lower to the soft. Hence using (3.13) and equations analogous to (3.14), we obtain

$$u_d(P; E_i) = u_i(E_i)\begin{Bmatrix}0\\1\end{Bmatrix}\left(-e^{2ika_j}p(0) + \frac{e^{-i\frac{1}{4}\pi}e^{2ika_j}}{\pi^{\frac{1}{2}}}\right.$$
$$\times\left[\frac{ip'(0)}{(ka_j)^{\frac{1}{2}}} + \frac{1}{24}\left\{\frac{3}{2} + \frac{\pi^2}{\alpha_j^2}\left[3\csc^2\left(\frac{\pi^2}{\alpha_j}\right) - 1\right]\right\}\frac{p'(0)}{(ka_j)^{\frac{1}{2}}} + \frac{p'''(0)}{12(ka_j)^{\frac{3}{2}}}\right]\Bigg) - u_i(E_i)\begin{Bmatrix}1\\0\end{Bmatrix}\frac{\pi^{\frac{1}{2}}e^{-i\frac{1}{4}\pi}e^{2ika_j}}{4\alpha_j}$$
$$\times\left\{\frac{2ip(0)\cot\left(\frac{\pi^2}{\alpha_j}\right)}{(ka_j)^{\frac{1}{2}}} + \frac{p(0)\cot\left(\frac{\pi^2}{\alpha_j}\right)\left[1 + \frac{2\pi^2}{\alpha_j^2}\csc^2\left(\frac{\pi^2}{\alpha_j}\right)\right] + 4p''(0)\cot\left(\frac{\pi^2}{\alpha_j}\right)}{8(ka_j)^{\frac{3}{2}}}\right\} + O(ka_j)^{-5/2}. \tag{4.3}$$

Here the upper figures apply to hard cylinders and the lower to soft. This formula supplants (3.8) in this case. A particular case of the above is seen in the derivation of (3.48).

B. Computation of the Far Field

(1) In directions not along shadow or specular lines, and not parallel to a side: The far field is computed from the appropriate vertex fields using (3.8). According to the geometrical theory, only those vertices visible from the field point contribute.

(2) In directions along shadow or specular lines, but not parallel to side: Here the procedure is the same as described in paragraph B(1) above, except for those contributing vertices which are illuminated by the incident wave. Equation (3.8) is not valid for finding the components of the far field contributed by the incident wave excitation at these vertices. If E_i is such a vertex and P_i a field point in the neighborhood of a shadow or specular line, then, typically,

$$u_d(P_i; E_i) = u_i(E_i)[I(\delta, \alpha_i) + I(\omega_1, \alpha_i) \pm I(\omega_2, \alpha_i) \pm I(\omega_3, \alpha_i)],$$

where δ, ω_1, ω_2, and ω_3 are appropriate angles such that δ is small and each of the ω_n is of the form $\omega_n = \Omega_n \pm \delta$, $|\Omega_n| \gg 0$. Then, referring to Fig. 16, we use the following formulas to find the far field:

$$I(\delta, \alpha_i) = \operatorname{sgn}\delta e^{ik(r_i+\xi_i)} \times\left[-\frac{1}{2} + \frac{e^{-i\frac{1}{4}\pi}}{\pi^{\frac{1}{2}}}\frac{k|\eta_i|}{(2kr_i)^{\frac{1}{2}}} + O(kr_i)^{-\frac{1}{2}}\right]$$
$$I(\Omega \pm \delta, \alpha_i) = -\frac{\pi^{\frac{1}{2}}e^{i\frac{1}{4}\pi}e^{ik(r_i+\xi_i)}}{2^{\frac{1}{2}}\alpha_i(kr_i)^{\frac{1}{2}}} \times\cot\frac{\pi\Omega}{2\alpha_i} + O(kr_i)^{-\frac{1}{2}}. \tag{4.4}$$

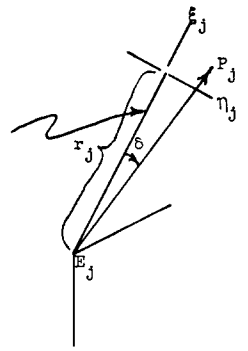


FIG. 16. Computation of far field along a shadow or specular line.

Equations (3.20) and (3.21) are particular cases of (4.4). This process is illustrated in detail, for $\alpha_j = \frac{3}{2}\pi$, in the derivation of (3.22).

(3) In directions parallel to sides: The far field in this case is determined by means of Green's theorem. The components of the field (or its normal derivative) along the sides of the cylinder are determined from the vertex fields [cf. (3.51)]. In analogy with (3.32) we have

$$u_s(r,\varphi) = \frac{e^{i(kr-\frac{1}{4}\pi)}}{2(2\pi kr)^{\frac{1}{2}}} \sum_{j=1}^{s} e^{-ik(x_j\cos\varphi + y_j\sin\varphi)} \times \int_0^{2a_j} \left[-ku\sin(\varphi-\theta_j) - i\frac{\partial u}{\partial n}\right] \times e^{-ik\rho_j\cos(\varphi-\theta_j)}\,d\rho_j + O(kr)^{-\frac{3}{2}}, \tag{4.5}$$

where n is the outer normal. Hence, as in the rectangular case, we must find asymptotic representations for integrals of the forms given in (3.33a) and (3.33b).

The methods used for finding these representations are illustrated in Sec. 3C of this paper for $\alpha = \frac{3}{2}\pi$. The same methods apply for any α which is a rational multiple of π. Indeed, assuming $\alpha = m\pi/n$, where m and n are positive integers, and

$$R_j = -\left(2m\cos\tau\left\{\cos\left[\frac{n}{m}\left(\tau+\frac{\pi}{2}+2\pi j\right)\right] - \cos\frac{n\delta}{m}\right\}\right)^{-1}, \tag{4.6}$$

$$T_h = -\left\{n\sin\frac{n\delta}{m} \times \left[\cos\left(\delta + \operatorname{sgn}\delta\,\frac{2\pi mh}{n}\right) + \sin\tau\right]\right\}^{-1};$$

then, as shown in the appendix,

$$I \equiv \int_0^\infty \frac{dx}{(\cosh x + \sin\tau)[\cosh(\pi x/\alpha) - \cos(\pi\delta/\alpha)]} = \sum_{j=0}^{m-1}[2\tau + \pi(4j-2m+1)]R_j + \sum_{h=0}^{n-1}\left[\delta + m\pi\operatorname{sgn}\delta\left(\frac{2h}{n}-1\right)\right]T_h, \tag{4.7}$$

where $\frac{3}{2}\pi > \tau > -\frac{1}{2}\pi$ and $2\pi m/n > |\delta| > 0$. Using this result, in place of its particular value when $\alpha = \frac{3}{2}\pi$, in the computation of the integral representations in Sec. 3C, will give the corresponding results for α any rational multiple of π.

ACKNOWLEDGMENT

The author wishes to express his appreciation to Professor Joseph B. Keller for suggesting this problem and for his aid and guidance in carrying out the research.

APPENDIX: EVALUATION OF AN INTEGRAL

We evaluate the integral I defined in (4.7). Letting $\alpha = m\pi/n$, m and n positive integers and $y = e^{x/m}$,

$$I = 4m\int_1^\infty \frac{y^{m+n-1}dy}{(y^{2m}+2y^m\sin\tau+1)(y^{2n}-2y^n\cos n\delta/m+1)}.$$

By partial fraction decomposition, the integrand equals

$$\sum_{j=0}^{m-1}\left[\frac{A_j}{y - e^{i/m(\tau+\frac{1}{2}\pi+2\pi j)}} + \frac{B_j}{y - e^{-i/m(\tau+\frac{1}{2}\pi+2\pi j)}}\right] + \sum_{h=0}^{n-1}\left[\frac{C_h}{y - e^{-i[\delta/m+\operatorname{sgn}\delta(2\pi h/n)]}} + \frac{D_h}{y - e^{-i[\delta/m+\operatorname{sgn}\delta(2\pi h/n)]}}\right],$$

where

$$A_j = \frac{-i}{4m\cos\tau\{\cos[n/m(\tau+\frac{1}{2}\pi+2\pi j)] - \cos(n\delta/m)\}}, \qquad B_j = -A_j,$$

$$C_h = \frac{-i}{4n\sin(n\delta/m)\{\cos[\delta+\operatorname{sgn}\delta(2\pi mh/n)] + \sin\tau\}}, \qquad D_h = -C_h,$$

$$\therefore\quad I = 4m\left\{\sum_{j=0}^{m-1} A_j\log\frac{1-e^{-i/m(\tau+\frac{1}{2}\pi+2\pi j)}}{1-e^{i/m(\tau+\frac{1}{2}\pi+2\pi j)}} + \sum_{h=0}^{n-1} C_h\log\frac{1-e^{-i[\delta/m+\operatorname{sgn}\delta(2\pi h/n)]}}{1-e^{i[\delta/m+\operatorname{sgn}\delta(2\pi h/n)]}}\right\}.$$

Hence if we let R_j and T_h be as defined in (4.6), then the result stated in (4.7) follows.

Note added in proof. In a forthcoming paper, the method of Sec. 4 will be used to obtain specific formulas for the scattering cross sections of arbitrary convex polygonal cylinders.

Accuracy of the Semicircle Approximation for the Density of Eigenvalues of Random Matrices

BURT V. BRONK
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 6 August 1963)

Certain statistical properties of the energy levels of complex physical systems have been found to coincide with those for distributions of eigenvalues derived from ensembles of random matrices. However, if ensembles of random matrices give a fair representation for the Hamiltonian of a complex physical system, the density of the characteristic values at the lower end of the spectrum should show some similarity with the exponential dependence found in nuclear spectra. The limiting distribution of the density for very high-dimensional random matrices is a semicircle, i.e., concave from below if plotted against the characteristic value which represents, in this case, the energy. Hence the deviations from the limiting distribution are investigated and it is shown that there is a region, at the very lowest part of the spectrum, where the density is convex from below, similar to an exponential function. The region of convexity is called the tail of the distribution. It is shown, however, that the avergae number of roots in the tail is very small, of the order of 1. It is concluded that those ensembles of random matrices which have been studied up to now, do not give a fair representation of Hamiltonians of complex systems.

I. INTRODUCTION

THE eigenvalue distribution for an ensemble of random Hermitian matrices represented by the Wishart-like density function

$$D(H) = A \exp\left[(-\text{trace } H^2)/4a^2\right] \tag{1}$$

has been determined by Wigner.[1] $D(H)$ gives the probability of finding a matrix in the ensemble with elements in the neighborhood around the elements of a given matrix H. The real and imaginary parts of each matrix element are independent random variables, each having a Gaussian distribution with the variance of the off-diagonal elements equal to a^2. A is a normalization constant, and $4a^2$ is set equal to one for convenience. The dimension of the matrices is considered specified, and will be denoted by n. The probability for the appearance of an eigenvalue in the neighborhood of ϵ was found to be

$$P_n(\epsilon) = (1/n) \sum_{\nu=1}^{n} \phi_\nu^2(\epsilon), \tag{2}$$

where $\phi\nu$ is the νth Hermite function, normalized in the interval $(-\infty, \infty)$.

$$\phi_\nu(\epsilon) = \pi^{-\frac{1}{4}}(2^{\nu-1}(\nu-1)!)^{-\frac{1}{2}} \exp(-\epsilon^2/2) H_{\nu-1}(\epsilon),$$

$$H_\nu(\epsilon) = (-)^\nu \exp(\epsilon^2) d^\nu(\exp(-\epsilon^2))/d\epsilon^\nu. \tag{3}$$

The asymptotic form for $P_n(\epsilon)$, as n becomes large, was found to be

$$nP_n(\epsilon) = (2n - \epsilon^2)^{\frac{1}{2}} \quad \text{for} \quad \epsilon^2 < 2n, \tag{4a}$$

$$nP_n(\epsilon) = 0 \quad \text{for} \quad \epsilon^2 \geq 2n. \tag{4b}$$

The present article is concerned with the precision of these equations near the lower end of the spectrum, which is of particular interest. This is the neighborhood of $\epsilon = -(2n)^{\frac{1}{2}}$, but since the distribution is an even function of ϵ, we may as well investigate the neighborhood of $\epsilon = +(2n)^{\frac{1}{2}}$, since keeping track of signs will be easier for positive argument. The corrections which we expect must be small compared with $n^{\frac{1}{2}}$—otherwise (4) would not be valid—but the *total* number of roots in the tail could still increase with increasing n. Suppose for instance that $P_n(\epsilon)$ were proportional to $n^{\frac{1}{2}}$ in the neighborhood of $\epsilon = (2n)^{\frac{1}{2}}$ extending over an interval which is possibly also proportional to $n^{\frac{1}{2}}$. In this case, the number of roots which are not accounted for by the limiting law (4), would increase with increasing n. However, as was mentioned in the abstract, it turns out that this is not the case. Actually, it will turn out that the accurate nP_n differs from that given only by $n^{\frac{1}{2}}$, and that the region in which $nP_n(\epsilon)$ is convex from below—and thus may show similarity to an exponential—is proportional to $n^{-\frac{1}{2}}$.

The fact that, for very large (or very small) ϵ, the distribution (2) is convex from below, can be

[1] E. P. Wigner, "Distribution Laws for Roots of a Random Hermitian Matrix" (unpublished). The notation of this reference is used in the present article. It should be pointed out that the semicircle distribution is actually appropriate for a very large class of ensembles, which contains the real symmetric ensemble (Sec. IV) as a special case. This is discussed by Wigner in Ann. Math. **67**, 325 (1958). Cf. *ibid.* **62**, 552 (1955).

[2] (a) Bateman Manuscript Project, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, Sec. 10.13. (b) The correct expression for the Hermite function near the turning point is well known, and could also be obtained by the connection of the Hermite Functions with the Laguerre functions. See Sec. 10.13, Eqs. (2) and (3), and Sec. 10.15. Eqs. (10)–(13), part (a), this reference.

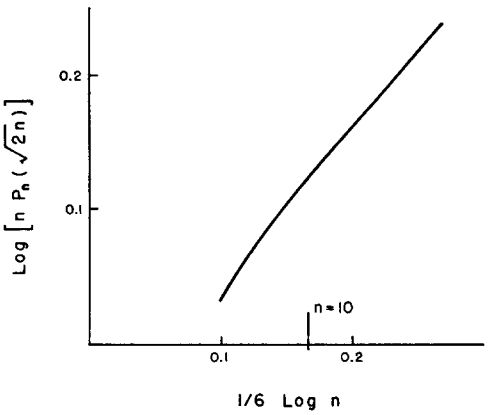


FIG. 1. Logarithmic plot showing rapid approach of numerical value of $\eta P[2\eta)^{\frac{1}{2}}]$ to asymptotic behavior ($\sim\eta^{1/6}$).

inferred already from (2) and (3). For very large ϵ, each φ_ν is convex, i.e., $\varphi_\nu\varphi_\nu''$ is positive. This follows also from the differential equation (5) for these functions. The second derivative of P_n is a sum of expressions $2\varphi_\nu'^2 + 2\varphi_\nu\varphi_\nu''$ which are all positive. Hence P_n is convex for large ϵ—as (5) shows it is merely convex beyond $\epsilon = (2_n - 1)^{\frac{1}{2}}$. However, it is also very small beyond that point.

II. CALCULATION OF THE TAIL OF THE DISTRIBUTION

We note that ϕ_ν is the solution of the one-dimensional harmonic-oscillator equation, with ϕ_1 corresponding to the ground state,

$$-\phi_\nu'' + \epsilon^2\phi_\nu = (2\nu - 1)\phi_\nu. \tag{5}$$

The asymptotic form for ϕ_ν is conveniently obtained in the neighborhood of $\epsilon = (2\nu - 1)^{\frac{1}{2}}$, by the transformation $\epsilon \to \epsilon_t + X$, where

$$\epsilon_t = (2\nu - 1)^{\frac{1}{2}}. \tag{6}$$

This gives for $\tilde{\phi}_\nu(X) = \phi_\nu(\epsilon)$

$$-\tilde{\phi}_\nu'' + 2\epsilon_t X\tilde{\phi}_\nu(X) = 0, \quad \text{for} \quad X^2 \ll 2\epsilon_t X. \tag{7}$$

This is Airy's equation. Since $\phi_\nu(\epsilon_t + X)$ must clearly approach zero very closely, even in the region of validity of (7), $\tilde{\phi}_\nu(X)$ is the decreasing solution of (7). This leaves only the normalization of this solution to be determined.[2] This is obtained by comparing the average value of $[\tilde{\phi}_\nu(X)]^2$ for $X < 0$, $|X| \gg 1$ with the classical approximation for the quantum mechanical probability density for a harmonic oscillator in that region. (This is the reciprocal of the velocity). The solution of (7) which goes to zero for large X is[3]

$$\tilde{\phi}_\nu(X) = i(\tfrac{1}{6}X)^{\frac{1}{2}} \times \exp(\tfrac{1}{6}i\pi)H_{\frac{1}{3}}^{(1)}(\tfrac{2}{3}i(2\epsilon_t)^{\frac{1}{2}}X^{\frac{1}{2}}) \qquad X \geq 0, \tag{8a}$$

or

$$\tilde{\phi}_\nu(X) = [\tfrac{1}{6}(-X)]^{\frac{1}{2}}/\sin\tfrac{1}{3}\pi[J_{-\frac{1}{3}}(\tfrac{2}{3}(-2\epsilon_t X^3)^{\frac{1}{2}}) + J_{\frac{1}{3}}(\tfrac{2}{3}(-2\epsilon_t X^3)^{\frac{1}{2}})], \qquad X \leq 0. \tag{8b}$$

$H_{\frac{1}{3}}^{(1)}$ is a Hankel function of the first kind, $J_{\frac{1}{3}}$ and $J_{-\frac{1}{3}}$ are Bessel functions. The form (8b) is connected with (8a) by

$$H_K^{(1)}(Z) = [J_{-K} - e^{-K\pi i}J_K(Z)]/i\sin K\pi \tag{9}$$

at $X = 0$, and the fact that once a real solution of (7) is chosen, it remains real for all X.

The asymptotic form for (8b) which was used in determining the constant in (8) is

$$\tilde{\phi}_\nu(X) \propto (-X)^{-\frac{1}{4}}\cos(\beta(-X)^{\frac{3}{2}} - \tfrac{1}{4}\pi) \quad \text{for} \quad X < 0. \tag{10}$$

The asymptotic form for (8a) is

$$\tilde{\phi}_\nu(X) = (3\pi\beta)^{-\frac{1}{2}}X^{-\frac{1}{4}}\exp(-\beta X^{\frac{3}{2}}) \quad \text{for} \quad X > 0, \tag{11}$$
$$\beta = (\tfrac{1}{3}2^{\frac{3}{2}})(2\nu)^{\frac{1}{2}}.$$

These are valid for $|x| \gg (2\nu)^{-\frac{1}{3}}$.

For $|x| \ll 1$, we obtain an approximate expression for $\tilde{\phi}_\nu$ by using the series form for $J_{\frac{1}{3}}$ in (8b).

$$\tilde{\phi}_\nu(X) = 1/(6^{\frac{1}{2}}\sin\tfrac{1}{3}\pi)[3^{\frac{1}{3}}/(2\epsilon_t)^{\frac{1}{2}}(-\tfrac{1}{3})! - (2\epsilon_t)^{\frac{1}{2}}X/3^{\frac{1}{3}}(\tfrac{1}{3})!],$$
$$\tilde{\phi}_\nu^{(2)}(X) \approx (C_1/\epsilon_t^{\frac{1}{6}})\exp(-C_2\epsilon_t^{\frac{1}{2}}X), \tag{12}$$
$$C_1 \approx 0.45, \qquad C_2 \approx 0.92,$$
$$\phi_\nu^{(2)}(\epsilon) \equiv \tilde{\phi}_\nu^{(2)}(X) = \tilde{\phi}_\nu^{(2)}(\epsilon - \epsilon_t).$$

It is clear that (12) becomes a good approximation as X approaches zero. We must however establish its validity as an upper bound for ϕ_ν throughout our region of interest, which turns out to be that part of the ϵ axis to the right of

$$\epsilon_i = \epsilon_t - 1.86/(2\nu)^{\frac{1}{2}}. \tag{13}$$

This will be done in the appendix. The expression for $\tilde{\phi}_\nu^{(2)}$ given in (12) will be used for $\tilde{\phi}_\nu$ in the rest of the article.

In order to evaluate $P_n(\epsilon)$ at points near $\epsilon = (2n)^{\frac{1}{2}}$, we must find the correct argument to use in (12) for each term in the sum (2). Set $\nu = (n - K)$; then

$$\epsilon_{t\nu} = (2\nu - 1)^{\frac{1}{2}} \to [2(n - K) - 1]^{\frac{1}{2}} \approx (2n)^{\frac{1}{2}} - \tfrac{1}{2}(2K + 1)/(2n)^{\frac{1}{2}}, \tag{14}$$

so

$$(2n)^{\frac{1}{2}} + y = \epsilon_{t\nu} + X\nu = \epsilon_{t\nu} + [\tfrac{1}{2}(2K + 1)/(2n)^{\frac{1}{2}} + y]. \tag{15}$$

[3] G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1944), 2nd ed. Paragraphs 6.4, 7.2, and 7.21.

Expanding and dropping terms $O(K/n)$ from both the coefficient and the exponential, we get

$$[\phi_{n-K}(X_K)]^2 \approx .2/(2n)^{\frac{1}{6}} \times \exp\{-1.84(2n)^{\frac{1}{2}}[K/(2n)^{\frac{1}{3}} + y]\}, \quad (16)$$

where we noticed that only terms with $K \ll n$ contribute appreciably to (2).

Using (16) in (2),

$$nP_n(\epsilon) \lesssim .2/(2n)^{\frac{1}{6}} \times \exp(-1.84(2n)^{\frac{1}{2}}y) \int_0^n \exp(-1.84K/(2n)^{\frac{1}{3}})\, dK \quad (17)$$

$$\approx [(2n)^{\frac{1}{3}}/9.2] \exp(-1.84(2n)^{\frac{1}{2}}y) \times \{1 - O[\exp(-n^{\frac{1}{3}})]\} \approx \tfrac{1}{10}(2n)^{\frac{1}{6}}, \quad \text{for large } n.$$

The numerical value of $nP_n[(2n)^{\frac{1}{3}}]$ for various values between 2 and 40, was computed on Princeton's IBM 7090 computer using the following exact expression[2] for $\phi_\nu(\epsilon)$:

$$\phi_\nu(\epsilon) = [n!/(2^n\pi^{\frac{1}{2}})]^{\frac{1}{2}} \sum_{k=0}^{\frac{1}{2}(n-1)\,(\text{or }\frac{1}{2}(n-2)\text{, } n \text{ even})} (-)^K(2\epsilon)^{n-2K} \times [\exp \tfrac{1}{2}(-\epsilon^2)]/[K!\,(n-2K)!]. \quad (18)$$

It was found that $nP_n[(2n)^{\frac{1}{3}}]$ indeed varies as $n^{\frac{1}{6}}$, with a numerical constant about 0.8 times that estimated in (17). Hence (17) is an overestimate as expected.

Now, define T_n as the total number of roots one would expect to find with values larger than $(2n)^{\frac{1}{2}}$.

$$T_n \approx \int_{(2n)^{\frac{1}{2}}}^{\infty} P_n(\epsilon)\, d\epsilon \approx \left[\frac{0.2}{(1.84)^2}\right](2n)^{\frac{1}{6}}/(2n)^{\frac{1}{6}} \approx 0.06 \quad \text{for large } n. \quad (19)$$

That is, T_n is independent of n, the dimension of the matrix, if n is large.

III. LOCATION OF THE INFLECTION POINT IN THE DISTRIBUTION FUNCTION (2)

We shall use the convexity of P_n as the criterion that it may show some similarity with the exponential function. Hence, we shall determine the inflection point ϵ_i of P_n outside of which it is convex from below. It was noted already that $|\epsilon_i| < (2n-1)^{\frac{1}{2}}$ but we need a lower limit for $|\epsilon_i|$.

An alternative form of (2) is arrived at by use of the Christoffel–Darboux formula,[4(a)]

$$P_n(\epsilon) = [1/(2n)^{\frac{1}{6}}][\phi'_{n+1}(\epsilon)\phi_n(\epsilon) - \phi'_n(\epsilon)\phi_{n+1}(\epsilon)]. \quad (20)$$

Notice that (20), although compact in appearance, simplifies neither numerical nor analytical calculation of (17), as the right side is the difference of two large numbers. Differentiating (20), and substituting (5) for $\phi''_n(\epsilon)$ gives

$$P'_n(\epsilon) = [-2/(2n)^{\frac{1}{6}}]\phi_{n+1}\phi_n \quad (21)$$

$$P''_n(\epsilon) = [-2/(2n)^{\frac{1}{6}}][\phi'_{n+1}(\epsilon)\phi_n(\epsilon) + \phi'_n(\epsilon)\phi_{n+1}(\epsilon)]. \quad (22)$$

We wish to find the greatest zero of $P''_n(\epsilon)$.

Call the expression inside the brackets, B. At $\epsilon \gg \epsilon_t$, $\phi_n(\epsilon)$ and $\phi_{n+1}(\epsilon)$ are defined so as to be positive, and since they are both decreasing, both terms of B are negative. The derivative of $\phi_n(\epsilon)$ is given by

$$\phi'_n(\epsilon) = [\tfrac{1}{2}(n-1)]^{\frac{1}{2}}\phi_{n-1}(\epsilon) - (\tfrac{1}{2}n)^{\frac{1}{2}}\phi_{n+1}(\epsilon). \quad (23)$$

From (23), we see that as long as ϕ_{n+1} is greater than ϕ_{n-1}, ϕ'_n is negative. As we proceed left, we first come to the outermost maximum of ϕ_{n+1}. Just to the left of this, ϕ'_{n+1} is positive, and ϕ_n is still positive, so the first term of B is now positive. The greatest zero of ϕ_{n+1} is to the right of the greatest zero of ϕ_n, so that the first term of B is still positive when ϕ_{n+1} goes through its greatest zero. Therefore B is positive at this point. Thus the point of inflection is located between the largest zero of ϕ_{n+1}, and the largest zero of ϕ'_{n+1}. From (23), we see that the latter is approximately at the largest zero of ϕ_{n+2}.[4(b)] These two points are very close together, and using Eq. 6.32.5 of reference 4, we locate the inflection point:

$$\epsilon_i \approx (2n)^{\frac{1}{2}} - 1.86/(2n)^{\frac{1}{6}}. \quad (24)$$

Adjusting the lower limit of the integral in (19) in accordance with (24) has only the effect of multiplying T by a constant of order unity.

We could also have located ϵ_i by noticing that it is also to the right of the largest zero of ϕ'_n. This can be found by integrating Schrödinger's equation (5), obtaining

$$\left.\frac{d\phi_n(\epsilon)}{d\epsilon}\right|_{\epsilon_i} = 0 = \int_{(2n)^{\frac{1}{2}}}^{\epsilon_i} [\epsilon^2 - (2n-1)]\phi_n(\epsilon)\, d\epsilon + \int_{\infty}^{(2n)^{\frac{1}{2}}} [\epsilon^2 - (2n-1)]\phi_n(\epsilon)\, d\epsilon. \quad (25)$$

Substitution of the asymptotic form for ϕ_n, (10), into the integrand in (25) leads to the same location for ϵ_i.

IV. REAL SYMMETRIC ENSEMBLE

We consider next the ensemble of real symmetric Hamiltonians which is the appropriate one to use

[4] G. Szegö, *Orthogonal Polynomials* (American Mathematical Society Colloquium Publications, Providence, Rhode Island), Vol. XXIII: (a) Paragraph 3.2 (Christoffel–Darboux formula); (b) Theorem 6.32 (location of largest zero). (c) Theorem 7.31.1 (relative max of $|\phi_n|$).

for the case of time-reversal-invariant physical systems. A formula corresponding to (2) has been given already by Mehta and Gaudin.[5] In the notation of this article the density of roots is given by

$$P_{n\ \text{real}}(Z/\sqrt{2}) = \sqrt{2}\sum_{i=1}^{n-1}\phi_i^2(Z) + (n-1)^{\frac{1}{2}}\phi_n(Z)\int_0^Z \phi_{n-1}(X)\,dX, \quad (26a)$$

$$n/\sqrt{2}\,P_{n\ \text{real}}(Z/\sqrt{2}) = (n-1)P_{n-1\ \text{Herm}}(Z) + [\tfrac{1}{2}(n-1)]^{\frac{1}{2}}\phi_n(Z)\int_0^Z \phi_{n-1}(X)\,dX. \quad (26b)$$

In the real symmetric case the radius of the semicircle is $n^{\frac{1}{2}}$. We therefore have for the expected number of roots in the tail

$$T_{n\ \text{real}} = \int_{(n)^{\frac{1}{2}}}^{\infty} nP_{n\ \text{real}}(Z')\,dZ'$$
$$= \int_{(2n)^{\frac{1}{2}}}^{\infty} nP_{n\ \text{real}}(Z/\sqrt{2})\,dZ/\sqrt{2},$$
$$= \int_{(2n)^{\frac{1}{2}}}^{\infty} (n-1)P_{n-1\ \text{Herm}}(Z)\,dZ + [\tfrac{1}{2}(n-1)]^{\frac{1}{2}} \times \int_{(2n)^{\frac{1}{2}}}^{\infty} dZ\phi_n(Z)\int_0^Z \phi_{n-1}(X)\,dX. \quad (27)$$

The first term on the right is $\lesssim 0.06$. Let us call the second term I, and change the order of integration to evaluate

$$I = \left[\int_0^{(2n)^{\frac{1}{2}}} \phi_{n-1}(X)\,dX\right]\left[\int_{(2n)^{\frac{1}{2}}}^{\infty} \phi_n(Z)\,dZ\right] + \int_{(2n)^{\frac{1}{2}}}^{\infty} dX\left[\phi_{n-1}(X)\int_x^{\infty} \phi_n(Z)\,dZ\right]. \quad (28)$$

For the second term, and the second integral of the first term, the asymptotic form (16) (its square root) is appropriate throughout the region of integration. This gives

$$I \approx 0.5(2n)^{-\frac{1}{2}}\int_0^{(2n)^{\frac{1}{2}}} \phi_{n-1}(X)\,dX + 0.12(2n)^{-\frac{1}{2}}. \quad (29)$$

It only remains to evaluate

$$J = \int_0^{(2n)^{\frac{1}{2}}} \phi_{n-1}(X)\,dX \approx \int_0^{\infty} \phi_n(X)\,dX. \quad (30)$$

Let us designate by $X_1 > X_2 > \cdots X_i > \cdots X_0$ the positive zeros of ϕ_{n+1}, X_1 being the largest. Well-known theorems tell us that the absolute value of successive extrema[4(c)] going out from zero along the positive real axis form an increasing sequence, and that the distance between zeros in the same interval[6(a)] increases in the same direction. By a suitable modification of a theorem of numerical comparison,[6(b)] the following can be established:

$$\left|\int_{X_i}^{X_{i+1}} \phi_n(X)\,dX\right| > \left|\int_{X_{i+1}}^{X_{i+2}} \phi_n(X)\,dX\right|. \quad (31)$$

Since ϕ_n is alternately positive and negative,

$$J < \int_0^{X_0} \phi_n(X)\,dX + \int_{X_1}^{\infty} \phi_n(X)\,dX < 2\int_{X_1}^{\infty} \phi_n(X)\,dX. \quad (32)$$

Using the asymptotic form for ϕ_n, and integrating from ϵ_i to infinity, we get $J \sim n^{-\frac{1}{2}}$. Therefore, the right-hand side of (27) is smaller than one and we have essentially the same result for real symmetric matrices as we had for Hermitian.

V. CONNECTION WITH PHYSICS

Various theoretical considerations,[7] as well as empirical evidence lead us to expect that the density of nuclear energy levels will obey an exponential law for a reasonably large range. We have seen above, however, that our study of n by n random matrices shows that the part of distribution which may show a similarity with the exponential is, in fact, completely negligible. No matter how large n is, as a rule, not even one root will be in the exponential part of the distribution. In the case of an n by n Hamiltonian which is invariant with respect to time reversal (real), Porter and Rosenzweig showed[8] that if one assumes only:

(a) The various matrix elements are statistically independent;

(b) The density function is invariant under real orthogonal transformations;

then one is *forced* to choose a density function of the form

$$D_n(H) = A\exp(-tr\,H^2/2a). \quad (33)$$

It is hard to imagine that the basis we choose for our ensemble should affect the eigenvalue distribution. The requirement at fault must therefore be (a). The next question will be, what sort of correlations among the elements of a general

[5] M. L. Mehta and M. Gaudin, Nucl. Phys. **18**, 420 (1960).

[6] F. Tricomi, *Differential Equations*, (Blackie & Son Ltd., London, 1961): (a) Sec. 21 (distance between successive zeros); (b) Sec. 20 (theorem of numerical comparison).

[7] D. ter Haar, *Elements of Statistical Mechanics* (Holt, Rinehart and Winston, Inc., New York, 1960), Chap. XIII.

[8] C. E. Porter and N. Rosenzweig, Ann. Acad. Sci. Finnicae Ser. A VI 44, (1960).

Hamiltonian should be assumed, and how should a density function with these correlation be chosen so as to satisfy (b). Since (33) does give us reasonable results with respect to local properties of the energy spectrum, (level spacings[9]) the most promising procedure would seem to be to look for a reasonable modification of that equation.

Let us look at (33) from another point of view. Consider an infinite-dimensional Hamiltonian matrix, with one row and one column labeled with a zero, the remainder numbered consecutively with the integers as they go out from zero. Let $(i \cup j)$ denote the maximum of i and j, where i designates the row of a matrix, and j the column. Let $f(i \cup j) = 1$ if $(i \cup j) \leq n$, and $f(i \cup j) = \infty$ otherwise. Then $D_n(H)$ is a product of terms of the form $\exp [-CH_{ij}^2 f(i \cup j)]$. Hence we have actually introduced a drastic correlation for elements with $(i \cup j) > n$, saying that the probability of such elements having finite value is identically zero. This has led to an almost equal abrupt cutoff in the eigenvalues at $n^{\frac{1}{2}}$. While this assumption may be appropriate in certain cases, where special knowledge of the dimension of the Hamiltonian is available, it is not valid in general. One possibility of avoiding such a drastic assumption would be to replace $P_n(\epsilon)$ by some suitably weighted average of $P_m(\epsilon)$ for $m \geq n$. This would have some analogy to the grand canonical ensemble of ordinary statistical mechanics.

ACKNOWLEDGMENTS

I would like to thank Professor E. P. Wigner for suggesting and motivating the problem, as well as for advice throughout. I would also like to thank Professor B. Bayman for helpful discussions. The calculation was started during the summer of 1962, during which time I received support from the United States Atomic Energy Commission.

APPENDIX

We want to establish the validity of using $\tilde{\phi}_\nu^{(2)}$, given by (12) as an approximation for $\tilde{\phi}_\nu$. We do this by showing that it is an upper bound for $\tilde{\phi}_\nu$ throughout the region of interest, i.e., to the right of ϵ_t as given by (13).

In Sec. III [e.g., (25)] it was shown that the outermost maximum of $\phi_\nu(\epsilon)$ is located at ϵ_t. If we write (5) as an equation for ϕ_n'', we see that except for zeros of ϕ_ν, the only inflection point of ϕ_ν, and thus the maximum of the slope (in this section, we are interested only in the absolute value of the slope) for our region, is at ϵ_t. From the same equation, we see that the magnitude of ϕ_n'' at a point to the right of ϵ_t is always somewhat less (since ϕ_n is monotonic there) than the magnitude of ϕ_n'' evaluated an equal distance to the left, as long as the distance is less than $(2n)^{-\frac{1}{2}}$. Therefore the slope to the left is less than that an equal distance to the right. Thus since ϕ_n never reaches zero on the right, its value at the maximum point on the left is less than twice its value at ϵ_t.

The asymptotic formula (11) is valid for $X \gg (2\nu)^{-\frac{1}{2}}$. For such an X, we can write the ratio

$$\tilde{\phi}_\nu(X)/\tilde{\phi}_\nu^{(2)}(X) = \frac{(2^{\frac{1}{2}}\pi)^{-\frac{1}{2}}(2\nu)^{-\frac{1}{4}}X^{-\frac{1}{4}} \exp(-\beta X^{\frac{3}{2}})}{C_1(2\nu)^{-1/12} \exp[-C_2(2\nu)^{\frac{1}{2}}X]}. \quad (34)$$

This ratio goes to zero rapidly for large X. [For $X > (2\nu)^{-\frac{1}{2}}$ it is already a decreasing function of ν.]

Now from (5) and (12), we can write the equations in a form suitable for comparison:

$$\phi_\nu''(\epsilon) + [(2\nu - 1) - \epsilon^2]\phi_\nu(\epsilon) = 0, \quad (35a)$$

$$\phi_\nu^{(2)\prime\prime}(\epsilon) - 0.8\epsilon_t^{\frac{1}{2}}\phi_\nu^{(2)}(\epsilon) = 0. \quad (35b)$$

Multiplying the first equation by $\phi_\nu^{(2)}$ and subtracting it from the second multiplied by ϕ_ν, we obtain after integrating

$$[\phi_\nu'(\epsilon)\phi_\nu^{(2)}(\epsilon) - \phi_\nu(\epsilon)\phi_\nu^{(2)\prime}(\epsilon)] = \int_\epsilon^\infty \{0.8\epsilon_t^{\frac{1}{2}} - [\epsilon^2 - (2\nu - 1)]\}\phi_\nu(\epsilon')\phi_\nu^{(2)}(\epsilon')\, d\epsilon', \quad (36)$$

since both terms on the left vanish at infinity. Rearranging and changing variables,

$$[\tilde{\phi}_\nu^{(2)}(X)]^2 \frac{d}{dX}\left[\frac{\tilde{\phi}_\nu(X)}{\tilde{\phi}_\nu^{(2)}(X)}\right] = \int_X^\infty [0.8\epsilon_t^{\frac{1}{2}} - (2\epsilon_t X' + X'^2)]\tilde{\phi}_\nu(X')\tilde{\phi}_\nu^{(2)}(X')\, dX'. \quad (37)$$

For X greater than $0.4(2\nu)^{-\frac{1}{2}}$, the integrand is always negative, therefore the ratio is decreasing to the right. Since the sign of the integrand can change only once for positive X, the two functions can be equal at most twice. If, proceeding left, the two functions are first equal at a point to the right of ϵ_t, the reason is because the slope of the ratio of the two functions can change sign only once, and we already know that the graphs cross at ϵ_t. If going left, the two functions are first equal at ϵ_t, they cross again at a point left of ϵ_t. In either case, once $\tilde{\phi}_\nu^{(2)}$ is greater than $\tilde{\phi}_\nu$ at or to the left of ϵ_t, it stays greater, because its slope (absolute value) is increasing while that of $\tilde{\phi}_\nu$ is decreasing.

[9] N. Rosenzweig and C. E. Porter, Phys. Rev. 120, 1698 (1960).

We have three possibilities:

(a) The graphs are tangent at ϵ_t;
(b) The graphs cross at ϵ_t, and one other point ϵ_c, to the left of ϵ_t;
(c) The graphs cross at ϵ_t, and one other point ϵ_c, to the right of ϵ_t.

In Cases (b) and (c), $\phi_\nu^{(2)}$ is greater than ϕ_ν, everywhere outside the region bounded by ϵ_c and ϵ_t. Consider Case (b). Let $\phi_\nu^{(2)} \to \alpha\phi_\nu^{(2)}$, where

$$1 \leq \alpha < 2.$$

Call the point to the left of b, where the slopes are equal, ϵ_s. As α is increased continuously, both $\phi_\nu^{(2)}$ and its slope (absolute value) increase everywhere. The point corresponding to ϵ_s moves right, while the crossing corresponding to ϵ_t moves left. In this manner, we fix the magnitude of α so that the two curves are tangent at some ϵ_0. Case (c) is handled identically. For clarity, we reexamine the situation to the right when we have tangency at ϵ_0 near ϵ_t. Now we have

$$\begin{aligned} \phi_\nu(\epsilon_0) &= \phi_\nu^{(2)}(\epsilon_0), \\ \phi_\nu'(\epsilon_0) &= \phi_\nu^{(2)\prime}(\epsilon_0). \end{aligned} \tag{38}$$

We can now in the same way we obtained (37) get

$$[\phi_\nu(\epsilon)]^2 \frac{d}{d\epsilon}\left[\frac{\phi_\nu^{(2)}(\epsilon)}{\phi_\nu(\epsilon)}\right]$$

$$= \int_{\epsilon_0}^{\epsilon} \{0.8\epsilon_t^{\frac{1}{2}} - [\epsilon^2 - (2\nu - 1)]\}\phi_\nu\phi_\nu^{(2)}\, d\epsilon. \tag{39}$$

The only place to the right of ϵ_0, where the integrand can become negative is at X equal to $0.4(2\nu)^{-\frac{1}{2}}$. But if $\phi_\nu^{(2)}$ once crossed or touched ϕ_ν to the right of that point, it could not become greater than ϕ_ν further to the right as required by (34), because the integrand does not change signs again. Hence $\phi_\nu^{(2)}$ is now an upper bound for ϕ_ν.

Theory of Vibrational Structure in Optical Spectra of Impurities in Solids. I. Singlets

D. E. McCUMBER
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received 21 June 1963)

General expressions are derived describing the modifications of the optical spectra of isolated impurities induced by the subclass of interactions with host-lattice phonons which shift the energy eigenvalues of the internal impurity state but which do not mix those internal states. The results apply to localized phonon modes as well as to extended modes (whose individual coupling decreases with increasing lattice volume), to anharmonic- as well as harmonic-phonon systems, and to phonon–impurity couplings which need not be linear in the dynamic local strain field. They take the form of "linked-cluster" expansions which, in the linear-coupling harmonic-phonon case, terminate after the second term. The relation of these expansions to moment methods is indicated. As a simple application of the linked-cluster expansions, we compute the dependence of the sharp "no-phonon" spectral line on impurity mass. We conclude that the no-phonon line should not exhibit isotopic-mass displacements in those impurity sites for which the lattice is inversion symmetric.

1. INTRODUCTION

A NUMBER of ions, notably those of the transition-metal and rare-earth sequences, exhibit one or more sharp optical lines when they are present substitutionally in host lattices. The gross features of such spectra can often be quantitatively interpreted by means of static-crystal-field theory. The detailed properties of the spectra—that is, the line shapes and the temperature dependence of the line positions—depend upon the dynamic interactions of the impurity with its environment. In this paper we present results relative to an important subclass of such environmental interactions, lattice-vibration (phonon) interactions which shift the energy eigenvalues of the internal states of an isolated impurity but which do not mix those internal states. In this introductory paper we implicitly consider an impurity system whose internal states are nondegenerate singlets. In the second paper of this series we shall discuss generalizations of our present results which apply to transitions between degenerate or nondegenerate impurity multiplets. Other features of impurity spectra will be discussed elsewhere.[1]

Special interactions of the type we consider have previously been treated in relation to trapped-electron spectra by many authors.[2–8] The Hamiltonians they considered had the basic structure

$$H = \varepsilon\psi^+\psi + \psi^+\psi \sum_q (C_q a_q^+ + C_q^* a_q) + \sum_q \hbar\omega_q (a_q^+ a_q + \tfrac{1}{2}), \quad (1.1)$$

where the $\{a_q, a_q^+\}$ are phonon annihilation–creation operators having the familiar commutation relations

$$[a_q, a_{q'}] = [a_q^+, a_{q'}^+] = 0, \quad [a_q, a_{q'}^+] = \delta(q, q'), \quad (1.2)$$

and where ψ, ψ^+ are similar operators relevant to a *single* internal "electronic" state of the isolated impurity having the Fermi–Dirac anticommutation relations

$$[\psi, \psi]_+ = [\psi^+, \psi^+]_+ = 0, \qquad [\psi, \psi^+]_+ = 1. \quad (1.3)$$

The phonon index q ranges over a complete set of lattice-vibration modes (not necessarily plane-wave modes). In this Hamiltonian, the phonon interaction shifts the energy level of the electronic state but does not mix that state with other electronic states.

Making the assumption that $|C_q|^2 \sim 1/V$ as the volume V of the lattice system increases, it is possible to demonstrate[2,6] that, to within corrections of order $1/V$,

$$f(t)_{N\{n_q\}} \equiv \langle N, \{n_q\} \,|[\psi(t), \psi^+]_+|\, N, \{n_q\}\rangle = \exp(-i\varepsilon t/\hbar) \exp(-\gamma) \exp\left\{\sum_q \frac{|C_q|^2}{(\hbar\omega_q)^2} \times [(1+n_q)\exp[-i(-1)^N\omega_q t] + n_q \exp[i(-1)^N\omega_q t]]\right\}, \quad (1.4a)$$

[1] D. E. McCumber, Bull. Am. Phys. Soc. **8**, 256 (1963).
[2] M. Lax, J. Chem. Phys. **20**, 1752 (1952).
[3] M. Lax and E. Burstein, Phys. Rev. **100**, 592 (1955).
[4] H. Gummel and M. Lax, Ann. Phys. (N. Y.) **2**, 28 (1957).
[5] B. S. Gourary and A. A. Maradudin, J. Phys. Chem. Solids **13**, 88 (1960).
[6] E. O. Kane, Phys. Rev. **119**, 40 (1960); W. E. Lamb, Jr., *ibid.* *55*, 190 (1939).
[7] J. J. Hopfield, *Proceedings of the International Conference on Semiconductor Physics, Exeter, 1962* (The Institute of Physics and the Physical Society, London, 1962), p. 75.
[8] R. C. O'Rourke, Phys. Rev. **91**, 265 (1953); S. Kiode, Z. Naturforsch. **15A**, 123 (1960).

where

$$\gamma = \sum_q |C_q|^2 (1 + 2n_q)/(\hbar\omega_q)^2,$$
$$\bar{\mathcal{E}} = \mathcal{E} - \sum_q |C_q|^2/(\hbar\omega_q), \tag{1.4b}$$

and where $\psi(t)$ is the Heisenberg operator $\psi(t) = \exp(iHt/\hbar)\psi \exp(-iHt/\hbar)$. The expectation value in (1.4) is taken with respect to $N\{n_q\}$ states for which $\psi^+\psi$ has the eigenvalue $N(= 0 \text{ or } 1)$ and the the set of operators $\{a_q^+ a_q\}$ the eigenvalues $\{n_q\}$.

The function $f(t)$ describes the spectral properties of the interacting impurity state ψ relative to some fixed energy reference. With minor modifications its Fourier transform is the optical absorption spectrum appropriate to the impurity-doped solid.[1,2,7]

In this paper we derive several important generalizations of (1.4). As a special application of our more general result, we verify that, if the $N\{n_q\}$ ensemble is replaced by a canonical $N\beta$ ensemble characterized by the quantum number N of $\psi^+\psi$ and by a phonon temperature $T = \hbar/k\beta$, then the result (1.4) is exact for *arbitrary* C_q provided only that n_q is replaced by

$$n_q = n(\omega_q) \equiv (e^{\beta\omega_q} - 1)^{-1}. \tag{1.5}$$

The importance of this result (previously reported by O'Rourke[8]) is that it extends Eqs. (1.4) to localized phonon modes for which the $|C_q|^2 \sim 1/V$ argument does not obtain. Such modes are of importance in real systems.

Equations (1.4) are valid only if the system Hamiltonian has the rather idealized form (1.1). We have derived a generalization of (1.4) appropriate to the Hamiltonian

$$H = H_e + H_p + \psi^+\psi A, \tag{1.6}$$

where H_e is any "electronic" Hamiltonian diagonal in the electronic occupation operator $\psi^+\psi$ (and independent of $\{a_q, a_q^+\}$), where H_p is any phonon Hamiltonian (not necessarily quadratic in $\{a_q, a_q^+\}$ but independent of $\psi^+\psi$), and where A is any Hermitian function of the phonon operators $\{a_q, a_q^+\}$.[9] As before, the phonon interaction shifts the energy level of the impurity state appropriate to the operators ψ, ψ^+, but does not mix that state with any other impurity states. An effective Hamiltonian of this structure can often be used to describe the lattice interactions of those impurity levels which are well separated (relative to phonon energies) from other impurity levels. In addition, those components of a more general Hamiltonian which have the basic structure of the Hamiltonian (1.6) can profitably be treated separately from those components which mix electronic impurity states. We shall discuss this feature in detail in a subsequent paper.

Using the Hamiltonian (1.6), we establish in Sec. 5 that

$$f(t)_{N\beta} \equiv \langle N\beta\,|[\psi(t),\,\psi^+]_+|\,N\beta\rangle = f_e(t)_{N\beta} g(t)_{N\beta}, \tag{1.7}$$

where

$$f_e(t)_{N\beta} = \langle N\beta\,|[\psi(t)^e,\,\psi^+]_+|\,N\beta\rangle$$
$$= \langle N\beta\,|[e^{iH_e t/\hbar}\psi e^{-iH_e t/\hbar},\,\psi^+]_+|\,N\beta\rangle, \tag{1.8}$$

and

$$\begin{aligned} g(t)_{N\beta} &= 1 - \left(\frac{i}{\hbar}\right)\int_0^t dt_1 \langle N\beta\,|A(t_1)|\,N\beta\rangle \\ &+ \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2 \int_0^t dt_1\,dt_2 \\ &\times \langle N\beta\,|(A(t_1)A(t_2))_N|\,N\beta\rangle \\ &- \frac{1}{3!}\left(\frac{i}{\hbar}\right)^3 \int_0^t dt_1 \cdots dt_3 \\ &\times \langle N\beta\,|(A(t_1)A(t_2)A(t_3))_N|\,N\beta\rangle + \cdots \qquad (1.9a) \\ &= \exp\left\{-\frac{i}{\hbar}\int_0^t dt_1 \langle N\beta\,|A(t_1)|\,N\beta\rangle\right. \\ &+ \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2 \int_0^t dt_1\,dt_2 [\langle N\beta\,|(A(t_1)A(t_2))_N|\,N\beta\rangle \\ &- \langle N\beta\,|A(t_1)|\,N\beta\rangle\langle N\beta\,|A(t_2)|\,N\beta\rangle] \\ &- \frac{1}{3!}\left(\frac{i}{\hbar}\right)^3 \int_0^t dt_1 \cdots dt_3 \\ &\times [\langle N\beta\,|(A(t_1)A(t_2)A(t_3))_N|\,N\beta\rangle \\ &- \langle N\beta\,|A(t_1)|\,N\beta\rangle\langle N\beta\,|(A(t_2)A(t_3))_N|\,N\beta\rangle \\ &- \langle N\beta\,|(A t_2)|\,N\beta\rangle\langle N\beta\,|(A(t_3)A(t_1))_N|\,N\beta\rangle \\ &- \langle N\beta\,|A(t_3)|\,N\beta\rangle\langle N\beta\,|(A(t_1)A(t_2))_N|\,N\beta\rangle \\ &+ 2\langle N\beta\,|A(t_1)|\,N\beta\rangle\langle N\beta\,|A(t_2)|\,N\beta\rangle \\ &\times \left.\langle N\beta\,|A(t_3)|\,N\beta\rangle] + \cdots\right\}. \qquad (1.9b) \end{aligned}$$

Here $A(t)$ is the Heisenberg operator $A(t) = \exp(iHt/\hbar)A \exp(-iHt/\hbar)$. In Eqs. (1.9) and below, $\langle N\beta| \cdots |N\beta\rangle$ designates the expectation value taken with respect to the $N\beta$ ensemble (indicated in greater detail in Sec. 2). Also, the abbreviation $(\cdots)_N$ indicates that the operators contained within the parentheses are to be time

[9] R. C. O'Rourke (Ref. 8) also computed lowest-order corrections for the case when the phonon frequencies are functions of $\psi^+\psi$.

ordered with times closest to t (greater times for $t > 0$, lesser times for $t < 0$), to the left when $N = 0$, and to the right when $N = 1$. [For example, if $t_1 > t_2$ for $t > 0$, then $(A(t_1)A(t_2))_N = A(t_1)A(t_2)$ when $N = 0$, and $A(t_2)A(t_1)$ when $N = 1$. If $t_1 < t_2$, the situation is reversed.]

The expansion (1.9b) is a type of "linked-cluster" or Baker–Hausdorff expansion.[10] Alternatively, if we view Eq. (1.9a) as a type of generalized moment expansion, then Eq. (1.9b) is the corresponding generalized cumulant (or semi-invariant) expansion. Such expansions and many of their formal properties have been discussed by Kubo.[11]

The usefulness of the expansion (1.9b) stems from the fact that we expect the first few terms to dominate the series. The nth-order term of the exponent series involves what are intrinsically nth-order correlations. In fact, if A is linear in $\{a_q, a_q^+\}$, as in (1.1) where

$$A = \sum_q (C_q a_q^+ + C_q^* a_q), \tag{1.10}$$

and if the phonon system is harmonic—that is, if H_p is at most bilinear in the $\{a_q, a_q^+\}$—then only the first two terms in the exponent (1.9b) are different from zero. The intrinsic third- and higher-order correlations of the operator A in the harmonic-phonon ensemble vanish. Equations (1.4) result from this harmonic case if in addition $\langle 0\beta|\ A\ |0\beta\rangle = 0$ and $f_e(t)_{N\beta} = \exp(-i\mathcal{E}t/\hbar)$. If A is not linear in the $\{a_q, a_q^+\}$, or if H_p is not harmonic, the higher-order terms of the expansion (1.9b) are not strictly zero. Nevertheless it is still frequently useful to retain only a few terms of the correlation expansion (1.9b).

In Sec. 2 we briefly discuss properties of the optical absorption and emission spectra, noting especially those properties which obtain when the series in (1.9b) is approximated by its first two terms. By rearranging the components of those two terms and by expanding one exponential in powers of its argument, we can identify the "no-phonon" and "vibrational structure" spectral components. The absorption and emission spectra appropriate to initially thermalized phonon ensembles of temperature T differ only by a factor of the type exp $[\beta(\omega - \mu)]$, where $\beta = \hbar/kT$, and μ is a temperature-dependent "chemical potential."

In Sec. 3 we use Eq. (1.9a) to derive sum rules or moment relations.[2-5] These are particularly useful in those strong-interaction cases for which it is not feasible to use Eq. (1.9b).[6]

Restricting the general Hamiltonian (1.6) to be the simple Hamiltonian (1.1), we verify in Sec. 4 that Eqs. (1.4) follow from Eqs. (1.9). In Sec. 5 we prove Eqs. (1.9).

Using results previously derived, we compute in Sec. 6 the shift in the position of the no-phonon line which arises because the impurity mass is different from the mass of the lattice ion it replaces. For simplicity we restrict the unperturbed Hamiltonian to the special harmonic Hamiltonian (1.1) with the linear coupling (1.10). The mass difference generates a kinetic-energy perturbation of the phonon spectrum which will be reflected in the position of the no-phonon line (as well as in detailed features of the vibrational structure, which are more difficult to compute). Experimentally the mass-difference effects can most easily be determined by comparing spectra from different impurity isotopes. It is an important result that for a Hamiltonian of the type (1.1) the position of the no-phonon line is *not* dependent upon the isotopic mass of the impurity when the impurity site is a center of inversion symmetry of the lattice.

2. EMISSION AND ABSORPTION SPECTRA

If we indicate the structure of the $N\beta$ ensemble more explicitly than in the first of Eqs. (1.9), then[12]

$$f(t)_{N\beta} = \mathrm{tr}_N\,(e^{-\beta H/\hbar}[\psi(t),\ \psi^+]_+)/\mathrm{tr}_N\,(e^{-\beta H/\hbar}), \tag{2.1}$$

where tr_N indicates the trace over all states for which the eigenvalue of $\psi^+\psi$ is $N = 0, 1$. Relative to the absorption and emission spectra of isolated impurities, we note that

$$f(t)_{0\beta} = \mathrm{tr}_0\,(e^{-\beta H/\hbar}e^{iHt/\hbar}\psi e^{-iHt/\hbar}\psi^+)/\mathrm{tr}_0\,(e^{-\beta H/\hbar}) \tag{2.2}$$

describes the (absorption) spectrum appropriate to the excitation of the impurity state $N = 1$ from the initial impurity reference (ground) state $N = 0$.[2] [For simplicity we assume in this paper that the initial reference state does not interact dynamically with phonons.] The time dependence of this function reflects the weighted energy differences between initial and final states; more precisely, its Fourier transform is the observed spectrum.

In (2.2) the initial phonon states are distributed according to the thermal weight factor exp $(-\beta H/\hbar)$; the final phonon states are not *a priori* restricted. The function

[10] G. H. Weiss and A. A. Maradudin, J. Math. Phys. **3**, 771 (1962); R. Englman and T. Levy, *ibid.* **4**, 105 (1963).
[11] R. Kubo, J. Phys. Soc. Japan **17**, 1100 (1962).
[12] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959), and references cited therein.

$$f(t)_{1\beta} = \mathrm{tr}_1\,(e^{-\beta H/\hbar}e^{-iHt/\hbar}\psi^+e^{iHt/\hbar}\psi)/\mathrm{tr}_1\,(e^{-\beta H/\hbar}) \quad (2.3)$$

describes the spectral properties of the inverse (emission) process in which the impurity goes from the state $N = 1$ to that with $N = 0$. Again the phonons are initially ($N = 1$) distributed as in a thermal ensemble. It is a trivial consequence of the cyclic property tr (AB) = tr (BA) of the trace that the two functions (2.2) and (2.3), appropriate to initially thermalized phonon ensembles, are related by

$$\begin{aligned} f(t)_{1\beta} &= f(t - i\beta)_{0\beta}/f(-i\beta)_{0\beta}, \\ f(t)_{0\beta} &= f(t + i\beta)_{1\beta}/f(i\beta)_{1\beta}. \end{aligned} \quad (2.4)$$

Each function $f(t)_{N\beta}$, $N = 0$ and 1, thus has relevance to both absorption and emission. Equations (2.4) also apply separately to the functions $f_s(t)_{N\beta}$.

If we define Fourier transforms $s(\omega)_{N\beta}$, $s_s(\omega)_{N\beta}$, and $p(\omega)_{N\beta}$ such that

$$\begin{aligned} f(t)_{N\beta} &= \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,s(\omega)_{N\beta}e^{-i\omega t}, \\ f_s(t)_{N\beta} &= \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,s_s(\omega)_{N\beta}e^{-i\omega t}, \end{aligned} \quad (2.5)$$

and

$$g(t)_{N\beta} = \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,p(\omega)_{N\beta}e^{-i\omega t}, \quad (2.6)$$

then Eq. (1.7) takes the form

$$s(\omega)_{N\beta} = \int_{-\infty}^{\infty}\frac{d\bar{\omega}}{2\pi}\,s_s(\omega - \bar{\omega})_{N\beta}p(\bar{\omega})_{N\beta}. \quad (2.7)$$

In terms of these Fourier transforms, Eqs. (2.4) take the form

$$\begin{aligned} s(\omega)_{0\beta} = e^{\beta(\omega-\mu)}s(\omega)_{1\beta}, \quad s_s(\omega)_{0\beta} &= e^{\beta(\omega-\mu)}s_s(\omega)_{1\beta}, \\ p(\omega)_{0\beta} &= e^{\beta(\omega-\mu)}p(\omega)_{1\beta}, \end{aligned} \quad (2.8)$$

where the temperature-dependent "chemical potential" μ is defined such that

$$\exp(\beta\mu) = \mathrm{tr}_0[\exp(-\beta H/\hbar)]/\mathrm{tr}_1[\exp(-\beta H/\hbar)].$$

From the fact that $g(t = 0) = 1$, we have alternatively

$$e^{\beta\mu} = \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,e^{\beta\omega}p(\omega)_{1\beta} = \left[\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,e^{-\beta\omega}p(\omega)_{0\beta}\right]^{-1}. \quad (2.9)$$

Equations (2.8) imply that Eqs. (2.4) apply separately to the functions $g(t)_{N\beta}$.

For a harmonic crystal in which A has the linear form (1.10), the only nonvanishing components of the exponential in Eq. (1.9b) are, as we indicate in Sec. 4, the first two terms

$$\begin{aligned} &-\frac{i}{\hbar}\int_0^t dt_1\langle N\beta\,|A(t_1)|\,N\beta\rangle + \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2 \\ &\quad\times\int_0^t dt_1\,dt_2[\langle N\beta\,|(A(t_1)A(t_2))_N|\,N\beta\rangle \\ &\quad - \langle N\beta\,|A(t_1)|\,N\beta\rangle\langle N\beta\,|A(t_2)|\,N\beta\rangle] \\ &= -it\langle N\beta\,|A|\,N\beta\rangle/\hbar + \int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\rho_{2N}(\omega)n(\omega) \\ &\quad\times\frac{\exp[i(-1)^N\omega t] - 1 - i(-1)^N\omega t}{(\hbar\omega)^2}. \end{aligned} \quad (2.10)$$

Here $n(\omega)$ is defined in (1.5), and

$$\rho_{2N}(\omega) = \int_{-\infty}^{\infty}dt\,e^{i\omega t}\langle N\beta\,|[A(t),\,A]|\,N\beta\rangle \quad (2.11a)$$

$$= -\rho_{2N}(-\omega) = \rho_{2N}(\omega)^*. \quad (2.11b)$$

The $\rho_{2N}(\omega)$ spectral representation in Eq. (2.10) is valid for the thermal phonon ensemble and has been extensively discussed in the literature.[12] The proof of its validity parallels that of Eq. (2.4). The two terms (2.10) are often an excellent approximation to the complete exponent in (1.9b) even when A contains small nonlinear terms and H_p contains small anharmonic components, in which case the higher-order terms are not strictly zero.

In the integrand of the last term in (2.10), the $\omega = 0$ singularity in $n(\omega)$ is canceled by the $\rho_{2N}(\omega)$ zero ensured by the symmetry relations (2.11b). Since the integrand is therefore regular at $\omega = 0$, we may replace the ω^2 denominator by $(\omega \pm i\epsilon)^2$, $\epsilon = 0^+$, without thereby changing the value of the integral. We choose the sign of $i\epsilon$ so that the exp $(i\omega t)$ component of (2.10) will not contribute in the limit $|t| \to \infty$, and so that all terms of order unity and of order t will derive respectively from the integrand numerator components $(1 \pm i\omega t)$. In order that this be true, the $\omega = 0$ pole present in the *separate* terms of the integrand (but not in the integrand as a whole) must not contribute to the exp $(i\omega t)$ integral when for $t \neq 0$ the contour is closed at infinity about a suitable half plane. This will be true if we rewrite (2.10) in the form

$$\begin{aligned} &-it\langle N\beta\,|A|\,N\beta\rangle/\hbar + \frac{1}{\hbar^2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\rho_{2N}(\omega)n(\omega) \\ &\quad\times\frac{\exp[i(-1)^N\omega t] - 1 - i(-1)^N\omega t}{(\omega + i\epsilon(-1)^N\,\mathrm{sgn}\,t)^2} \\ &= -\gamma_{Nt} - it\Delta\mathcal{E}_{Nt}/\hbar + \frac{1}{\hbar^2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi} \\ &\quad\times\frac{\rho_{2N}(\omega)n(\omega)\exp[i(-1)^N\omega t]}{(\omega + i\epsilon(-1)^N\,\mathrm{sgn}\,t)^2}, \end{aligned} \quad (2.12a)$$

where, for $\epsilon = 0^+$,

$$\gamma_{Nt} = \frac{1}{\hbar^2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{\rho_{2N}(\omega)n(\omega)}{(\omega + i\epsilon(-1)^N \operatorname{sgn} t)^2}, \qquad (2.12b)$$

and

$$\Delta\mathcal{E}_{Nt} = \Delta\mathcal{E}_N - i\tfrac{1}{2}\hbar\Gamma_N \operatorname{sgn} t$$
$$= \left[\langle N\beta\,|A|\,N\beta\rangle - (-1)^N\int_0^{\infty}\frac{d\omega}{2\pi}\frac{\rho_{2N}(\omega)}{\hbar\omega}\right]$$
$$- \frac{i}{2\beta\hbar}\frac{\partial\rho_{2N}(\omega)}{\partial\omega}\bigg|_{\omega=0}\operatorname{sgn} t. \qquad (2.12c)$$

With respect to Eqs. (2.4), we define $\operatorname{sgn}(t \pm i\beta) = \operatorname{sgn} t$.

If A has the linear form (1.10), phase–space considerations generally ensure that in three-dimensional lattices $\rho_{2N}(\omega)$ vanishes at least as fast as ω^3 as $\omega \to 0$. In that case, we may set $\epsilon = 0$ and ignore the sgn t parts of (2.12). For more general A operators, such components cannot be neglected. If $\Gamma_N \neq 0$, we note in particular that the energy shift $\Delta\mathcal{E}_{Nt}$ in (2.12c) is complex. The sign of the imaginary part of $\Delta\mathcal{E}_{Nt}$ will always be such as to damp the time response when the results of (2.12) are used in (1.9b):

$$g(t)_{N\beta}|_{(2.12)} = \exp\{-\gamma_{Nt} - it\Delta\mathcal{E}_N/\hbar - \tfrac{1}{2}\Gamma_N|t|\}$$
$$\times \exp\left\{\frac{1}{\hbar^2}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{\rho_{2N}(\omega)n(\omega)\exp[i(-1)^N\omega t]}{(\omega + i(-1)^N\epsilon \operatorname{sgn} t)^2}\right\}. \qquad (2.13)$$

The degree to which equations of this type are consistent with Eqs. (2.4) or (2.8) is a measure of the validity of approximations based upon truncations of the series (1.9b). For example, we find after some manipulations involving the definition (1.5) of $n(\omega)$ and the properties (2.11b) of $\rho_{2N}(\omega)$, that Eqs. (2.8) and (2.13) are completely self-consistent only if

$$\rho_{20}(\omega) = \rho_{21}(\omega) \equiv \rho_2(\omega), \qquad (2.14a)$$

and

$$\langle 1\beta\,|A|\,1\beta\rangle - \langle 0\beta\,|A|\,0\beta\rangle = -2\int_0^{\infty}\frac{d\omega}{2\pi}\frac{\rho_2(\omega)}{\hbar\omega}. \qquad (2.14b)$$

These conditions do obtain in the linear-A harmonic-H_p case for which Eqs. (2.13) are exact. Notice that they do *not* impose any condition on the no-phonon width Γ_N of Eq. (2.12).

Following Kane,[6] we may rewrite the last exponential in Eq. (2.13) in the form $\exp(\cdots) = 1 + [\exp(\cdots) - 1]$. The first term generates a sharp "no-phonon" component; the second generates what is commonly known as "vibrational structure." This separation is particularly appropriate when the phonon-impurity coupling is not large [$\exp(-\gamma) \gtrsim 0.1$]. In that case it is sometimes useful to continue the expansion of the exponential in powers of its argument. For the linear A operator (1.10), we would identify the nth-order terms in that expansion as n-phonon vibrational structure. For more general A operators, the n-phonon identification is less direct.

When Eqs. (2.14) obtain, the vibrational structure predicted by Eqs. (2.13) in absorption is the mirror image about the no-phonon line of that in emission. A typical case for which Eqs. (2.14) are not exactly met and for which the emission and absorption spectra are not mirror images about the no-phonon line is that in which the phonon frequencies ω_q are functions of $\psi^+\psi$ or, equivalently, the operator A is bilinear in the $\{a_q, a_q^+\}$. If such effects are small, the order-A^3 corrections to Eqs. (2.13) may often be inconsequential.

3. MOMENT RELATIONS (SUM RULES)

While it is often convenient and satsifactory to analyze the spectral function $g(t)_{N\beta}$ by expanding the last exponential in (2.13) in powers of its argument, this is feasible without machine computation only if the phonon-impurity interaction is relatively weak, as it is for example in the R-line spectrum of ruby.[13] In F-center spectra and in the green and violet broadband components of the ruby spectrum,[14] the interaction is not weak and the exponential expansion does not converge rapidly. In such cases one can usefully characterize the optical spectrum by means of its moments or semi-invariants, as Lax has noted.[2-4] The relevant expressions are easily derived from Eq. (1.9) and the definition (2.6).

The first few moments of the function $p(\omega)_{N\beta}$ are

$$\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}p(\omega)_{N\beta} = g(0)_{N\beta} = 1, \qquad (3.1a)$$

$$\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\omega p(\omega)_{N\beta} = i\frac{\partial}{\partial t}g(t)_{N\beta}\bigg|_{t=0} = \langle N\beta\,|(A/\hbar)|\,N\beta\rangle, \qquad (3.1b)$$

and

$$\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\omega^2 p(\omega)_{N\beta} = i^2\frac{\partial^2}{\partial t^2}g(t)_{N\beta}\bigg|_{t=0} = \langle N\beta\,|(A/\hbar)^2|\,N\beta\rangle. \qquad (3.1c)$$

Equation (2.12c) gives a related result for $1/\omega$.

[13] G. Burns and M. I. Nathan, J. Appl. Phys. **34**, 703 (1963).
[14] D. S. McClure, J. Chem. Phys. **36**, 2757 (1962).

The semi-invariants χ_m of the spectral function $p(\omega)_{N\beta}$ are defined by the expression

$$g(t) = \exp\left\{\sum_{m=1}^{\infty} \frac{\chi_m}{m!}(-it)^m\right\}. \tag{3.2}$$

For the special case of the harmonic Hamiltonian (1.1) with A of the linear form (1.10), we can easily establish from Eqs. (1.4) that for integral $m \geq 1$,

$$\chi_0 = \chi_1 = 0, \tag{3.3a}$$

$$\chi_{2m} = \sum_q |C_q|^2 \omega_q^{2m-2}[2n(\omega_q) + 1]/\hbar^2, \tag{3.3b}$$

$$\chi_{2m+1} = \sum_q |C_q|^2 \omega_q^{2m-1}/\hbar^2. \tag{3.3c}$$

The ratios $r_m \equiv \chi_m/(\chi_2)^{\frac{1}{2}m}$ for $m \geq 3$ are measures of the deviation of $p(\omega)_{N\beta}$ from a Gaussian line shape, for which $r_m = 0$ if $m \geq 3$. [The ratios r_3 and r_4 are the coefficients of skewness and of excess, respectively, for the spectral distribution $p(\omega)_{N\beta}$.]

The first-moment expression (3.1b) is particularly relevant to static-crystal-field analyses of phonon-broadened absorption or emission bands.[15,16] It implies that the unperturbed crystal-field level is to be identified with the *mean* frequency of the absorption band (less $\langle 0\beta|\ (A/\hbar)\ |0\beta\rangle$) and not with the frequency of the remnant "no-phonon" line located near the low-frequency edge of the band. According to Eq. (2.12c) the no-phonon line is displaced a distance

$$\Delta\mathcal{E}_0/\hbar = \langle 0\beta\ |(A/\hbar)|\ 0\beta\rangle - \int_0^{\infty} \frac{d\omega}{2\pi}\, \rho_{20}(\omega)/(\hbar^2\omega) \tag{3.4}$$

from the unperturbed frequency appearing in $f_e(t)_{0\beta}$, and a distance $(\Delta\mathcal{E}_0 - \langle 0\beta|\ A\ |0\beta\rangle)/_N$ from the mean band frequency.

Moment methods do not adequately reflect fine structure in the vibrational structure. When such structure is important and the phonon-impurity coupling is strong,[17] a machine computation based on an approximate expression of the type (2.13) is most satisfactory. Hopfield has given one spectral equation based on Eq. (1.4) which is suitable for such a computation.[7]

4. THE HARMONIC HAMILTONIAN (1.1)

To verify that Eqs. (1.7)–(1.9) reduce to Eqs. (1.4) when the Hamiltonian (1.6) has the special form (1.1), we find it convenient to introduce in place of the fields $\{a_q,\ a_q^+\}$ in (1.1), new fields $\{\bar{a}_q,\ \bar{a}_q^+\}$ defined by the canonical transformation

$$\bar{a}_q = a_q - NC_q/\hbar\omega_q, \quad \bar{a}_q^+ = a_q^+ - NC_q^*/\hbar\omega_q, \tag{4.1}$$

where N is a constant. This transformation preserves the commutation relations (1.2). Relative to states for which the eigenvalue of $\psi^+\psi$ is N, the Hamiltonian (1.1) becomes

$$H_N = (\mathcal{E}N - N^2 \sum_q |C_q|^2/\hbar\omega_q) + \sum_q \hbar\omega_q(\bar{a}_q^+\bar{a}_q + \tfrac{1}{2}). \tag{4.2}$$

With respect to the $N\beta$ ensemble the operators $\{\bar{a}_q,\ \bar{a}_q^+\}$ have the property[18]

$$\langle N\beta\ |(\bar{a}_{q_1}^+)^{m_1}(\bar{a}_{q_2}^+)^{m_2}\cdots(\bar{a}_{q_i})^{\bar{m}_2}(\bar{a}_{q_1})^{\bar{m}_1}|\ N\beta\rangle = \prod_i \{\delta(m_i, \bar{m}_i)m_i!\,(n(\omega_{q_i}))^{m_i}\}, \tag{4.3}$$

where $n(\omega)$ is defined in (1.5) and where $\{m_i\}$ and $\{\bar{m}_i\}$ are two sets of nonnegative integers. In the Heisenberg picture, when the Hamiltonian (4.2) governs the dynamics of the $\{a_q,\ a_q^+\}$ fields,

$$A(t) = \sum_q (C_q a_q^+ + C_q^* a_q)(t) = \sum_q (C_q\bar{a}_q^+ e^{i\omega_q t} + C_q^*\bar{a}_q e^{-i\omega_q t}) + 2N\sum_q |C_q|^2/\hbar\omega_q, \tag{4.4a}$$

$$\equiv \bar{A}(t) + 2N\sum_q |C_q|^2/\hbar\omega_q. \tag{4.4b}$$

Using these results in the right-hand side of Eq. (1.9a), we find that

$$g(t)_{N\beta} = \left\{1 - \frac{i}{\hbar}\int_0^t dt_1 2N\sum_q |C_q|^2/\hbar\omega + \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2\int_0^t dt_1\,dt_2 \times (2N\sum_q |C_q|^2/\hbar\omega_q)^2 + \cdots\right\} \times \left\{1 + \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2\int_0^t dt_1\,dt_2 \times \langle N\beta\ |(\bar{A}_1\bar{A}_2)_N|\ N\beta\rangle + \cdots\right\}, \tag{4.5a}$$

[15] S. Sugano and Y. Tanabe, J. Phys. Soc. Japan **13**, 880 (1958).
[16] S. Sugano and M. Peter, Phys. Rev. **122**, 381 (1961).
[17] A notable example is the exciton spectrum reported by R. E. Dietz, D. G. Thomas, and J. J. Hopfield, Phys. Rev. Letters **8**, 391 (1962).
[18] This may be verified by direct computation using $\langle N\beta\ |\theta|\ N\beta\rangle = \mathrm{tr}\ [\theta\ \exp\ (-\beta H_N/\hbar)]/\mathrm{tr}\ [\exp\ (-\beta H_N/\hbar)]$. Alternatively, see Eq. (6.7) of Martin and Schwinger, Ref. 10, which more directly applies to our Eq. (4.5).

$$= \{\exp(-it2N\sum_q |C_q|^2/\hbar^2\omega_q)\}$$
$$\times\Big\{1+\frac{1}{2!}\Big(\frac{i}{\hbar}\Big)^2\int_0^t dt_1\,dt_2\langle N\beta\,|(\bar{A}_1\bar{A}_2)_N|\,N\beta\rangle$$
$$+\frac{1}{4!}\Big(\frac{i}{\hbar}\Big)^4\int_0^t dt_1\cdots dt_4[\langle N\beta\,|(\bar{A}_1\bar{A}_2)_N|\,N\beta\rangle$$
$$\times\langle N\beta\,|(\bar{A}_3\bar{A}_4)_N|\,N\beta\rangle+\langle N\beta\,|(\bar{A}_1\bar{A}_3)_N|\,N\beta\rangle$$
$$\times\langle N\beta\,|(A_2A_4)_N|\,N\beta\rangle+\langle N\beta\,|(\bar{A}_1\bar{A}_4)_N|\,N\beta\rangle$$
$$\times\langle N\beta\,|(\bar{A}_2\bar{A}_3)_N|\,N\beta\rangle]+\cdots\Big\}, \tag{4.5b}$$
$$= \{\exp(-it2N\sum_q |C_q|^2/\hbar^2\omega_q)\}$$
$$\times\Big\{1-\frac{1}{2\hbar^2}\int_0^t dt_1\,dt_2\langle N\beta\,|(\bar{A}_1\bar{A}_2)_N|\,N\beta\rangle$$
$$+\frac{1}{2!}\Big[\frac{1}{2\hbar^2}\int_0^t dt_1\,dt_2$$
$$\times\langle N\beta\,|(\bar{A}_1\bar{A}_2)_N|\,N\beta\rangle\Big]^2-\cdots\Big\}, \tag{4.5c}$$
$$= \exp(-it2N\sum_q |C_q|^2/\hbar^2\omega_q)\exp\Big[-\frac{1}{2\hbar^2}$$
$$\times\int_0^t dt_1\,dt_2\langle N\beta\,|(\bar{A}_1\bar{A}_2)_N|\,N\beta\rangle\Big], \tag{4.5d}$$

where for notational simplicity we have written $\bar{A}_i$ for $\bar{A}(t_i)$. In (4.5a) we used the property (4.3) to eliminate from the second factor the expectation values of all odd powers of $\bar{A}$. In (4.5b) we again used (4.3) to factor the fourth- and higher-order correlation functions. That such functions factor into products of second-order functions is characteristic of the linear operator A for a harmonic-phonon system. Equation (4.5c) obtains from (4.5b) by interchange of dummy integration variables.

From Eq. (4.3) and Eq. (4.4), it follows that

$$\langle N\beta\,|\bar{A}(t_1)\bar{A}(t_2)|\,N\beta\rangle = \sum_q |C_q|^2\,[n(\omega_q)e^{i\omega_q(t_1-t_2)} + (1+n(\omega_q))e^{-i\omega_q(t_1-t_2)}] \tag{4.6}$$

or

$$\int_0^t dt_1\int_0^{t_1} dt_2\langle N\beta\,|\bar{A}(t_1)\bar{A}(t_2)|\,N\beta\rangle = \sum_q\frac{|C_q|^2}{\omega_q^2}\times\{n(\omega_q)e^{i\omega_q t}+(1+n(\omega_q))e^{-i\omega_q t}+i\omega_q t-(1+2n(\omega_q))\}. \tag{4.7}$$

Substituting this into Eq. (4.5d) and using the result in turn with Eq. (1.7), we have Eqs. (1.4).

Equations (1.4) could have been verified by an extension of the method used by Kane,[6] although the algebraic manipulations are very involved. As an illustration of what is encountered, we indicate the analogue of (4.5) for the factor $\exp(-\gamma)$ in the no-phonon component of (1.4). If $n_q = \bar{a}_q^+\bar{a}_q$, that factor would appear in the form

$$\Big\langle N\beta\,|\prod_q\Big\{1-\frac{|C_q|^2}{(\hbar\omega_q)^2}(2n_q+1)+\frac{1}{2!}\Big[\frac{|C_q|^2}{(\hbar\omega_q)^2}\Big]^2\times[2n_q(n_q-1)+4n_q+1]+\cdots\Big\}\Big|\,N\beta\Big\rangle$$
$$= \prod_q\Big\{1-\frac{|C_q|^2}{(\hbar\omega_q)^2}(2\bar{n}_q+1)+\frac{1}{2!}\Big[\frac{|C_q|^2}{(\hbar\omega_q)^2}(2\bar{n}_q+1)\Big]^2+\cdots\Big\}$$
$$= \exp\{-\sum_q |C_q|^2(2\bar{n}_q+1)/(\hbar\omega_q)^2\}, \tag{4.8}$$

where $\bar{n}_q$ is defined in (1.5). To obtain the second of Eqs. (4.8), we were required to use the property (4.3). The factors $m!$ in (4.3) played essential roles. In contrast, Kane used the argument that

$$\prod_q\Big\{1-\frac{|C_q|^2}{(\hbar\omega_q)^2}(2n_q+1)+\cdots\Big\} = \exp\{-\sum_q |C_q|^2(2n_q+1)/(\hbar\omega_q)^2+O(VC_q^4)\} \underset{V\to\infty}{\to} \exp\{-\sum_q |C_q|^2(2n_q+1)/(\hbar\omega_q)^2\} \tag{4.9}$$

if $|C_q|^2 \sim 1/V$ for V large. Our result does not depend for its validity upon this property of the coefficients C_q; it therefore obtains even for localized phonon modes for which the approximation (4.9) is invalid.[8]

5. PROOF OF THE EXPANSIONS (1.9)

It remains to verify the fundamental expansions (1.9). For this purpose we introduce the "unperturbed" Hamiltonian

$$H_0 = H_e + H_p, \tag{5.1}$$

whose components appear in the Hamiltonian H of (1.6). If $\mathcal{O}$ is an arbitrary operator, we introduce time-dependent Heisenberg operators according to the following conventions:

$$\mathcal{O}(t) = e^{iHt/\hbar}\mathcal{O}e^{-iHt/\hbar};\quad \mathcal{O}(t)^0 = e^{iH_0t/\hbar}\mathcal{O}e^{-iH_0t/\hbar};\quad \{\mathcal{O}(t)^0\}(t') = e^{iHt'/\hbar}e^{iH_0t/\hbar}\mathcal{O}e^{-iH_0t/\hbar}e^{-iHt'/\hbar}; \tag{5.2}$$

etc. If $H - H_0 = \psi^+\psi A$, as follows from (1.6) and (5.1), it is easy to verify by differentiation that

$$\mathcal{O}(t) = \mathcal{O}(t)^0 + \frac{i}{\hbar}\int_0^t dt'\{[A\psi^+\psi, \mathcal{O}(t-t')^0]\}(t'). \tag{5.3}$$

Applying this result successively to the field $\psi(t)$ in $f(t)_{N\beta}$, we obtain the two equivalent results [$A(t_i) = A_i$ for conciseness]

$$\begin{aligned} f(t)_{N\beta} &\equiv \langle N\beta\,|[\psi(t),\,\psi^+]_+|\,N\beta\rangle \\ &= \langle N\beta\,|[\psi(t)^0,\,\psi^+]_+|\,N\beta\rangle \\ &\quad - \frac{i}{\hbar}\int_0^t dt_1\langle N\beta\,|[A_1\psi(t)^0,\,\psi^+]|\,N\beta\rangle \\ &\quad + \left(\frac{i}{\hbar}\right)^2\int_0^t dt_1\int_0^{t_1} dt_2 \\ &\quad \times \langle N\beta\,|[A_1A_2\psi(t)^0,\,\psi^+]_+|\,N\beta\rangle \\ &\quad - \cdots, \end{aligned} \tag{5.4a}$$

$$\begin{aligned} &= \langle N\beta\,|[\psi(t)^0,\,\psi^+]_+|\,N\beta\rangle \\ &\quad - \frac{i}{\hbar}\int_0^t dt_1\langle N\beta\,|[\psi(t)^0A_1,\,\psi^+]|\,N\beta\rangle \\ &\quad + \left(\frac{i}{\hbar}\right)^2\int_0^t dt_1\int_0^{t_1} dt_2 \\ &\quad \times \langle N\beta\,|[\psi(t)^0A_2A_1,\,\psi^+]_+|\,N\beta\rangle \\ &\quad + \cdots, \end{aligned} \tag{5.4b}$$

depending upon whether we move A to the right or left each time (5.3) is applied. If we next observe that, for the F–D fields ψ, ψ^+ for which Eqs. (1.3) obtain,

$$\begin{aligned} &\mathrm{tr}_0\,([A_1\cdots A_n\psi(t)^0,\,\psi^+]_+) \\ &\qquad = \mathrm{tr}_0\,(A_1\cdots A_n[\psi(t)^0,\,\psi^+]_+), \\ &\mathrm{tr}_1\,([\psi(t)^0A_n\cdots A_1,\,\psi^+]_+) \\ &\qquad = \mathrm{tr}_1\,([\psi(t)^0,\,\psi^+]_+A_n\cdots A_1), \end{aligned} \tag{5.5}$$

and that for the $N\beta$ ensemble

$$\begin{aligned} &\langle N\beta\,|A_1\cdots A_n[\psi(t)^0,\,\psi^+]_+|\,N\beta\rangle \\ &\quad = \langle N\beta\,|A_1\cdots A_n|\,N\beta\rangle\langle N\beta\,|[\psi(t)^0,\,\psi^+]_+|\,N\beta\rangle, \end{aligned} \tag{5.6}$$

then by using relations of the type

$$\begin{aligned} &\int_0^t dt_1\int_0^{t_1} dt_2\langle N\beta\,|A_1A_2|\,N\beta\rangle \\ &\quad = \frac{1}{2!}\left[\int_0^t dt_1\int_0^{t_1} dt_2 + \int_0^t dt_2\int_{t_2}^t dt_1\right] \\ &\qquad \times \langle N\beta\,|A_1A_2|\,N\beta\rangle \\ &\quad = \frac{1}{2!}\int_0^t dt_1\,dt_2\langle N\beta\,|(A_1A_2)_0|\,N\beta\rangle, \end{aligned} \tag{5.7}$$

we may rewrite Eqs. (5.4) in the form

$$\begin{aligned} f(t)_{N\beta} &= f_e(t)_{N\beta}\Big\{1 - \frac{i}{\hbar} \\ &\quad \times \int_0^t dt_1\langle N\beta\,|A_1|\,N\beta\rangle + \frac{1}{2!}\left(\frac{i}{\hbar}\right)^2 \\ &\quad \times \int_0^t dt_1\,dt_2\langle N\beta\,|(A_1A_2)_N|\,N\beta\rangle - \cdots\Big\} \end{aligned} \tag{5.8a}$$

$$= f_e(t)_{N\beta}\left\langle N\beta\,\middle|\left(\exp\left\{-i\int_0^t dt'\,A(t')\right\}\right)_N\middle|\,N\beta\right\rangle. \tag{5.8b}$$

Equations (5.8) are equivalent to Eq. (1.9a). Equation (1.9b) follows from Eq. (5.8a) if we write

$$\begin{aligned} &\left\{1 - \frac{i}{\hbar}\int_0^t dt_1\langle N\beta\,|A_1|\,N\beta\rangle + \cdots\right\} \\ &\quad = \exp\left\{\ln\left[1 - \frac{i}{\hbar}\int_0^t dt_1\langle N\beta\,|A_1|\,N\beta\rangle + \cdots\right]\right\}, \end{aligned} \tag{5.9}$$

expand the logarithm in powers of its argument less unity, and rearrange the resulting series to collect equal powers of A in the same term. The proof is straightforward and requires no special techniques.[11]

Since Eqs. (1.9a,b) are simply algebraic rearrangements of the same result, one might ask why Eq. (1.9b) is generally to be preferred except in the moment calculations of Sec. 3. The answer is indicated fairly clearly by the linear-A harmonic-phonon case. In that case, only the first two terms of the series in Eq. (1.9b) are different from zero, whereas the series in Eq. (1.9a) does not terminate. Since Eq. (1.9b) is an expansion in terms of intrinsically higher-order correlations, we expect that, even for a general operator A and a general Hamiltonian (1.6), the first few terms of the series (1.9b) yield a more accurate representation of $g(t)_{N\beta}$ than do an equal number of terms from the series (1.9a).

6. ISOTOPE SHIFT IN NO-PHONON LINE

As a simple application of the preceding results, we compute the dependence of the position of the no-phonon spectral line upon impurity mass. For simplicity we restrict ourselves to the Hamiltonian (1.1), modified by the addition of a kinetic-energy perturbation deriving from the mass difference between the impurity and the lattice ion it replaces.

We assume that the index q in (1.1) is now a vector $\mathbf{q}$ of the reciprocal lattice, and that it identifies plane-wave phonon modes. Different phonon branches are taken into account by a suitable extended-zone scheme in the reciprocal lattice.

With a convenient choice of $\{a_q, a_q^+\}$ phases, the mass perturbation of H may be written:

$$\Delta H = \tfrac{1}{2}[\Delta(m_b^{-1})/m_b^{-1}] \sum_{qq'} \left(\frac{\hbar}{2nV}\right)(\omega_q\omega_{q'})^{\frac{1}{2}} \times (\mathbf{e}_{qb}a_q + \mathbf{e}_{qb}^* a_q^+)\cdot(\mathbf{e}_{q'b}a_{q'} + \mathbf{e}_{q'b}^* a_{q'}^+), \quad (6.1)$$

where $\Delta(m_b^{-1})/m_b^{-1} = [(m_b/m_I) - 1]$ measures the relative difference in impurity-ion (m_I) and lattice-ion (m_b) masses at the impurity site **b**, where $\mathbf{e}_{qb} = \mathbf{e}_{(-q)b}^*$ is the displacement vector associated with phonons **q** at site **b**, where $\omega_q = \omega_{-q}$ is the angular frequency in (1.1) of the phonons **q**, and where n is the number of unit cells per unit crystal volume. The vectors $\mathbf{e}_{qb}$ are normalized such that $\sum_b \mathbf{e}_{qb}^*\cdot\mathbf{e}_{qb} = 1$ when the **b** summation ranges over all atoms within a unit cell.[19]

The phases in Eq. (6.1) have been chosen such that the displacement from equilibrium $\mathbf{u}_{lb}$ of atom **b** in lattice cell **l** (relative to the impurity cell **l** = 0) is

$$\mathbf{u}_{lb} = i \sum_q \left(\frac{\hbar}{2m_b\omega_q nV}\right)^{\frac{1}{2}} \times \{e^{i\mathbf{q}\cdot\mathbf{l}}\mathbf{e}_{qb}a_q - e^{-i\mathbf{q}\cdot\mathbf{l}}\mathbf{e}_{qb}^* a_q^+\}. \quad (6.2)$$

It is important to specify these phases since they also enter (1.1) through the coupling coefficients C_q. The relative phases reflect whether the impurity or some other ion is responsible for the mass modification (6.1).

In the system we consider, the total Hamiltonian remains harmonic. The no-phonon line is described exactly by the parameters of Eq. (2.12) with $\langle 0\beta|\ A\ |0\beta\rangle = 0$ and $\Gamma_0 = 0$. To determine the mass shift of the no-phonon line it is convenient to compute the retarded correlation function $P_{20}(t) \equiv \langle 0\beta|\ [A(t), A(0)]\ |0\beta\rangle\ \theta(t)$. In terms of the spectral function $\rho_{20}(\omega)$ defined in (2.11), the Fourier transform of $P_{20}(t)$ is

$$p_{20}(\omega) = \int_0^\infty dt\, e^{i\omega t}\langle 0\beta\ |[A(t), A(0)]|\ 0\beta\rangle = i\int_{-\infty}^{\infty} \frac{d\bar{\omega}}{2\pi}\frac{\rho_{20}(\bar{\omega})}{\omega - \bar{\omega} + i\epsilon}, \quad (6.3)$$

$\epsilon = 0^+$. Since $\langle 0\beta|\ A\ |0\beta\rangle = 0$, it follows easily from the properties (2.11) and from Eq. (2.12c) that the displacement of the no-phonon line from the unperturbed electronic level $\mathcal{E}$ is

$$\Delta\mathcal{E}_0 = p_{20}(\omega = 0)/2i\hbar. \quad (6.4)$$

[19] Our phonon notation parallels that of J. M. Ziman, *Electrons and Phonons* (Clarendon Press, Oxford, England, 1962), Chap. I.

More generally,

$$\rho_{20}(\omega) = 2\,\mathrm{Re}\,p_{20}(\omega). \quad (6.5)$$

The Heisenberg equation of motion

$$i\frac{\partial}{\partial t}a_q = \omega_q a_q + [(m_b/m_I) - 1]\sum_{q'}[(\omega_q\omega_{q'})^{\frac{1}{2}}/(2nV)] \times \mathbf{e}_{qb}^*\cdot(\mathbf{e}_{q'b}a_{q'} + \mathbf{e}_{q'b}^* a_{q'}^+) \quad (6.6)$$

follows immediately from the Hamiltonians (1.1) and (1.6). A corresponding equation for a_q^+ follows by Hermitian conjugation. Using this equation with the definition (1.10) of the operator A, we find in a straightforward fashion that ($\epsilon = 0^+$)

$$\begin{aligned}\frac{1}{i}p_{20}(\omega) = {} & \{\sum_q 2\omega_q\,|C_q|^2/[(\omega + i\epsilon)^2 - \omega_q^2]\} \\ & + \{[(m_b/m_I) - 1]/(2nV)\} \\ & \times \{\sum_q \omega_q^{\frac{1}{2}}[\omega(C_q^*\mathbf{e}_{qb}^* - C_q\mathbf{e}_{qb}) \\ & + \omega_q(C_q^*\mathbf{e}_{qb}^* + C_q\mathbf{e}_{qb})]/[(\omega + i\epsilon)^2 - \omega_q^2]\} \\ & \cdot\mathsf{D}(\omega)\cdot\{\sum_q \omega_q^{\frac{1}{2}}[\omega(\mathbf{e}_{qb}C_q - \mathbf{e}_{qb}^*C_q^*) \\ & + \omega_q(\mathbf{e}_{qb}C_q + \mathbf{e}_{qb}^*C_q^*)]/[(\omega + i\epsilon)^2 - \omega_q^2]\}. \end{aligned} \quad (6.7)$$

The dyadic $\mathsf{D}(\omega)$ is the solution of the equation

$$\mathsf{D}(\omega) = \mathsf{I} + \mathsf{M}_b(\omega)\cdot\mathsf{D}(\omega) \quad (6.8a)$$

with

$$\mathsf{M}_b(\omega) = \{[(m_b/m_I) - 1]/(2nV)\} \times \{2\sum_q \omega_q^2\mathbf{e}_{qb}\mathbf{e}_{qb}^*/[(\omega + i\epsilon)^2 - \omega_q^2]\}. \quad (6.8b)$$

Using this result in Eq. (6.4), we find the no-phonon displacement to be

$$\Delta\mathcal{E}_0 = -\{\sum_q |C_q|^2/\hbar\omega_q\} + \tfrac{1}{2}[(m_b/m_I) - 1]\boldsymbol{\pi}_b(C_q)\cdot\mathsf{D}(0)\cdot\boldsymbol{\pi}_b(C_q), \quad (6.9)$$

where the vector

$$\boldsymbol{\pi}_b(C_q) = \sum_q (C_q\mathbf{e}_{qb} + C_q^*\mathbf{e}_{qb}^*)/(2nV\hbar\omega_q)^{\frac{1}{2}} = \sum_q \mathbf{e}_{qb}(C_q + C_{-q}^*)/(2nV\hbar\omega_q)^{\frac{1}{2}}. \quad (6.10)$$

In the limit $\omega \to 0$, Eq. (6.8b) reduces to

$$\mathsf{M}_b(0) = -\{[(m_b/m_I) - 1]/nV\}\sum_q \mathbf{e}_{qb}\mathbf{e}_{qb}^* = [1 - (m_b/m_I)]\mathsf{I}, \quad (6.11a)$$

where the second expression obtains because the **q** summation is over a complete set of phonon modes.

For the special case (6.11a), the solution of Eq. (6.8a) is trivial:

$$\mathsf{D}(0) = (m_I/m_\mathrm{b})\mathsf{I}. \tag{6.11b}$$

Thus,

$$\Delta\mathcal{E}_0 = -\{\sum_\mathrm{q} |C_\mathrm{q}|^2/\hbar\omega_\mathrm{q}\} - \tfrac{1}{2}[(m_I - m_\mathrm{b})/m_\mathrm{b}]\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})\cdot\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q}), \tag{6.12}$$

where $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})$ is defined in Eq. (6.10).

Using the Schwarz inequality with Eq. (6.12), we find successively that

$$|\Delta\mathcal{E}_0 + \sum_\mathrm{q} |C_\mathrm{q}|^2/\hbar\omega_\mathrm{q}| \leq \tfrac{1}{4}[|m_I - m_\mathrm{b}|/m_\mathrm{b}] \sum_\mathrm{q} |C_\mathrm{q} + C^*_{-\mathrm{q}}|^2/\hbar\omega_\mathrm{q}, \tag{6.13a}$$

$$\leq [|m_I - m_\mathrm{b}|/m_\mathrm{b}] \sum_\mathrm{q} |C_\mathrm{q}|^2/\hbar\omega_\mathrm{q}. \tag{6.13b}$$

The factor $\sum_\mathrm{q} |C_\mathrm{q}|^2/\hbar\omega_\mathrm{q}$ in (6.13b) can be evaluated experimentally if one measures the distance between the no-phonon line and the mean spectral frequency.

It is interesting to evaluate the mass shift in (6.12) when the phonon-impurity coupling depends upon the dilation (trace) of the site-b "strain tensor"

$$\mathsf{S}_\mathrm{b} = -\tfrac{1}{2} \sum_\mathrm{q} \{(\hbar/2m_\mathrm{b}\omega_\mathrm{q}nV)^{\frac{1}{2}} \times [(\mathbf{q}\mathbf{e}_{\mathrm{q}\mathrm{b}} + \mathbf{e}_{\mathrm{q}\mathrm{b}}\mathbf{q})a_\mathrm{q} + \mathrm{c.\,c.}]\} \tag{6.14}$$

appropriate to the displacement (6.2). For this case we take

$$A \equiv \sum_\mathrm{q} (C_\mathrm{q} a^+_\mathrm{q} + C^*_\mathrm{q} a_\mathrm{q}) = C \sum_{i=1}^{3} (\mathsf{S}_\mathrm{b})_{ii}, \tag{6.15}$$

where C is a real coupling constant. For this coupling, which can provide a useful phenomenological representation of the real phonon-impurity coupling,[20,21] the constants C_q are

$$C_\mathrm{q} = -C\mathbf{q}\cdot\mathbf{e}^*_{\mathrm{q}\mathrm{b}}(\hbar/2m_\mathrm{b}\omega_\mathrm{q}nV)^{\frac{1}{2}}. \tag{6.16a}$$

Since $\mathbf{e}^*_{\mathrm{q}\mathrm{b}} = \mathbf{e}_{(-\mathrm{q})\mathrm{b}}$, this implies that

$$C_\mathrm{q} + C^*_{-\mathrm{q}} = 0 \quad \text{and} \quad \boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q}) = 0. \tag{6.16b}$$

The no-phonon line exhibits *no* isotopic frequency shift in spite of the fact that the spectral function $\rho_{20}(\omega)$ determined by Eq. (6.5) will display mass difference $(m_I - m_\mathrm{b})$ perturbations.

It is notable that the observed isotopic shift of the no-phonon components of the R lines in pink ruby[22] is about forty times smaller than is predicted by the right-hand side of Eq. (6.13b). It is very likely that this very large reduction from the estimate of Eq. (6.13b) results from a cancellation of the type appearing in Eq. (6.16b).

Crystal symmetry often restricts the properties of the vector $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})$ in Eqs. (6.10) and (6.12). If G_b is the group of crystal symmetry operations which leave the impurity site **b** invariant, the group G_b is a point group and a subgroup of the full crystal symmetry group. Its operations leave the electronic, the lattice-phonon, and the interaction parts of the system Hamiltonian separately invariant. If the group G_b contains the site-**b** inversion operator I_b, we expect the vector $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})$ to vanish and the no-phonon line *appropriate to the Hamiltonian* (1.1) to exhibit no isotopic frequency shift.

Without developing the full proof by group theory, we can indicate by the following simple argument why $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q}) = 0$ when G_b contains I_b. Invert the impurity electronic configuration, leaving the lattice and coordinate systems unchanged. Because for I_b in G_b the electronic Hamiltonian is inversion invariant and the wavefunctions are therefore eigenfunctions of the parity operator P, the inverted wavefunctions must be $P' = \pm 1$ times the original wavefunctions. Since the phonon-impurity interaction in Eq. (1.1) is diagonal in the electronic states, the parity eigenvalue P' enters the matrix elements C_q through its square $(P')^2 = +1$. That is, because the C_q depend upon the absolute value squared of the electronic wavefunction, the coefficients C_q and hence the vector $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})$ are invariant with respect to electronic inversion. However, since in Eq. (6.10) we indiscriminately sum over *all* vectors $\mathbf{q}$ which refer to the lattice phonons, any directional preference in the vector $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q})$ must arise from the electronic configuration. In particular, we expect $\boldsymbol{\pi}_\mathrm{b} \rightarrow -\boldsymbol{\pi}_\mathrm{b}$ when the electronic configuration is inverted. This is consistent with our previous remarks only if $\boldsymbol{\pi}_\mathrm{b}(C_\mathrm{q}) = 0$.

ACKNOWLEDGMENTS

I am indebted to C. G. B. Garrett and M. D. Sturge for the stimulation to undertake this work and for constructive comments regarding its application to real crystal systems. I also wish to acknowledge profitable conversations with D. A. Kleinman, M. Lax, and P. A. Wolff.

[20] D. E. McCumber and M. C. Sturge, J. Appl. Phys. 34, 1682 (1963).

[21] D. E. McCumber, Phys. Rev. 130, 2271 (1963).

[22] A. L. Schawlow, *Advances in Quantum Electronics*, edited by J. R. Singer (Columbia University Press, New York, 1961), p. 50.

Stroboscopic-Perturbation Procedure for Treating a Class of Nonlinear Wave Equations

MARTIN D. KRUSKAL
Princeton University Plasma Physics Laboratory, Princeton, N. J.
AND
NORMAN J. ZABUSKY
Bell Telephone Laboratories, Incorporated, Whippany, N. J.
(Received 9 July 1963)

A new perturbation procedure is presented for treating initial-value problems of nonlinear hyperbolic partial differential equations. The *characteristic variables* of the partial differential equation and the functions of these variables are expanded in powers of ϵ, and the formal solution is uniformly valid over time intervals $O(1/\epsilon)$. The uniform first-order solution is evaluated for the equation

$$y_{tt} = (1 + \epsilon y_x)y_{xx},$$

subject to the standing-wave initial conditions: $y(x, 0) = a \sin \pi x$, $y_t(x, 0) = 0$. This equation is the lowest continuum limit of an equation for which numerical computations are available. The uniform zero-order solution breaks down after a time $t_B = 4/\epsilon a\pi$. A detailed study of the solution is made in the vicinity of the breakdown region of the (x, t) plane, and it demonstrates that the formal solution for y_x and y_t goes from a single-valued to a triple-valued function while y_{xx} and y_{tt} become singular. To compare the solutions with the available numerical computations, the y_x and y_t waveforms are decomposed into spatial Fourier modes. The effect of breakdown is manifest in the modal amplitudes $\propto J_n(nT)/nT$. The modal amplitudes change their asymptotic behavior, from exponentially decreasing as $n \to \infty$, to algebraically decreasing when t goes from smaller to larger than t_B. In the time interval up to breakdown, $t < t_B$, the modal energies are in excellent agreement with the modal energies of the numerical computations, whereas for $t > t_B$ they diverge. For $t < t_B$, the total energy calculated from the uniform zero-order solution is conserved and equal to the initial value,

$$E_C^{(0)} = 2a^2\pi^2 \sum_{n=1}^{\infty} \left[\frac{J_n(nT)}{(nT)}\right]^2 = \tfrac{1}{2}a^2\pi^2 = E_C\,|_{t=0},\ t \leq t_B.$$

Thus, the lowest-continuum-limit equations describe the dynamics of a discrete model for a finite time. A heuristic discussion is given which suggests that the time of description can be extended beyond t_B by including higher spatial derivatives in the continuum model.

1. INTRODUCTION AND SUMMARY

THE representation of wave motion by means of signals propagated along the characteristics of the wave system has proved to be a powerful tool.[1,2,3] The method of characteristics is especially useful for treating initial-value problems. The concept of "domains of dependence" which influence a point in the (r, t) volume and the concept of "range of influence" of a point along the initial-condition surface provide a cogent image for appreciating the physical significance of the results which underlie the mathematics.

The subject of this paper is the development and application of a uniform perturbation procedure for a large class of nonlinear hyperbolic equations. This is accomplished by resolving the partial differential equation into a set of first-order ordinary differential equations, each describing a one-parameter family of characteristics. The characteristic variables are used as the basis for the perturbation expansion.[4]

Many of the nonlinear wave equations of mathematical physics are derivable from Lagrangians which are basically analogous. The equations of one-dimensional elasticity, hydrodynamics, magnetohydrodynamics, and electrodynamics all yield equations which can be treated by the above-mentioned perturbation procedure. In all these cases one finds similar physical phenomena, as, for example, the exchange of energy among the

[1] R. Courant, *Methods of Mathematical Physics II (Partial Differential Equations)* (Interscience Publishers, Inc., New York, 1962).
[2] T. Taniuti, Progr. Theoret. Phys. (Kyoto) Suppl. 9, 69 (1958).
[3] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948).
[4] P. A. Fox, J. Math. and Phys. 34, 133 (1955). This paper applies a perturbation method to the study of a one-dimensional polytropic fluid, that is, a fluid whose sound speed $c \propto (\text{density})^{\frac{1}{2}(\gamma-1)}$. The spirit of the method is the same as that contained in the present work because the characteristic variables are used as the basis for the perturbation expansion. However, the formal mathematical development of the method, its general usefulness, and its detailed application to a physical problem are distinctly different in the two. See also the related work by C. C. Lin, J. Math. and Phys. 33, 117 (1954). The authors wish to thank Dr. Joseph B. Keller for bringing these references to their attention.

various modes of oscillation of the system. Of particular interest is the development of discontinuities or "shocks" in the solution after times (t_B) which are related to the strength of the nonlinearity and the initial excitation. The perturbation solution will exhibit this discontinuity or "breakdown" in zero order. The partial differential equation does not describe the physical processes beyond this time, and recourse must be made to higher-order processes to resolve the nature of the breakdown.[5]

At first, the perturbation procedure will be illustrated for a one-dimensional system whose Lagrangian density is the ordinary "linear" energy density difference[6] $\frac{1}{2}(y_t^2 - y_x^2)$ plus two small terms, one being an arbitrary function of y_t and the other an arbitrary function of y_x. We limit our discussion to nonlinearities which are functions only of derivatives of y.

We will apply the stroboscopic-perturbation procedure in detail to the Lagrangian density,[7]

$$\mathcal{L} = \tfrac{1}{2}[y_t^2 - y_x^2] - \epsilon[(p+2)(p+1)]^{-1} y_x^{(p+2)}. \qquad (1.1)$$

The work of Fermi, Pasta, and Ulam (FPU)[8] on the standing longitudinal oscillations of a *nonlinear* beaded string motivated the choice of the Lagrangian [Eq. (1.1)]. It represents the typical beaded Lagrangian which they studied, in the "lowest" continuum limit, that is, the limit of infinitely dense, infinitesimal beads and springs.

The numerical computations of FPU for standing oscillations showed unexpected recurrence phenomena. The present authors attempted to describe these phenomena with the Lagrangian density of (1.1), and this led to the development of the procedure which constitutes this paper. It is found, however, that after a finite time of order $1/\epsilon$, the perturbation solutions of the equations of motion obtained from (1.1) exhibit a breakdown, in that the second derivative y_{xx} becomes singular. This conclusion was rigorously established by one of the authors (N.J.Z.)[9] for the original exact partial differential equation with $p = 1$. By using an inversion or hodograph transformation, he put the nonlinear wave equation into an equivalent linear form. For $p = 1$ he obtained an *exact* solution in terms of known functions and proved rigorously that breakdown occurs.

Up until the breakdown time t_B, the results of the zero-order perturbation procedure agree with the numerical computations.[10] At about this time, the numerical solutions cannot be described by the Lagrangian density given in (1.1). To extend the continuum analysis beyond t_B, one must modify the lowest-order continuum Lagrangian to include higher-order (spatial derivative) terms. This aspect of the problem is discussed in Sec. 6 and is presently under investigation.

2. DECOMPOSITION OF THE EQUATIONS OF MOTION INTO THEIR CHARACTERISTIC REPRESENTATION

A. The Stroboscopic Concept[11]

In treating linear phenomena one almost always invokes the principle of superposition. This allows one to decompose a complicated motion into a sum of simpler motions, called states or modes, each of which can be analyzed on a more elementary level. For nonlinear systems the principle is of course invalid.

For nearly linear systems[12] we can develop a

[5] R. Courant, Ref. 3. See, for example, Appendix II to Chapter II, and Chapter V, paragraph 9. The occurrence of discontinuities in the solution of nonlinear second-order hyperbolic equations is a well known phenomenon. Much of the recent work in this field has been motivated by hydrodynamic considerations. To resolve the discontinuities, "viscositylike" terms have been included in the equations of motion. In essence, the included term is of one higher spatial derivative than those which characterize the wave motion (a third-spatial-derivative term $\epsilon' \partial^3 y/\partial x^3$ would be a typical term). For propagation in solids, a more natural term to add would be of an "elastic" nature, for example, $\epsilon' \partial^4 y/\partial x^4$. This renders the system dispersive, whereas with viscositylike terms the system is made dissipative.

[6] Subscripts are used to designate partial derivatives, $y_t = \partial y/\partial t$.

[7] This Lagrangian also describes the longitudinal elastic vibrations of a bar whose stress (σ)–strain (s) law is characterized by the relation

$$d\sigma/ds = 1 + \epsilon y_x^p.$$

[8] E. Fermi, J. R. Pasta, and S. Ulam, Los Alamos Rept. No. 1940, May, 1955. The problem studied in this report is described briefly in S. Ulam, *A Collection of Mathematical Problems* (Interscience Publishers, Inc., New York, 1960) Chap. 7, paragraph 8. Recently, Joseph Ford [J. Math. Phys. **2**, 787 (1961)] and E. Atlee Jackson [J. Math. Phys. **4**, 551, 686 (1963)] have attempted to explain the recurrence phenomena described in the FPU report. For further discussion see N. J. Zabusky in *Mathematical Models in Physical Sciences*, edited by S. Drobot (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963), p. 99 and M. D. Kruskal, in a talk at International Business Machines, Inc., Conference on Numerical Computations in Large-Scale Physical Phenomena, Dec. 1963 (to be published).

[9] N. J. Zabusky, J. Math. Phys. **3**, 1028 (1962).

[10] The method of comparing the analytical and numerical results will be made clear in Sec. 5, to follow. Essentially, it involves a comparison of the energy in the normal modes obtained by decomposing the waveform into spatial Fourier modes.

[11] The stroboscopic concept is implicit in the more comprehensive work of one of the authors. [M. D. Kruskal, J. Math. Phys. **3**, 806 (1962)]. In this paper an asymptotic theory is presented for treating physical systems described by coupled first-order differential equations whose solutions are nearly periodic.

[12] In a nearly linear system, the nonlinearity enters with a small parameter ϵ. This is exhibited in (1.1). The "corresponding linear system" is defined as the one obtained by setting $\epsilon = 0$.

perturbation procedure applicable to both standing and progressive oscillation phenomena. This is accomplished by decomposing the partial differential equations of the system into a set of ordinary differential equations, each of the latter describing a one-parameter family of characteristics. Moving along a particular characteristic direction, one finds that variations associated with opposite characteristics appear as rapid oscillations. These rapid oscillations are associated with the corresponding linear system and can be "averaged out" in each order of the perturbation calculation, which, in effect, decouples the phenomena along oppositely directed characteristics. One performs the averaging over one period of the linear oscillation and obtains a period-by-period or stroboscopic view of the waveform. This stroboscopic waveform will vary "slowly," that is, in accordance with a time scale governed by the nonlinear coupling constant.

It may be mentioned that the final solution is constructed from a set of relationships that give the (x, t) path of each characteristic *implicitly* in terms of the characteristic variables. This is unlike a linear analysis, where one merely superimposes the results from elementary states.

B. Decomposition of the One-Dimensional Wave Equation into a Characteristic Representation

Consider the partial differential equation

$$[\Gamma(y_t)]^2 y_{tt} = [\Phi(y_x)]^2 y_{xx}, \tag{2.1}$$

where

$$\Phi = 1 + \epsilon\mathfrak{F}(y_x), \qquad \Gamma = 1 + \epsilon\mathfrak{G}(y_t). \tag{2.2}$$

In general, $\mathfrak{F}$ and $\mathfrak{G}$ are written as power series in ϵ beginning with terms $O(1)$. The linear sound speed has been normalized to unity and we have chosen the nonlinearity to be separable into a product of a function of y_x by a function of y_t, a form frequent in applications.

If we define[13]

$$u = y_x, \qquad v = y_t, \tag{2.3}$$

then (2.1) can be written as two coupled partial differential equations,

$$u_t - v_x = 0, \tag{2.4}$$

$$\Gamma^2(v)v_t - \Phi^2(u)u_x = 0, \tag{2.5}$$

where (2.4) is a consistency condition. We now introduce new dependent variables which serve to symmetrize the coupled pair of equations. This is a standard transformation in the method of characteristics, and in our case is made by multiplying (2.4) by $\pm\Phi\Gamma$ and adding the result to (2.5). We obtain the symmetric pair of equations,

$$\Gamma^2 v_t \pm \Phi\Gamma u_t \mp (\Phi\Gamma v_x \pm \Phi^2 u_x) = 0. \tag{2.6}$$

Dividing by $\Gamma(v)$ we obtain

$$(r_\pm)_t \mp [\Phi(u)/\Gamma(v)](r_\pm)_x = 0, \tag{2.7}$$

where

$$r_\pm = \tfrac{1}{2}\left[\pm\int_0^v \Gamma(v')\,dv' + \int_0^u \Phi(u')\,du'\right]$$
$$= \tfrac{1}{2}[\pm A(v) + B(u)]. \tag{2.8}$$

The new dependent variables $r_\pm$ are known as the *Riemann invariants* of the given partial differential equation (2.1). They and any function of either of them are constant along the solutions of their respective characteristic ordinary differential equations,

$$dx/dt = \mp\Phi(u)/\Gamma(v) \quad \text{for} \quad r_\pm. \tag{2.9}$$

Instead of $r_\pm$ on the left side of (2.8), we could have taken an arbitrary function of $r_\pm$. We have chosen the form shown, with the multiplicative constant $\frac{1}{2}$ and zero on the lower limit of the integrals (corresponding to an additive constant), so that $r_\pm$ goes over to the conventional expression for the invariants in the corresponding linear system. The coefficient Φ/Γ in (2.7) can be expressed as the ratio of a function of $(r_+ + r_-)$ to a function of $(r_+ - r_-)$ by using the inverse of the relations obtained by adding and subtracting the two equations (2.8). Thus,

$$u = B^{-1}(r_+ + r_-), \qquad v = A^{-1}(r_+ - r_-). \tag{2.10}$$

C. Representation of the Exact Solution by a Pair of Implicit Relations

If one knew, *a priori*, the exact path of the characteristics in the (x, t) plane, one could write down the exact solution to an initial-value problem. In this section we derive the solution in the form of a pair of implicit relations. These will serve as the starting point of the perturbation procedure.

We observe that the coefficient of $(r_\pm)_x$ in (2.7) can be written as

$$\mp[\Phi/\Gamma] = \mp[1 + \epsilon F(r_+ + r_-, r_+ - r_-)], \tag{2.11}$$

since Φ is a function only of u, and Γ is a function only of v. For the perturbation method, F can be expanded in a power series in ϵ if $r_\pm$ remain of the same order of magnitude as initially. This is the

[13] N. J. Zabusky, Ref. 9, paragraph 3, where $\Gamma = 1$.

case for the FPU problem, where it is only the derivatives of $r_\pm$ that become unbounded.

We next observe that the left sides of Eqs. (2.7) are in the form of convective derivatives, that is, a time derivative plus a multiple of a space derivative. This suggests that if we transform to a suitable moving reference frame, the equations obtained will be more amenable to analysis. A Lagrangian transformation[14] will replace convective derivatives in one reference frame by partial time derivatives in another reference frame. Thus, we make a different Lagrangian transformation of the independent variables for each equation (2.7), namely

$$t = \tau,$$
$$x = \xi_\pm \mp \left[\tau + \epsilon \int_0^\tau F(r_+, r_-)\, d\tau'_\pm\right], \tag{2.12}$$

and obtain the pair of equations

$$\partial(r_\pm)/\partial\tau = 0, \tag{2.13}$$

where the upper and lower signs always correspond to r_+ and r_- respectively. Thus, $r_\pm$ are independent of τ, so

$$r_\pm = R_\pm(\xi_\pm), \tag{2.14}$$

where[15] $R_\pm(x')$ are the initial values ($t = 0$) of the Riemann invariants $r_\pm$.

No subscripts are appended to x and t as they are the generic space-time variables. The integration in (2.12) is performed *along characteristics*. This is represented symbolically by appending the subscript $\pm$ to τ'. Since $r_\pm$ in (2.12) can be expressed in terms of $\xi_\pm$ by (2.14), we find it convenient to change the variables of integration from $\tau'_\pm$ to $\xi'_\mp$,

$$d\tau'_\pm = J_\mp\, d\xi'_\mp = \partial t/\partial\xi'_\mp|_{\xi_\pm}\, d\xi'_\mp. \tag{2.15}$$

In other words, $\xi_\mp$ varies if we move along the $\xi_\pm$ characteristic. Thus, (2.12) takes the form

$$\xi_\pm = x \pm t \pm \Xi_\pm, \tag{2.16}$$

where

$$\Xi_\pm = \epsilon \int_{\xi_\pm}^{\xi_\mp} F J_\mp\, d\xi'_\mp \tag{2.17}$$

is a function of ξ_+ and ξ_-. That is, one expresses F in (2.12) in terms of ξ_+ and ξ_- by using (2.14). The lower limit of the integral follows from (2.12) at $\tau = t = 0$, namely $\xi_\mp|_{t=0} = \xi_\pm = x'$.

We obtain $J_\mp$ by first eliminating x from the pair of equations (2.16),

$$t = \pm\tfrac{1}{2}(\xi_\pm - \xi_\mp) - \tfrac{1}{2}(\Xi_\pm + \Xi_\mp), \tag{2.18}$$

and then taking the partial derivative of t with respect to $\xi_\mp$, holding $\xi_\pm$ fixed, or

$$J_\mp = \mp\tfrac{1}{2} - \tfrac{1}{2}(\Xi_\pm + \Xi_\mp)_{\xi_\mp}. \tag{2.19}$$

Hence, the problem of solving a nonlinear partial differential equation has been cast into one involving the solution of a pair of coupled linear integral equations of the second kind, namely (2.17).

The step-by-step manner of computing $J_\mp$ and of evaluating the integrals is the essence of the perturbation procedure. We make the Ansatz that the function $\Xi_\pm$ can be calculated in nth order by using the function $J_\mp$ from the previous order, or

$$\Xi_\pm^{(n)} = \int_{\xi_\pm}^{\xi_\mp} \epsilon F J_\mp^{(n-1)}\, d\xi'_\mp + O(\epsilon^{n+1}), \tag{2.20}$$

where

$$J_\mp^{(n-1)} = \mp\tfrac{1}{2} - \tfrac{1}{2}[\Xi_\pm^{(n-1)}]_{\xi_\mp} - \tfrac{1}{2}[\Xi_\mp^{(n-1)}]_{\xi_\mp}, \qquad n \geq 0. \tag{2.21}$$

It is not at all obvious that this procedure yields a convergent solution, for although $\Xi_\pm$ is defined with the coefficient ϵ, the range of integration in (2.17) is $O(1/\epsilon)$. In fact, we must validate the assumed procedure in each order by showing that $\Xi_\pm^{(n)}$ is unchanged if $J_\mp$ is determined from it in an iterative fashion.

In Secs. 3B and 3C we will show that the initial conditions as well as the nonlinearity govern the validation calculation.

Equation (2.14) with $\xi_\pm$ defined by (2.16) gives an implicit representation of the solution and thus will not always define a single-valued function $\xi_\pm(x, t)$. The implicit function theorem[16] states that if the Jacobian J does not vanish, then the relations (2.16) do define a single-valued function, where

$$J = \frac{\partial[(x + t), (x - t)]}{\partial(\xi_+, \xi_-)} = \{1 - (\Xi_+)_{\xi_+}\} \times \{1 + (\Xi_-)_{\xi_-}\} + (\Xi_+)_{\xi_-}(\Xi_-)_{\xi_+} \tag{2.22}$$

and $(x \pm t)$ are defined in (2.16) and are treated as functions of ξ_+ and ξ_-.

In the problem to be studied in Sec. 3, it will be

[14] The Lagrangian representation of the equations of hydrodynamics is sometimes preferred to the Eulerian form. The form of these equations and their meaning are given by J. Serrin, "Mathematical Principles of Classical Fluid Mechanics," *Handbuch der Physik* (Springer Verlag, Berlin, 1959), Vol. VIII, Pt. 1, p. 125. If a partial differential equation is presented in convective (Eulerian) form, it may be transformed to a stationary form by using an appropriate Lagrangian transformation.

[15] For clarity, we use x' to designate the independent spatial variable at $t = 0$.

[16] T. M. Apostol, *Mathematical Analysis* (Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1957), p. 146, paragraph 7-5.

shown that the last term of the right side of (2.22) is of higher order in ϵ than the first is. Thus, to $O(\epsilon)$, the Jacobian vanishes whenever either of the relations holds:

$$1 = \pm(\Xi_{\pm})_{\xi_{\pm}} + O(\epsilon^2). \tag{2.23}$$

The earliest time that this occurs is called the *breakdown time*, t_B. In the (x, t) plane, breakdown occurs when two neighboring characteristics intersect. Beyond this time the formal solution is multivalued in certain regions (and therefore physically meaningless), since there is more than one characteristic of the same family passing through a point. We will show below that at $t \approx t_B$, $(r_+)_x$ or $(r_-)_x$ becomes singular at isolated points along the x axis.

3. A STROBOSCOPIC-PERTURBATION PROCEDURE AND ITS APPLICATION TO THE FERMI-PASTA-ULAM PROBLEM

A. Preliminaries

We will now develop a uniform order-by-order perturbation procedure for calculating the path of the separate characteristics in the physical plane. By a uniform procedure, we mean one which yields a valid solution over time intervals $O(1/\epsilon)$, that is we treat the quantity $\epsilon^{n+1}t$ as $O(\epsilon^n)$. This procedure starts from the equations derived in Sec. 2 and will be applied to the wave equation[17]

$$y_{tt} = (1 + \epsilon y_x)y_{xx}, \tag{3.1}$$

obtained from (2.1) by setting $\Gamma = 1$ and $\Phi = (1 + \epsilon y_x)^{\frac{1}{2}}$. This corresponds to the Lagrangian (1.1) with $p = 1$. Equation (3.1) with periodic initial conditions and fixed boundary conditions has been solved exactly.[9] The power of the perturbation procedure will become evident by the facility with which it generates the power series expansion of the exact solution.

We examine, as FPU did, the standing-wave oscillations resulting from the initial conditions

$$y(x', 0) = a \sin \pi x', \quad \text{and} \quad y_t(x', 0) = 0. \tag{3.2}$$

The new independent variables defined in (2.8) become

$$r_{\pm} = \tfrac{1}{2}\{\pm v + (2/3\epsilon)[(1 + \epsilon u)^{\frac{3}{2}} - 1]\},$$
$$= \tfrac{1}{2}\{\pm v + u + (\tfrac{1}{4}\epsilon)u^2 - (\tfrac{1}{24}\epsilon^2)u^3\} + O(\epsilon^3), \tag{3.3}$$

so that their initial values ($v = 0$, $u = a\pi \cos \pi x'$) are

$$R_{\pm}(x') = (1/3\epsilon)[(1 + \epsilon a\pi \cos \pi x')^{\frac{3}{2}} - 1]$$
$$= \tfrac{1}{2}[a\pi \cos \pi x'] + (\tfrac{1}{8}\epsilon)[a\pi \cos \pi x']^2$$
$$- (\tfrac{1}{48}\epsilon^2)[a\pi \cos \pi x']^3 + \cdots. \tag{3.4}$$

An equivalent initial-value problem over the infinite x space is obtained by replacing the fixed boundary conditions $y(0, t) = y(1, t) = 0$ by the requirement that the solution $y(x, t)$ be *periodic* over a length of 2 and *odd* in x.

Φ can be written as a function of $(r_+ + r_-)$ by inverting the sum of Eqs. (3.3),

$$\Phi = (1 + \epsilon F) = (1 + \epsilon u)^{\frac{1}{2}}$$
$$= [1 + (\tfrac{3}{2}\epsilon)(r_+ + r_-)]^{\frac{1}{3}}. \tag{3.5}$$

Expanding the right side of (3.5) and comparing yields

$$\epsilon F = \tfrac{1}{2}\epsilon(r_+ + r_-) - \tfrac{1}{4}\epsilon^2(r_+ + r_-)^2$$
$$+ (\tfrac{5}{24})\epsilon^3(r_+ + r_-)^3 + \cdots. \tag{3.6}$$

Substituting for $r_{\pm}$ the relations given in (2.14) and (3.4), one finds

$$\epsilon F = \beta(\cos \pi\xi_+ + \cos \pi\xi_-) - 2\beta^2(\cos \pi\xi_+ \cos \pi\xi_-)$$
$$- \beta^3(\cos^3 \pi\xi_+ - 3 \cos^2 \pi\xi_+ \cos \pi\xi_-$$
$$- 3 \cos \pi\xi_+ \cos^2 \pi\xi_- + \cos^3 \pi\xi_-) + O(\beta^4), \tag{3.7}$$

where

$$\beta = \tfrac{1}{4}\epsilon a\pi. \tag{3.8}$$

B. The Uniform Zero-Order Solution[18] $S^{(n)}$

$S^{(0)}$ is obtained by including terms $O(\epsilon t)$ or $O(\beta t)$ and neglecting terms $O(\epsilon)$. If the leading terms of ϵF as defined in (3.7) are substituted into (2.20), we obtain

$$\Xi_{\pm} = \beta \int_{\xi_{-*}}^{\xi_{-}} (\cos \pi\xi_{\pm} + \cos \pi\xi'_{\mp})$$
$$\times (\mp\tfrac{1}{2})\, d\xi'_{\mp} + O(\beta), \tag{3.9}$$

where we have taken $J^{(-1)}_{\mp} = \mp\frac{1}{2}$, according to the Ansatz made in Sec. 2C. This is the value one obtains from the linear partial differential equation ($\epsilon = 0$), where

$$\xi_{\mp} = \xi_{\pm} \mp 2t. \tag{3.10}$$

[17] This equation was used by Ornstein and Zernike as a model for studying the effects of nonlinear elastic phenomena on the thermal conductivity of nonconducting crystals. However, they linearized the equation early in the analysis and thereby failed to observe breakdown. Thus, they did not realize that this equation could be used only to describe short-time phenomena. See L. S. Ornstein and F. Zernike, Proc. Roy. Acad. Soc., Amsterdam **19**, 1295 (1917).

[18] We designate by $S^{(n)}$ the phrase "uniform nth-order solution."

Then

$$\Xi_\pm = \pm\tfrac{1}{2}\beta(\xi_\pm - \xi_\mp)\cos\pi\xi_\pm + O(\beta). \quad (3.11)$$

The second term of the integrand of (3.9) contributes a term $O(\beta)$.

We now verify the Ansatz that $\Xi_\pm^{(0)}$ given in (3.11) is not affected to $O(1)$ if we use $J_\mp^{(0)}$ instead of $J_\mp^{(-1)}$.

$$\Xi_\pm = \beta\int_{\xi_\pm}^{\xi_\mp}(\cos\pi\xi_\pm + \cos\pi\xi'_\mp)$$
$$\times[\mp\tfrac{1}{2} \pm \tfrac{1}{4}\beta\pi(\xi_\pm - \xi'_\mp)\sin\pi\xi'_\mp]\,d\xi'_\mp + O(\beta), \quad (3.12)$$

where the quantity in brackets is the new value of $J_\mp$, calculated by substituting (3.11) into (2.19). This can be simplified to

$$\Xi_\pm = \pm\tfrac{1}{2}\beta(\xi_\pm - \xi_\mp)\cos\pi\xi_\pm \pm \tfrac{1}{4}\beta^2\pi\int_{\xi_\pm}^{\xi_\mp}(\xi_\pm - \xi'_\mp)$$
$$\times[\cos\pi\xi_\pm\sin\pi\xi'_\mp + \tfrac{1}{2}\sin 2\pi\xi'_\mp]\,d\xi'_\mp + O(\beta), \quad (3.13)$$

and one verifies (as will be demonstrated below) that the integral in (3.13) is $O(\beta)$. This validates the Ansatz in this order. $S^{(0)}$ is given by the pair of relations

$$\xi_\pm = x \pm t + \tfrac{1}{2}\beta(\xi_\pm - \xi_\mp)\cos\pi\xi_\pm + O(\beta), \quad (3.14)$$

and

$$r_\pm = \tfrac{1}{2}a\pi\cos\pi\xi_\pm + O(\beta). \quad (3.15)$$

The first space and time derivatives of the waveform are obtained directly in lowest order by taking the sum or difference of r_+ and r_- as defined in (3.15). Thus,

$$u = y_x(x, t) = r_+ + r_- + O(\epsilon)$$
$$= \tfrac{1}{2}a\pi(\cos\pi\xi_+ + \cos\pi\xi_-) + O(\epsilon)$$
$$= a\pi[\cos\tfrac{1}{2}\pi(\xi_+ + \xi_-)$$
$$\times\cos\tfrac{1}{2}\pi(\xi_+ - \xi_-)] + O(\epsilon), \quad (3.16)$$

and

$$v = y_t(x, t) = r_+ - r_- + O(\epsilon)$$
$$= \tfrac{1}{2}a\pi(\cos\pi\xi_+ - \cos\pi\xi_-) + O(\epsilon)$$
$$= -a\pi[\sin\tfrac{1}{2}\pi(\xi_+ + \xi_-)$$
$$\times\sin\tfrac{1}{2}\pi(\xi_+ - \xi_-)] + O(\epsilon). \quad (3.17)$$

One can rewrite (3.16) and (3.17) as the coupled pair of implicit relations

$$u = a\pi\cos\pi(x + \tfrac{1}{4}\epsilon tv)$$
$$\times\cos\pi(t + \tfrac{1}{4}\epsilon tu) + O(\epsilon),$$
$$v = -a\pi\sin\pi(x + \tfrac{1}{4}\epsilon tv)$$
$$\times\sin\pi(t + \tfrac{1}{4}\epsilon tu) + O(\epsilon), \quad (3.18)$$

if we substitute for $\xi_+ \pm \xi_-$ from (3.14).

If t is smaller than the breakdown time t_B, the implicit relations given in (3.18) will define unique functions u and v of x and t.

C. The Uniform First-Order Solution $S^{(1)}$

$J_\mp^{(0)}$ is obtained by substituting (3.11) into (2.19), or

$$J_\mp = \mp\tfrac{1}{2}[1 - \tfrac{1}{2}\beta\pi(\xi_\pm - \xi_\mp)\sin\pi\xi_\mp] + O(\beta). \quad (3.19)$$

This expression was used in (3.12) to verify the Ansatz in zero order. Hence $\Xi_\pm$ can be calculated from (2.17) or (2.20) as

$$\Xi_\pm = \mp\tfrac{1}{2}\beta\int_{\xi_\pm}^{\xi_\mp}[1 - \tfrac{1}{2}\beta\pi(\xi_\pm - \xi'_\mp)\sin\pi\xi'_\mp]$$
$$\times[\cos\pi\xi_\pm + \cos\pi\xi'_\mp]\,d\xi'_\mp + O(\beta^2), \quad (3.20)$$

or

$$\Xi_\pm = \pm\tfrac{1}{2}\beta(\xi_\pm - \xi_\mp)\cos\pi\xi_\pm \pm (\beta/2\pi)$$
$$\times(\sin\pi\xi_\pm - \sin\pi\xi_\mp) \mp (\tfrac{1}{16}\beta^2)(\xi_\pm - \xi_\mp)$$
$$\times(\cos 2\pi\xi_\mp + 4\cos\pi\xi_\pm\cos\pi\xi_\mp) + O(\beta^2). \quad (3.21)$$

Again, one can verify that $\Xi_\pm^{(1)}$ given in (3.21) is unaffected if we use $J_\mp^{(1)}$, derived from (3.21), instead of $J_\mp^{(0)}$. $S^{(1)}$ is given by (2.16) [with (3.21) for $\Xi_\pm$] and the relation

$$r_\pm = \tfrac{1}{2}a\pi(\cos\pi\xi_\pm + \beta\cos^2\pi\xi_\pm) + O(\beta^2). \quad (3.22)$$

One can verify that (3.22) with (3.21) is the correct solution to $O(\epsilon)$ by direct substitution in the characteristic equations (2.7), which are written as

$$(r_\pm)_t \mp [1 + \tfrac{1}{2}\epsilon(r_+ + r_-) + O(\epsilon^2)](r_\pm)_x = 0. \quad (3.23)$$

The verification is straightforward but lengthy, and only an outline of it will be given.

We take the t and x partial derivatives of $r_\pm(\xi_\pm)$, substitute into (3.23), divide through by $(\xi_\pm)_x$, and obtain

$$(\xi_\pm)_t/(\xi_\pm)_x = \pm[1 + \tfrac{1}{2}\epsilon(r_+ + r_-)] + O(\epsilon^2)$$
$$= \pm[1 + \beta(\cos\pi\xi_+ + \cos\pi\xi_-)] + O(\beta^2). \quad (3.24)$$

The left-hand side is evaluated by taking partial derivatives of (2.16) with $\Xi_\pm$ given by (3.21). We obtain

$$(\xi_\pm^{(1)})_t/(\xi_\pm^{(1)})_x$$
$$= \pm\{(1 \pm \mu_\pm \mp \nu_\pm)/[1 \pm (\mu_\pm + \nu_\pm)]\}, \quad (3.25)$$

where we have defined $\mu_\pm$ and $\nu_\pm$ through the relations

$$(\Xi_\pm)_x = \mu_\mp(\xi_\pm)_x + \nu_\pm(\xi_\mp)_x, \quad (3.26)$$

and have determined them as

$$\mu_\pm = \mp\beta \cos \pi\xi_\mp \mp \tfrac{1}{4}\pi\beta(\xi_\pm - \xi_\mp)(\sin \pi\xi_\mp) \times [2 - \beta \cos \pi\xi_\pm] + O(\beta^2),$$

$$\nu_\pm = \mp\tfrac{1}{2}\beta(\cos \pi\xi_\pm + \cos \pi\xi_\mp) \pm \tfrac{1}{4}\beta^2\pi(\xi_\pm - \xi_\mp) \times (\sin \pi\xi_\mp)[\cos \pi\xi_\pm + \cos \pi\xi_\mp] + O(\beta^2). \quad (3.27)$$

If we substitute (3.27) into the right-hand side of (3.25) and consistently omit terms $O(\beta^2)$, we obtain the right-hand side of (3.24), thus verifying the solution. The solution for this problem given previously in Sec. 6 of reference 9 appears to be valid to terms $O(\epsilon)$. However, not all the terms $O(\epsilon)$ were explicitly given in (6.8) and (6.9) of reference 9 [compare with (3.21)], and hence the solution there is uniformly valid only to $O(1)$.

4. BREAKDOWN PROPERTIES OF THE UNIFORM FIRST-ORDER SOLUTION $S^{(1)}$

A. Introduction

In Sec. 1 we remarked that the exact solution exhibited a breakdown after an elapsed time of $O(1/\epsilon)$; that is, the solution of the problem is not everywhere defined beyond the breakdown time t_B. We will now demonstrate that $S^{(1)}$ also exhibits this behavior and we will proceed to study the properties of the solution for space and time intervals near breakdown. Such a study is a prerequisite for resolving the breakdown when appropriate higher-derivative terms are included in the original partial differential equation.

B. Breakdown Time

To study the properties of the Jacobian (2.22), we evaluate the partial derivatives $(\Xi_\pm)_{\xi_\pm}$ and $(\Xi_\pm)_{\xi_\mp}$ as

$$(\Xi_\pm)_{\xi_\pm} = -\beta\pi\delta \sin \pi\xi_\pm \pm \beta \cos \pi\xi_\pm + (\tfrac{1}{2}\beta^2\pi\delta)(\sin \pi\xi_\pm \cos \pi\xi_\mp) + O(\beta^2),$$

$$(\Xi_\pm)_{\xi_\mp} = \tfrac{1}{2}\beta[\mp(\cos \pi\xi_\pm + \cos \pi\xi_\mp) + (\beta\pi\delta) \times \sin \pi\xi_\pm(\cos \pi\xi_\pm + \cos \pi\xi_\mp)] + O(\beta^2), \quad (4.1)$$

where

$$\delta = \pm\tfrac{1}{2}(\xi_\pm - \xi_\mp), \quad (4.2)$$

and where $\Xi_\pm$ is the first-order quantity given in (3.21). From (4.1) it follows that the last term of (2.22) is $O(\beta^2)$. Thus, $J = 0$ if either of the pair of relations obtained by substituting (4.1) into (2.23) holds:

$$1 - \beta \cos \pi\xi_\pm = \pm\beta\pi\delta \sin \pi\xi_\pm \times [-1 + \tfrac{1}{2}\beta \cos \pi\xi_\mp] + O(\beta^2). \quad (4.3)$$

We express t in terms of δ by substituting (3.21) and (4.2) into (2.18),

$$t = \delta[1 - \tfrac{1}{2}\beta(\cos \pi\xi_\pm + \cos \pi\xi_\mp)] + O(\beta) + O(\beta^2\delta). \quad (4.4)$$

Rearranging (4.4) we obtain

$$\delta = t[1 + \tfrac{1}{2}\beta(\cos \pi\xi_\pm + \cos \pi\xi_\mp)] + O(\beta), \quad (4.5)$$

since terms $O(\beta^2\delta)$ are *a fortiori* $O(\beta)$. Substituting into (4.3) yields

$$T^{-1} = \mp[\sin \pi\xi_\pm + \tfrac{1}{4}\beta \times \sin 2\pi\xi_\pm]/[1 - \beta \cos \pi\xi_\pm] + O(\beta^2), \quad (4.6)$$

or

$$T^{-1} = \mp\sin \pi\xi_\pm \mp (\tfrac{3}{4}\beta) \sin 2\pi\xi_\pm + O(\beta^2), \quad (4.7)$$

where

$$T^{-1} = (\beta\pi t)^{-1} \quad (4.8)$$

is a monotonically decreasing function of t. Equation (4.3) will be first satisfied at the points which maximize the right side of (4.7). Taking the derivative of the right side of (4.7) with respect to $\xi_\pm$ and setting it equal to zero, we obtain the pair of relations

$$\cos \pi\xi_\pm = -(\tfrac{3}{2}\beta) \cos 2\pi\xi_\pm + O(\beta^2), \quad (4.9)$$

which are obviously satisfied by

$$\pi\xi_\pm = \mp(\tfrac{1}{2}\pi - \tfrac{3}{2}\beta) + O(\beta^2), \quad (4.10)$$

where the signs have been chosen to make T^{-1} a positive number. If (4.10) is substituted into (4.7) we find that

$$T^{-1} = 1 + O(\beta^2), \quad (4.11)$$

or that the breakdown time is given by

$$t_B = (\beta\pi)^{-1} + O(\beta) = 4/(\epsilon a\pi^2) + O(\beta). \quad (4.12)$$

We see that the expression for the initial breakdown point $\pi\xi_\pm^{(1)}$ differs from $\mp\frac{1}{2}\pi$ by $O(\beta)$, whereas the normalized breakdown time T_B differs from 1 by $O(\beta^2)$.

C. Multivaluedness of the Solution After Breakdown

We now investigate the properties of the breakdown region in the physical plane. We will show that at breakdown $\xi_\pm^{(1)}(x, t)$ goes from a single-valued function to a triple-valued function, so that

the quantities $y_t^{(1)}(x, t)$ or $y_x^{(1)}(x, t)$ behave the same way. If we multiply (3.14) by π and introduce the variables

$$q_\pm = \tfrac{1}{2}\pi \pm \pi\xi_\pm, \tag{4.13}$$

and

$$\theta_\pm = \tfrac{1}{2}\pi \pm \pi(x \pm t), \tag{4.14}$$

we obtain the expression

$$q_\pm = \theta_\pm + \beta\pi\delta \sin q_\pm - \tfrac{1}{2}\beta(\cos q_\pm + \cos q_\mp)$$
$$\mp (\tfrac{1}{8}\beta^2\pi\delta)(\cos 2q_\mp + 4 \sin q_\pm \sin q_\mp) + O(\beta^2). \tag{4.15}$$

If δ is obtained from (4.5) with (4.12), and the result substituted into (4.15), we obtain

$$q_\pm = \theta_\pm + T(\sin q_\pm + \tfrac{1}{2}\beta \sin^2 q_\pm)$$
$$- \tfrac{1}{2}\beta \cos q_\pm + \beta f_\mp + O(\beta^2), \tag{4.16}$$

where

$$f_\mp = \tfrac{1}{2}(-\cos q_\mp + \tfrac{1}{4}T \cos 2q_\mp). \tag{4.17}$$

If we take

$$\tilde{\theta}_\pm = \theta_\pm + \tfrac{1}{6}\beta(\tfrac{1}{2} + f_\mp) = O(\beta), \tag{4.18}$$

$$T = 1 + \sigma, \tag{4.19}$$

and we expand the trigonometric functions in (4.16) which involve $q_\pm$ (but not $q_\mp$), we obtain the cubic equation

$$q_\pm^3 - \tfrac{9}{2}(\beta)q_\pm^2 - 6\sigma q - 6\tilde{\theta}_\pm + O(\beta^2) = 0, \tag{4.20}$$

where we have omitted terms $O(\sigma q_\pm^3)$, $O(q_\pm^4)$, and $O(\beta\sigma q_\pm^2)$. We simplify the form of the cubic equation by introducing the transformation suggested in the breakdown calculation (4.10), namely

$$q_\pm = \tilde{q}_\pm + \tfrac{3}{2}\beta, \tag{4.21}$$

and also by replacing σ by

$$\sigma = \tilde{\sigma} - \tfrac{9}{8}\beta^2. \tag{4.22}$$

The result of these substitutions is

$$\tfrac{1}{3}\tilde{q}_\pm^3 - 2\tilde{\sigma}\tilde{q}_\pm - 2\tilde{\theta}_\pm + O(\beta^2) = 0, \tag{4.23}$$

which takes the standard form of a cubic equation[19]

$$Y^3 - 3pY = -2, \tag{4.24}$$

if we define

$$Y = -(\tilde{q}_\pm)(3\tilde{\theta}_\pm)^{-\frac{1}{3}}, \qquad p = 2\tilde{\sigma}(3\tilde{\theta}_\pm)^{-\frac{2}{3}}. \tag{4.25}$$

Equation (4.24) has one real root for $p < 1$ and three real roots for $p > 1$. Thus, as $\tilde{\sigma}$ increases from negative to positive values, the solution of the partial differential equation goes from a single-valued function to a triple-valued function. The value of $\tilde{\sigma}$ corresponding to breakdown is obtained by setting $p = 1$ in (4.25), or

$$\tilde{\sigma}_B = \tfrac{1}{2}(3\tilde{\theta}_\pm)^{\frac{2}{3}} = O(\beta^{\frac{2}{3}}). \tag{4.26}$$

That is, with the above analysis, one can study the properties of the breakdown region provided $\sigma > O(\beta^{\frac{2}{3}})$.

D. The Behavior of the Singularity of $(r_\pm)_x$ Near Breakdown

Since y_x passes from a single-valued to a triple-valued function in the breakdown region, $y_{xx}(x, t)$ has a singularity in this region. The nature of the singularity is obtained most readily from the behavior of the first partial derivatives of $r_\pm$.

For example, from (2.7) the spatial derivative is given by

$$(r_\pm)_x = \pm\Phi^{-1}(r_\pm)_t = \pm(r_\pm)_t + O(\epsilon), \tag{4.27}$$

where we have taken $\Gamma = 1$ and $\Phi = 1 + O(\epsilon)$ as in Sec. 3A. Substituting $r_\pm$ from (3.22) into (4.27),

$$(r_\pm)_x = \mp(\tfrac{1}{2}a\pi^2 \sin \pi\xi_\pm)(\xi_\pm)_t + O(\epsilon), \tag{4.28}$$

where we have assumed that ξ_+ and ξ_- are dependent only on t through the relations given in (2.18). We now follow a procedure which is similar to that used for evaluating the Jacobian J in Sec. 2, namely we take time derivatives of the left and right sides of (2.16) (treating x as a constant) and solve the resulting pair of equations to obtain

$$(\xi_\pm)_t = \pm J^{-1}[1 \pm (\Xi_\mp)_{\xi_\mp} \pm (\Xi_\pm)_{\xi_\mp}], \tag{4.29}$$

where J is given by (2.22). Thus,

$$(\xi_\pm)_t = \pm[1 \mp (\Xi_\pm)_{\xi_\mp} + O(\beta)]^{-1}$$
$$+ J^{-1}(\Xi_\pm)_{\xi_\mp} + O(\beta). \tag{4.30}$$

For our purposes we need only consider the first term of the right side of (4.30), which we rewrite as

$$(\xi_\pm)_t = [(\pm\xi_\pm - \Xi_\pm)_{\xi_\mp}]^{-1} + O(\beta) \tag{4.31}$$

or

$$(\xi_\pm)_t = \pm[(\tilde{q}_\pm - \pi\Xi_\pm)_{\tilde{q}_\mp}]^{-1} + O(\beta), \tag{4.32}$$

since $d(\pi\xi_\pm) = \pm d\tilde{q}_\pm$ by (4.13) and (4.21). The quantity in parentheses has been studied in Eqs. (4.15) and (4.16) and in expanded form in (4.23). Hence, taking the $\tilde{q}_\pm$ partial derivative of (4.23) and substituting into (4.32), we obtain

$$(\xi_\pm)_t = \pm[\tilde{q}_\pm^2 - 2\tilde{\sigma}]^{-1} + O(\beta),$$

[19] E. Jahnke and F. Emde, *Tables of Functions with Formulas and Curves* (Dover Publications, Inc., New York, 1945); see p. 20 at rear of book.

or

$$(\xi_\pm)_t = \pm[(3\tilde{\theta}_\pm)^{\frac{1}{2}}(Y^2 - p)]^{-1} + O(\beta), \qquad (4.33)$$

where we have eliminated $\tilde{q}_\pm$ and $\tilde{\sigma}$ by (4.25). The leading term of the right side of (4.33) is $O(\beta)^{\frac{1}{2}}$ since $\tilde{\theta}_\pm = O(\beta)$ and $(Y^2 - p) = O(1)$ when Y is a root of the cubic equation (4.24) with finite p. Substituting (4.33) into (4.21) and eliminating $\xi_\pm$ in favor of $\tilde{q}_\pm$, we obtain

$$(r_\pm)_x = \pm\tfrac{1}{2}a\pi^2[(3\tilde{\theta}_\pm)^{\frac{1}{2}}(Y^2 - p)]^{-1} \times \cos(\tilde{q}_\pm + \tfrac{3}{2}\beta) + O(\epsilon). \qquad (4.34)$$

Thus, in the breakdown region, $(r_\pm)_x$ varies essentially as $\pm[x \pm t + O(\beta)]^{-\frac{1}{2}}$.

5. SPECTRAL DECOMPOSITION OF THE $S^{(0)}$ WAVEFORM INTO SPATIAL FOURIER MODES

A. Introduction

We have emphasized the phenomenon of breakdown as inherent in those wave systems whose nonlinear terms are functions of the derivatives of the dependent variable. Another property of nonlinear systems is the flow of energy between the normal modes of oscillation of the corresponding linear system, the rate of flow being governed by the strength of the nonlinear coupling constant. To study this aspect of nonlinear wave systems, we will decompose the $S^{(0)}$ waveform into spatial Fourier modes. We can then compare our results directly with the FPU computations, which are given in the form of modal energy vs time.

It should be noted that we will be taking the Fourier decomposition of a periodic function given implicitly. Before breakdown this is a well-behaved function (continuous and possessing all derivatives), and equals the sum of its Fourier series. However, as t passes through t_B, y_x becomes infinitely steep and then multivalued. The spatial region over which y_x (as well as y) is multivalued increases with t. The multivalued waveform has no physical meaning for times $t > t_B$, and so the formal Fourier coefficients which we obtain are irrelevant although they depend on the parameter ϵt continuously and analytically at and beyond ϵt_B.

The computation of a Fourier coefficient is a smoothing process, since it forms a weighted average in a region where the curve is multivalued. For example, assume that the curve $f(x)$ is multivalued in the region (a, b), as shown in Fig. 1. The Fourier sine coefficients over the interval (0, 1) are

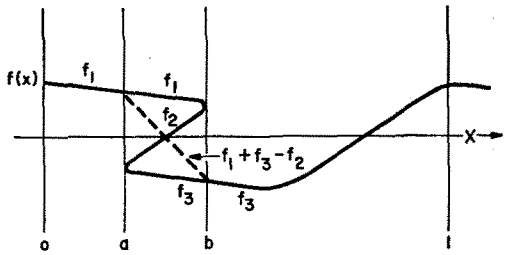


FIG. 1. Fourier decomposition of a multivalued function.

$$\tfrac{1}{2}B_n = \int_0^1 f(x)\sin\pi nx\,dx = \int_0^b f_1\sin\pi nx\,dx + \int_b^a f_2\sin\pi nx\,dx + \int_a^1 f_3\sin\pi nx\,dx = \int_0^a f_1\sin\pi nx\,dx + \int_a^b (f_1 + f_3 - f_2)\sin\pi nx\,dx + \int_b^1 f_3\sin\pi nx\,dx,$$

where the middle integral is the weighted average referred to above. If one sums the Fourier series determined by the coefficients B_n, then one obtains a single-valued function (the dashed line) with amplitude $f = f_1 + f_3 - f_2$ in the region (a, b).

B. Procedure for Decomposing into Spatial Fourier Modes

The total energy contained in the string vibrations is conserved. For the system described by (3.1), it is given by

$$E = \tfrac{1}{2}\int_{-1}^{1} dx[y_t^2 + y_x^2 + \tfrac{1}{3}\epsilon y_x^3], \qquad (5.1)$$

where the quantity in brackets is twice the energy density of the system.

As ϵ is a small number in the FPU calculations (smaller than $\frac{1}{32}$), and y_x has a maximum value of $a\pi$, we can consistently neglect those terms in the energy expression which appear with the explicit coefficient ϵ. Our results, for comparison with FPU, will be affected by no more than the thickness of a graph line.

Thus we will calculate

$$E^{(0)} = \tfrac{1}{2}\int_{-1}^{1} dx\{[y_t^{(0)}]^2 + [y_x^{(0)}]^2\}, \qquad (5.2)$$

always omitting terms which appear with the coefficient ϵ. Now, $y_t^{(0)} = v^{(0)}$ and $y_x^{(0)} = u^{(0)}$ are uniquely determined by (3.18), from which one verifies by inspection that u is even and v odd

under spatial inversion. Thus, we can write

$$u^{(0)} = \sum_{n=1}^{\infty} A_n \cos \pi n x, \quad v^{(0)} = \sum_{n=1}^{\infty} B_n \sin \pi n x, \tag{5.3}$$

where the coefficients A_n and B_n are functions of time only and are defined by

$$A_n = \int_{-1}^{1} u^{(0)} \cos \pi n x' \, dx', \qquad n \geq 1; \tag{5.4}$$

and

$$B_n = \int_{-1}^{1} v^{(0)} \sin \pi n x' \, dx', \qquad n \geq 1. \tag{5.5}$$

It will be seen that $A_0 = 0$.

It is convenient to replace the dummy variable x' by ξ_+ or ξ_-, for x can be written as an explicit function of ξ_+ and ξ_-. This is exhibited in (2.16) where t is held fixed throughout. To lowest order, we use (3.14) and replace $\pm\frac{1}{2}(\xi_\pm - \xi_\mp)$ by t. Thus,

$$\pi x' = \pi(\xi_\pm \mp t) \mp T \cos \pi \xi_\pm, \tag{5.6}$$

and

$$\pi \, dx' = \pi(1 \pm T \sin \pi \xi_\pm) \, d\xi_+, \tag{5.7}$$

where T is defined by (4.8). By (5.6), the limits of integration in (5.4) and (5.5) remain the same. Furthermore, $r_\pm$, u, and v are periodic functions of $\xi_\pm$ with period 2.

We now substitute into (5.4) and (5.5) the expressions for u and v given in (3.16) and (3.17),

$$A_n = A_n^+ + A_n^- \quad \text{and} \quad B_n = B_n^+ - B_n^-, \tag{5.8}$$

where

$$A_n^\pm = \int_{-1}^{1} r_\pm \cos \pi n x' \, dx'$$

$$= \tfrac{1}{2} a\pi \int_{-1}^{1} \cos \pi \xi (C_\mp)_n (1 \pm T \sin \pi \xi) \, d\xi, \tag{5.9}$$

and

$$B_n^\pm = \int_{-1}^{1} r_\pm \sin \pi n x' \, dx'$$

$$= \tfrac{1}{2} a\pi \int_{-1}^{1} \cos \pi \xi (S_\mp)_n (1 \pm T \sin \pi \xi) \, d\xi. \tag{5.10}$$

That is, the upper and lower signs on A and B indicate the contribution to these coefficients from r_+ and r_-, respectively. For convenience, we have introduced

$$(C_\mp)_n = \cos \pi n x' = \cos [\pi n(\xi \mp t) \mp \zeta],$$

$$(S_\mp)_n = \sin \pi n x' = \sin [\pi n(\xi \mp t) \mp \zeta], \tag{5.11}$$

where

$$\zeta = nT \cos \pi \xi. \tag{5.12}$$

Note that $A_0^\pm = 0$, because $(C_\mp)_0 = 1$.

After some computation, we find for each $n \geq 1$ that

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = (2a\pi) \left\{ \frac{J_n(nT)}{nT} \right\} \times \begin{bmatrix} \sin \frac{1}{2}\pi n & \cos \frac{1}{2}\pi n \\ \cos \frac{1}{2}\pi n & -\sin \frac{1}{2}\pi n \end{bmatrix} \begin{bmatrix} \cos \pi n t \\ \sin \pi n t \end{bmatrix}, \tag{5.13}$$

where matrix multiplication is understood. To arrive at this compact form, we have expanded $(C_\mp)_n$ and $(S_\mp)_n$ in infinite trigonometric series with Bessel function coefficients and derived the identities

$$2[J_n(nT)/nT](\sin \tfrac{1}{2}\pi n) \equiv \int_{-1}^{1} d\xi (\cos \pi \xi \cos \pi n \xi \cos \zeta) + \tfrac{1}{2} T \int_{-1}^{1} d\xi (\sin 2\pi \xi \sin \pi n \xi \sin \zeta), \tag{5.14}$$

and

$$2[J_n(nT)/nT](\cos \tfrac{1}{2}\pi n) \equiv -\int_{-1}^{1} d\xi (\cos \pi \xi \cos \pi n \xi \sin \zeta) + \tfrac{1}{2} T \int_{-1}^{1} d\xi (\sin 2\pi \xi \sin \pi n \xi \cos \zeta). \tag{5.15}$$

This is discussed in detail in the appendix.

The modal energy is obtained by substituting (5.3) into (5.2) and carrying out the integrations. Thus,

$$E^{(0)} = \tfrac{1}{2} \sum_{n=1}^{\infty} (A_n^2 + B_n^2) = \sum_{n=1}^{\infty} E_n^{(0)}, \tag{5.16}$$

where

$$E_n^{(0)} = 2[a\pi J_n(nT)/nT]^2, \qquad n \geq 1. \tag{5.17}$$

$E_n^{(0)}$ is the energy in mode n for a string of length 2, and is twice the value obtained in the FPU computations. By substituting the initial conditions (3.2) into (5.1), one finds the total energy residing in the string at $t = 0$, as $E|_0 = \frac{1}{2}a^2\pi^2$. The series for the total energy given in (5.16) is a particular case of the Kapteyn series of the second kind[20] and has been summed by G. A. Schott:

$$E^{(0)} = 2a^2\pi^2 \sum_{n=1}^{\infty} [J_n(nT)/(nT)]^2 = \tfrac{1}{2}a^2\pi^2 = E|_0,$$

[20] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1958). The Kapteyn series are discussed in Chap. 17. Equation 1, Sec. 17.6 gives the summation formula used.

for $T \leq 1$. Thus, the total energy calculated from the $S^{(0)}$ solutions is a conserved quantity and is equal to the initial total energy.

Table I lists properties of the function $J_n(nT)/(nT)$, which has its first maximum (for $n > 1$) at $t = T_n$.

Breakdown corresponds to $T = 1$. All modes greater than the first (mode 1 decreases monotonically) increase monotonically in energy during the time interval in which the analysis is meaningful. They have maxima in the region just beyond breakdown.

C. Behavior of the Modal Amplitude Function Near Breakdown

It is interesting to consider how the occurrence of breakdown is manifest in the properties of the modal amplitude functions given in (5.13). The singular nature (or sharpness) of u or v will be characterized by the behavior of $\lim_{n\to\infty} [J_n(n)/n]$, for breakdown is associated with $T = 1$. We apply Watson's formulas[21] in this transitional region

$x < n$:

$$J_n(x) = (\pi 3^{\frac{1}{2}})^{-1} w e^{n\alpha} K_{\frac{1}{3}}(nw^3/3) + O(n^{-1}), \tag{5.18}$$

$x > n$:

$$J_n(x) = (w/3^{\frac{1}{2}})[J_{\frac{1}{3}}(\tfrac{1}{3}nw^3)\cos\Delta - Y_{\frac{1}{3}}(\tfrac{1}{3}nw^3)\sin\Delta] + O(n^{-1}), \tag{5.19}$$

where

$$\Delta = n(w - \tfrac{1}{3}w^3 - \tan^{-1} w) + \tfrac{1}{6}\pi = -(\tfrac{1}{5}nw^5)[1 + \tfrac{5}{7}w^2 + O(w^4)] + \tfrac{1}{6}\pi, \tag{5.20}$$

$$\alpha = n(w + \tfrac{1}{3}w^3 - \tanh^{-1} w) = -(\tfrac{1}{5}nw^5)[1 - \tfrac{5}{7}w^2 + O(w^4)], \tag{5.21}$$

$$w = |(x^2/n^2) - 1|^{\frac{1}{2}}. \tag{5.22}$$

According to the calculation of the breakdown time given in (4.11), it is appropriate and convenient to replace T by

$$T = (1 \pm \rho^2)^{\frac{1}{2}} = 1 \pm \tfrac{1}{2}\rho^2 + O(\rho^4), \tag{5.23}$$

for the first correction to T is $O(\beta^2)$. The upper and lower signs correspond to after and before breakdown, respectively. If we substitute $x = nT = n(1 \pm \rho^2)^{\frac{1}{2}}$ in (5.22), we find in both cases

$$w = |\rho|. \tag{5.24}$$

Thus, $w = \rho$ is treated as a *positive* and *small* number in (5.18) and (5.19), which describe the situation before and after breakdown, respectively. If $\rho^2 \ll 1$, then

$$\Delta - \tfrac{1}{6}\pi = \alpha = -\tfrac{1}{5}n\rho^5 + O(\rho^7), \tag{5.25}$$

and we obtain from (5.18) and (5.19) the expressions

before:

$$J_n(nT) = (\pi 3^{\frac{1}{2}})^{-1}\rho \times \exp(-\tfrac{1}{5}n^2\rho^5)K_{\frac{1}{3}}(\tfrac{1}{3}n\rho^3) + O(1/n); \tag{5.26}$$

after:

$$J_n(nT) = (\rho 3^{-\frac{1}{2}})[J_{\frac{1}{3}}(\tfrac{1}{3}n\rho^3)\cos\Delta - Y_{\frac{1}{3}}(\tfrac{1}{3}n\rho^3)\sin\Delta] + O(1/n). \tag{5.27}$$

Now we take the limit $n\rho^3 \to \infty$; that is, the closer we are to the breakdown time, the larger must the mode number n be for our conclusion to be valid. Using the asymptotic forms of the Bessel functions we obtain

before:

$$J_n(nT) = (n\rho^5/18\pi)^{\frac{1}{2}} \exp(-\tfrac{1}{5}n^2\rho^5) + O(1/n); \tag{5.28}$$

after:

$$J_n(nT) = (2/\pi n\rho)^{\frac{1}{2}} \cos(\tfrac{1}{3}n\rho^3 - \tfrac{1}{4}\pi) + O(1/n). \tag{5.29}$$

Thus, before breakdown the maximum value of the modal amplitude function $J_n(nT)/nT$ decreases exponentially, predominantly as $\exp(-\frac{1}{5}n^2\rho^5)$. After breakdown it decreases algebraically as $(n^3\rho)^{-\frac{1}{2}}$. Hence, although the Fourier series for $y_x^{(0)}(x, t)$ is absolutely convergent, its first space derivative is not. That is,

$$y_{xx}^{(0)}(x, t) = \frac{\partial}{\partial x}\sum_{n=1}^{\infty} A_n \cos\pi nx = -\pi \sum_{n=1}^{\infty} nA_n \sin\pi nx \propto \sum_{n=1}^{\infty} (n\rho)^{-\frac{1}{2}} \cos(n\rho^3 - \tfrac{1}{4}\pi)\sin\pi nx. \tag{5.30}$$

D. Comparison with the Numerical Computations

In Figs. 2 through 5 we have compared the analytical result (5.17) with the numerical computations for strings composed of $N = 16$, 32, and 64

TABLE I. Numerical properties of the modal amplitude and modal energy functions.

Mode n	T_M	nT_M	$J_n(nT_M)/(nT_M)$	Max $[E_n^{(0)}/E_1 \mid_{t=0}]$
1	0	0	1/2	1
2	1.150	2.300	0.1799	0.1295
3	1.203	3.611	0.1105	0.0488
4	1.203	4.812	0.0786	0.0247
5	1.192	5.963	0.0607	0.0147

[21] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. II, paragraph 7.13, Eqs. 28 and 30.

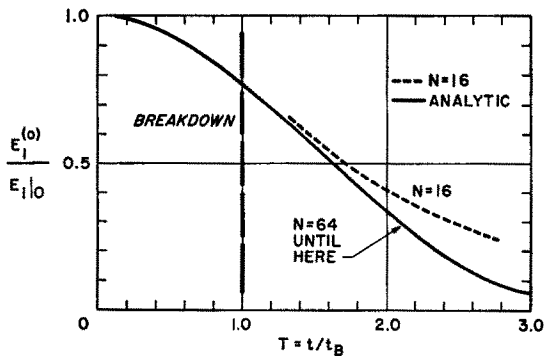


FIG. 2. The first-mode energies for the computational and analytic $S^{(0)}$ solutions.

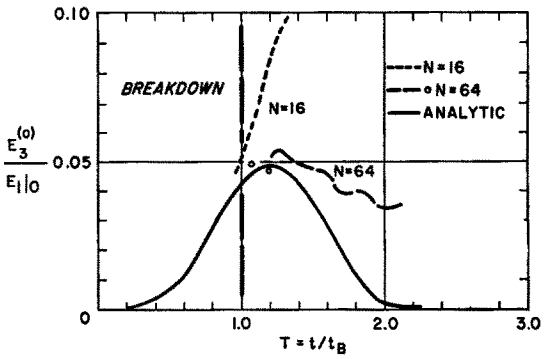


FIG. 4. The third-mode energies for the computational and analytic $S^{(0)}$ solutions.

particles.[22] The larger N is, the less "granular" or more nearly continuous is the string.

Fig. 2 compares the $N = 16$ and $N = 64$ first-mode energies with the analytic results. Up until breakdown the computations coincide with the analytic solutions (solid line). After $T = 1$, the $N = 64$ case continues to coincide with the analytic curve (up until the point shown), whereas the $N = 16$ case digresses. This digression is more vividly shown in Fig. 3, where the energies in mode 2 are plotted against normalized time. After breakdown, the large N (less granular) solutions are in better agreement with the analytic solution than

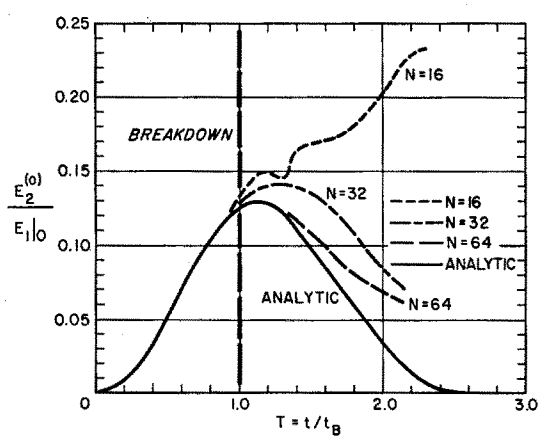


FIG. 3. The second-mode energies for the computational and analytic $S^{(0)}$ solutions.

[22] The original FPU report, Ref. 7, contained computations for the $N = 32$ string. The results for the $N = 16$ and $N = 64$ strings were obtained in 1961 by James L. Tuck of Los Alamos Scientific Laboratory, and were made available to the authors in April of 1962. It should be emphasized that the smallness parameter ϵ used in the present analysis is related to the parameters α and N of the numerical computation by

$$\epsilon = 2\alpha/N.$$

The nonlinear coupling constant α was taken to be $\frac{1}{4}$, and N is the number of particles on the string. This relation implies that the amount of time (or computation steps) required to reach breakdown increased linearly with N, because $t_B \propto (1/\epsilon) \propto N$.

are the small N solutions. The reason for this presumably lies in the behavior of the multivalued region. The percentage of the total string length occupied by the multivalued region increases more slowly for the large N string.

Figures 4 and 5 are similar comparisons of the energies of modes 3 and 5. For the higher mode numbers, the energies of the more granular strings diverge rapidly from the analytic solution—even before breakdown. This is because the properties of the normal modes of a beaded string differ from those of a continuous string—for example, the normal-mode frequencies of the former are not integral multiples of the fundamental frequency.

6. CONCLUSIONS AND RECOMMENDATIONS PERTINENT TO RESOLVING BREAKDOWN

In this paper, from the start, we have addressed ourselves to the study of nonlinear hyperbolic partial differential equations. These are typical of a wide range of physical phenomena and were mentioned as the "lowest-order continuum limit" of problems studied at the Los Alamos Scientific Laboratory on high-speed computing machinery.

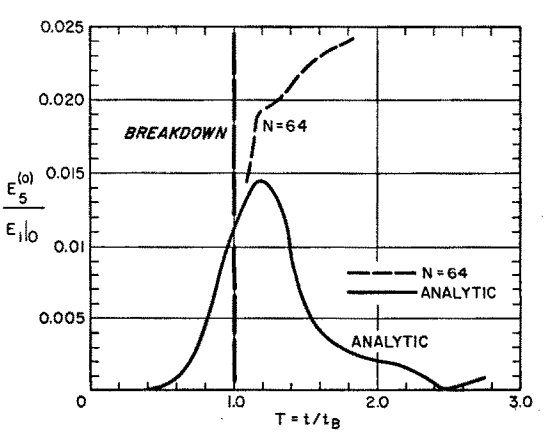


FIG. 5. The fifth-mode energies for the computational and analytic $S^{(0)}$ solutions.

A perturbation procedure was presented for treating these nonlinear wave phenomena, and it was shown that the solutions broke down for times greater than t_B, which is proportional to the ratio of the number of particles on the string to the nonlinear coupling constant. A Fourier decomposition of the analytic waveform gave good agreement with the Los Alamos computations.

Breakdown is associated with the development of a singularity in y_{xx}. This implies that, immediately before breakdown, the second and higher spatial derivatives will become very large. To study what happens around and *after* t_B one must include the higher spatial derivatives that were omitted in taking the lowest continuum limit. This heuristic attitude is similar to that in hydrodynamics, where one adds viscosity (a higher-derivative term) into the equations of motion and thereby prevents the solution from breaking down. Instead, one has a region of steep shock front whose thickness is related to the coefficient of viscosity.

The present problem is qualitatively different from the hydrodynamical one because of the nature of the higher-derivative terms.[5] This is readily seen if one takes the continuum limit of one of the models studied on the computing machine. Equation (6.1) presents the model for the "quadratic" nonlinearity,

$$(1/k^2)\delta_j^2 y_i^j = (1/h^2)\delta_i^2 y_i^j[1 + \alpha(y_{i+1}^j - y_{i-1}^j)], \qquad (6.1)$$

where δ^2 is the central second-difference operator. The subscript i on δ corresponds to spatial differences and j to temporal differences. Thus, the spatial central difference operator is defined by

$$\delta_i^2 y_i^j = y_{i+1}^j - 2y_i^j + y_{i-1}^j. \qquad (6.2)$$

If we assume that spatial and temporal differences can be expanded in Taylor series, then

$$y_{i\pm1} - y_i = \{\pm h y_x + (\tfrac{1}{2}h^2)y_{xx} \pm (h^3/3!)y_{xxx} + \cdots\}_{x=x_i}, \qquad (6.3)$$

where $h = 1/N$ is the particle spacing.

A similar relation holds for temporal differences, with h replaced by k. Expanding the terms in (6.1) in this manner yields

$$y_{tt} - (1 + \epsilon y_x)y_{xx} = (\tfrac{1}{2}h^2)[(1 + \epsilon y_x)y_{xxxx} + 2y_{xx}y_{xxx} - (k^2/h^2)y_{tttt}] + O(h^4) + O(k^4), \qquad (6.4)$$

where $\epsilon = 2\alpha/N$. For a typical case $N = 32$ and $\alpha = \frac{1}{4}$, $\epsilon = \frac{1}{64}$ and $\frac{1}{12}h^2 = 1/[3(64)^2]$. Thus, at $t = 0$ where the first spatial and temporal derivatives were $a\pi(a = 1)$, we were justified in taking the continuum limit $h = 1/N = 0$ and ϵ small. However, for large times, this approximation is no longer valid in certain spatial regions and we must include *some* or *all* of the terms given on the right-hand side of (6.4). The terms included will be those which are largest in magnitude.

7. ACKNOWLEDGMENTS

The work presented here was begun under the auspices of the Atomic Energy Commission when both authors were resident at the Princeton University Plasma Physics Laboratory. Since October 1961, the work of one of the authors (N.J.Z.) has been sponsored by the Bell Telephone Laboratories, Incorporated.

The authors wish to thank Dr. Stanislaw M. Ulam for his encouragement and enlightenment on various aspects of this problem, and Dr. James L. Tuck for making available to them recent computations performed at the Los Alamos Scientific Laboratory.

One of the authors (N.J.Z.) wishes to acknowledge helpful conversations with his colleagues at the Bell Telephone Laboratories, especially Dr. Yung-Chang Lee and Dr. James Alan Cochran.

The authors also acknowledge the patience and precision with which the various modifications of the work were rendered by Mrs. Nancy Campbell and her group at Whippany.

APPENDIX: CALCULATION OF THE FOURIER COEFFICIENTS

In this appendix we will carry out the details of the computation of the Fourier coefficients A_n and B_n given in (5.13). We begin with the definitions of $A_n^{\pm}$ and $B_n^{\pm}$ given in (5.9) and (5.10), substitute them into (5.8), and obtain

$$A_n = \tfrac{1}{2}a\left\{\int_{-\pi}^{\pi} dx \cos x[(C_+)_n + (C_-)_n] - \tfrac{1}{2}T\int_{-\pi}^{\pi} dx \sin 2x[(C_+)_n - (C_-)_n]\right\}, \qquad (A1)$$

$$B_n = \tfrac{1}{2}a\left\{\int_{-\pi}^{\pi} dx \cos x[-(S_+)_n + (S_-)_n] + \tfrac{1}{2}T\int_{-\pi}^{\pi} dx \sin 2x[(S_+)_n + (S_-)_n]\right\}, \qquad (A2)$$

where the dummy variable $\pi\xi$ of (5.9) and (5.10) has been replaced by x. If we expand the trigonometric functions $(C_\mp)_n$ and $(S_\mp)_n$ defined by (5.11) and (5.12), we can express the bracketed quantities in (A1) and (A2) as

$$[(C_+)_n + (C_-)_n]/\cos nx = [(S_+)_n + (S_-)_n]/\sin nx = 2(\cos \pi n t \cos \xi - \sin \pi n t \sin \xi), \qquad (A3)$$

and

$$[(C_+)_n - (C_-)_n]/\sin nx = [-(S_+)_n + (S_-)_n]/\cos nx$$
$$= -2(\cos \pi nt \sin \zeta + \sin \pi nt \cos \zeta), \quad \text{(A4)}$$

where

$$x = \pi\xi \quad \text{and} \quad \zeta = nT \cos x.$$

Equations (A3) and (A4) are substituted into (A1) and (A2) and the terms regrouped, to yield

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = a\pi \begin{bmatrix} a_n & b_n \\ b_n & -a_n \end{bmatrix} \begin{bmatrix} \cos \pi nt \\ \sin \pi nt \end{bmatrix}, \quad \text{(A5)}$$

where matrix multiplication is understood. The coefficients are defined by the relations

$$a_n = \text{Re}\,(I_{cn}) + \tfrac{1}{2}T\,\text{Im}\,(I_{sn}), \quad \text{(A6)}$$

$$b_n = -\text{Im}\,(I_{cn}) + \tfrac{1}{2}T\,\text{Re}\,(I_{sn}), \quad \text{(A7)}$$

where

$$I_{cn} = \pi^{-1} \int_{-\pi}^{\pi} dx(e^{i\zeta} \cos x \cos nx), \quad \text{(A8)}$$

$$I_{sn} = \pi^{-1} \int_{-\pi}^{\pi} dx(e^{i\zeta} \sin 2x \sin nx). \quad \text{(A9)}$$

We now use a modified form of the Jacobi series,[23] and express $e^{i\zeta}$ as a trigonometric series with Bessel function coefficients, namely

$$e^{i\zeta} = \exp(inT \cos x)$$
$$= \sum_{m=-\infty}^{\infty} J_m(nT) \exp[im(\tfrac{1}{2}\pi + x)]. \quad \text{(A10)}$$

If (A10) is substituted into (A9), the integrals can be evaluated easily because of the orthogonal properties of the exponential functions. We obtain

$$I_{cn} = i^{n+1}[J_{n+1}(nT) - J_{n-1}(nT)] = [-\sin \tfrac{1}{2}\pi n$$
$$+ i \cos \tfrac{1}{2}\pi n][J_{n+1}(nT) - J_{n-1}(nT)], \quad \text{(A11)}$$

$$I_{sn} = -i^{n+2}[J_{n+2}(nT) - J_{n-2}(nT)] = [\cos \tfrac{1}{2}\pi n$$
$$+ i \sin \tfrac{1}{2}\pi n][J_{n+2}(nT) - J_{n-2}(nT)]. \quad \text{(A12)}$$

If (A11) and (A12) are substituted into (A7) and (A8), we find that

$$a_n/\sin \tfrac{1}{2}\pi n = b_n/\cos \tfrac{1}{2}\pi n$$
$$= \{-J_{n+1}(nT) + J_{n-1}(nT)$$
$$+ \tfrac{1}{2}T[J_{n+2}(nT) - J_{n-2}(nT)]\}$$
$$= 2J_n(nT)/(nT), \quad \text{(A13)}$$

and therefore finally obtain (5.13).

[23] G. N. Watson, Ref. 20. The form given in Eq. (A10) is obtained from the first expression given by Watson in Sec. 2.22.

Generalization of Laplace's Expansion to Arbitrary Powers and Functions of the Distance between Two Points*

R. A. SACK
Department of Mathematics, Royal College of Advanced Technology, Salford, England
(Received 16 August 1963)

In analogy to Laplace's expansion, an arbitrary power r^n of the distance r between two points $(r_1, \vartheta_1, \varphi_1)$ and $(r_2, \vartheta_2, \varphi_2)$ is expanded in terms of Legendre polynomials of $\cos \vartheta_{12}$. The coefficients are homogeneous functions of r_1 and r_2 of degree n satisfying simple differential equations; they are solved in terms of Gauss' hypergeometric functions of the variable $(r_</r_>)^2$. The transformation theory of hypergeometric functions is applied to describe the nature of the singularities as r_1 tends to r_2 and of the analytic continuation of the functions past these singularities. Expressions symmetric in r_1 and r_2 are obtained by quadratic transformations; for $n = -1$ and $n = -2$; one of these has previously been given by Fontana. Some three-term recurrence relations between the radial functions are established, and the expressions for the logarithm and the inverse square of the distance are discussed in detail. For arbitrary analytic functions $f(r)$, three analogous expansions are derived; the radial dependence involves spherical Bessel functions of $(r_<\partial/\partial r_>)$ of of related operators acting on $f(r_>)$, $f(r_1 + r_2)$ or $f[(r_1^2 + r_2^2)^{\frac{1}{2}}]$.

1. INTRODUCTION

THE inverse distance r^{-1} between two points Q_1 and Q_2, specified by the polar coordinates $(r_1, \vartheta_1, \varphi_1)$ and $(r_2, \vartheta_2, \varphi_2)$ with reference to a common origin O, is given by the well-known Laplace expansion

$$r^{-1} = r_>^{-1} \sum_{l=0}^{\infty} (r_</r_>)^l P_l(\cos \vartheta_{12}), \tag{1}$$

where

$$r_< = \min(r_1, r_2), \qquad r_> = \max(r_1, r_2), \tag{2}$$

$$\cos \vartheta_{12} = \cos \vartheta_1 \cos \vartheta_2 + \sin \vartheta_1 \sin \vartheta_2 \cos(\varphi_1 - \varphi_2), \tag{3}$$

and the $P_l(x)$ are the Legendre polynomials. In many physical problems, the distance between Q_1 and Q_2 may be required to powers other than the inverse first, and an expansion analogous to (1) is required for such cases. One way of approaching the problem is to preserve the expansion in powers of $(r_</r_>)$; the expression

$$r^{-2\nu} = r_>^{-2\nu} \sum_{l=0}^{\infty} (r_</r_>)^l C_l^{\nu}(\cos \vartheta_{12}) \tag{4}$$

serves to define the angular dependence as Gegenbauer polynomials of the argument[1] (cf. B 3.15[2]); but for three-dimensional problems it is more convenient to preserve the dependence on the angles, and to redefine the dependence on the radii, and the writer is not aware that the corresponding expansion

$$V_n = r^n = \sum_l R_{nl}(r_1, r_2) P_l(\cos \vartheta_{12}) \tag{5}$$

has been given in the general case. If n is a positive even integer, V_n is the $\frac{1}{2}n$th power of

$$r^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \vartheta_{12}, \tag{6}$$

and the expansion (5) is a finite series terminating with $l = \frac{1}{2}n$; the form of the radial functions R_n is independent of the comparative values of r_1 and r_2. For odd positive values of n, recurrence relations based on (1) and (6) have occasionally been quoted; the expressions for $n = 1$ have been given explicitly by Jen.[3]

The purpose of the present paper is to derive the explicit terms in the expansion (5) for the general case. For variations of the positions of the points Q_1 and Q_2, the function V_n appears as the solution of the partial differential equation

$$\nabla_1^2 V_n = \nabla_2^2 V_n = n(n+1) V_{n-2}; \tag{7}$$

the corresponding differential equation for the radial functions R_{nl} following from (5) and (7), together with simple additional conditions of dimensionality and continuity, are solved in Sec. 2 in terms of Gauss' hypergeometric function

$$F(\alpha, \beta; \gamma, x) = 1 + \sum_1^{\infty} \frac{(\alpha)_s (\beta)_s}{(\gamma)_s s!} x^s, \tag{8}$$

where

$$(\alpha)_0 = 1; \quad (\alpha)_s = \alpha(\alpha+1) \cdots (\alpha+s-1) = \Gamma(\alpha+s)/\Gamma(\alpha). \tag{9}$$

* This work was begun at the Laboratory of Molecular Structure and Spectra, University of Chicago, supported by Office of Naval Research Contract Nonr-2121(01), continued at Salford, and completed at the Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin, supported by National Aeronautics and Space Administration Grant NsG-275-62(4180).

[1] L. Gegenbauer, Wien. Sitzung. **70**, 6, 434 (1874); **75**, 891 (1877).

[2] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953). Sections and formulas in this work are directly referenced by the letter B.

[3] C. K. Jen, Phys. Rev. **43**, 540 (1933).

In Sec. 3, the extensive transformation theory of the hypergeometric function is applied to express the R_n in a variety of forms and to study their behavior, expecially in the asymptotic case $r_1 \to r_2$. The results obtained are asymmetric in $r_<$ and $r_>$, but by means of quadratic transformations can be expressed in several symmetric forms; one of these transformations has recently been derived by Fontana[4] on the basis of group-theoretical arguments.

In Sec. 4, Gauss' relations between contiguous hypergeometric functions are used to establish recurrence relations between the R_{nl}, and the case of the logarithm and the inverse square are discussed in greater detail in Sec. 5.

The results obtained in Sec. 3 are rewritten in Sec. 6 in a symbolic form, independent of the power n, but involving powers or functions of differential operators; this yields an expansion theorem for an arbitrary analytic function $f(r)$. The more general problem that the function depends on the relative orientation of Q_1 and Q_2 as well as on their distance are considered in a separate paper.

2. MATHEMATICAL DERIVATION

Substitution of (7) into (5) leads to

$$\frac{\partial^2 R_{nl}}{\partial r_1^2} + \frac{2}{r_1}\frac{\partial R_{nl}}{\partial r_1} - l(l+1)\frac{R_{nl}}{r_1^2} = \frac{\partial^2 R_{nl}}{\partial r_2^2} + \frac{2}{r}\frac{\partial R_{nl}}{\partial r_2} - l(l+1)\frac{R_{nl}}{r_2^2}. \tag{10}$$

Furthermore, the R_{nl} are homogeneous functions of degree n in the variables r_1 and r_2, and since V_n is a continuous function if $r_< = 0$, they must contain the factor $r_<^l$ so that

$$R_{nl}(r_1, r_2) = r_<^l r_>^{n-l} G_{nl}(r_</r_>), \tag{11}$$

where $G_{nl}(x)$ is an analytic function for $0 \leq x < 1$. Expressing G_n as a power series,

$$G_{nl}(r_</r_>) = \sum_s c_{nls}(r_</r_>)^s, \tag{12}$$

and substituting (10) into (11), we obtain the recurrence relations

$$(s+2)(2l+s+3)c_{n,l,s+2} = (n-2l-s)(n-s+1)c_{nls}. \tag{13}$$

The sequence of coefficients thus begins with $s = 0$, as the other possibility $s = -2l-1$ would violate the continuity condition, and hence $c_{nls} = 0$ for odd s, and for even $s = 2\nu$,

$$c_{n,l,2\nu} = \frac{(l-\frac{1}{2}n)_\nu(-\frac{1}{2}n-\frac{1}{2})_\nu}{(l+\frac{3}{2})_\nu \nu!} c_{nl0}, \tag{14}$$

where $(\alpha)_\nu$ is defined in (9). Hence, with the definition (8) for Gauss' hypergeometric function, (11), (12), and (14) yield

$$R_{nl}(r_1, r_2) = K(n, l) r_<^l r_>^{n-l} \times F(l-\tfrac{1}{2}n, -\tfrac{1}{2}-\tfrac{1}{2}n; l+\tfrac{3}{2}; r_<^2/r_>^2). \tag{15}$$

The coefficients $K(n, l)$ are most easily determined by considering the case $\vartheta_{12} = 0$ when all the $P_l(\cos\vartheta_{12}) = 1$:

$$V_n = |r_> - r_<|^n = r_>^n \sum_\lambda \binom{n}{\lambda}\left(\frac{-r_<}{r_>}\right)^\lambda; \tag{16}$$

comparison of the coefficients of $r_<^\lambda r_>^{n-\lambda}$ in (15) and (16) yields

$$\frac{n(n-1)\cdots(n-\lambda+1)}{\lambda!} = K(n,\lambda) + K(n,\lambda-2)\frac{(\lambda-2-\frac{1}{2}n)(-\frac{1}{2}-\frac{1}{2}n)}{\lambda-\frac{1}{2}} + K(n,\lambda-4)\frac{(\lambda-4-\frac{1}{2}n)_2(-\frac{1}{2}-\frac{1}{2}n)_2}{(\lambda-\frac{5}{2})_2 2!} + \cdots. \tag{17}$$

Considered as a function of n, the left-hand side is a polynomial of degree λ; it follows by induction that each $K(n, l)$ must be a polynomial in n of degree l at most.

Now for positive even n, the series (17) breaks off at $l = \frac{1}{2}n$, and conversely for any value of l, $K(n, l)$ vanishes for $n = 0, 2, \cdots 2l-2$. Hence it must be a multiple of $n(n-2)\cdots(n-2l+2)$ or of $(-\frac{1}{2}n)_l$, and since by virtue of (1) all $K(-1, l)$ are unity, the general solution is

$$K(n, l) = (-\tfrac{1}{2}n)_l/(\tfrac{1}{2})_l. \tag{18}$$

3. SOLUTION FOR THE RADIAL FUNCTIONS AND THEIR TRANSFORMATIONS

The Eqs. (15) and (18) show that radial functions R_{nl} in the expansion (5) are given as

$$R_{nl}(r_1, r_2) = \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} r_>^n \left(\frac{r_<}{r_>}\right)^l \times F\left(l-\tfrac{1}{2}n, -\tfrac{1}{2}-\tfrac{1}{2}n; l+\tfrac{3}{2}; \frac{r_<^2}{r_>^2}\right). \tag{19}$$

The hypergeometric functions (8) are finite series, i.e., they are polynomials in x, if either α or β is a negative integer or zero. This implies that, for all positive odd integer values of n, the series for R_{nl} break off, and if $n = -1$, they consist of the leading

[4] P. R. Fontana, J. Math. Phys. 2, 825 (1961).

term only, in agreement with (1). For positive even n, the series are finite for $l \leq \frac{1}{2}n$; for $l > \frac{1}{2}n$, the factor $(-\frac{1}{2}n)_l$ ensures that R_{nl} vanishes identically.

Of the numerous transformations of the hypergeometric function, the following are especially relevant in the present context [cf. (B 2.9.1, 2); (B 2.10.1, 2)]:

$$F(\alpha, \beta; \gamma; x)$$

$$= (1-x)^{\gamma-\alpha-\beta} F(\gamma-\alpha, \gamma-\beta; \gamma; x), \tag{20a}$$

$$= \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \times F(\alpha, \beta; \alpha+\beta-\gamma+1; 1-x) + \frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)} (1-x)^{\gamma-\alpha-\beta} \times F(\gamma-\alpha, \gamma-\beta; \gamma-\alpha-\beta+1; 1-x), \tag{20b}$$

$$= \frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)} (-x)^{-\alpha} \times F(\alpha, 1-\gamma+\alpha; 1-\beta+\alpha; x^{-1}) + \frac{\Gamma(\gamma)\Gamma(\alpha-\beta)}{\Gamma(\alpha)\Gamma(\gamma-\beta)} (-x)^{-\beta} \times F(\beta, 1-\gamma+\beta; 1-\alpha+\beta; x^{-1}). \tag{20c}$$

The first, if applied to (19), yields

$$R_{nl}(r_1, r_2) = \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} \frac{r_<^l (r_>^2 - r_<^2)^{n+2}}{r_>^{l+n+4}} \times F\left[l+2+\tfrac{1}{2}n, \tfrac{3}{2}+\tfrac{1}{2}n; l+\tfrac{3}{2}; \frac{r_<^2}{r_>^2}\right], \tag{21}$$

which shows that the functions F are invariant against the substitution $n \to -n-4$. Thus the coefficients R are rational functions of r_1 and r_2 for odd integer n whatever its sign, and also for negative even n as long as $l < \frac{1}{2}|n| - 1$, though in the latter case, the expansion (5) does not break off as with positive even n.

The transformation (20b) applied to (19) yields

$$R_{nl}(r_1, r_2) = \frac{2^{n+1}(l+\frac{1}{2})(-\frac{1}{2}n)_l}{(1+\frac{1}{2}n)_{l+1}} r_<^l r_>^{n-l} \times F\left[l-\tfrac{1}{2}n, -\tfrac{1}{2}-\tfrac{1}{2}n; -1-n; \frac{r_>^2 - r_<^2}{r_>^2}\right] - \frac{2l+1}{2^{n+3}(n+2)} \frac{r_<^l (r_>^2 - r_<^2)^{n+2}}{r_>^{l+n+4}} \times F\left[l+\tfrac{1}{2}n+2, \tfrac{3}{2}+\tfrac{1}{2}n; n+3, \frac{r_>^2 - r_<^2}{r_>^2}\right]. \tag{22}$$

Here the gamma products have been simplified with the use of (9) and Legendre's duplication formula (B 1.2.15)

$$\Gamma(2z) = 2^{2z-1}\pi^{-\frac{1}{2}}\Gamma(z)\Gamma(z+\tfrac{1}{2}). \tag{23}$$

The expansion (22) shows the nature of the branch point for fractional n as $r_<$ approaches $r_>$; we see that for $n \leq -2$, the individual functions R_{nl} are divergent, though they remain integrable as long as $n > -3$.

For integer n, (22) needs special interpretation since either one series contains terms with the indeterminate factor 0/0, or else both series possess infinite coefficients. In particular, if the function F in (19) represents a polynomial in $r_<^2/r_>^2$, it transforms into a polynomial in the variable $(r_>^2 - r_<^2)/r_>^2$; this corresponds to the terminating part of that series in (22) which has negative parameters; the terms of this series resume when the denominator in (8) also vanishes, a passage to the limit shows that the ratio 0/0 is to be interpreted as $\frac{1}{2}$, and the resumed terms exactly cancels the other series (22). On the other hand, for the nonterminating series R_{nl} in (19), at negative even n the infinities of the two series cancel out, leading to logarithmic terms in agreement with (B 2.10.12, 13).

The transformation (20c) when applied to (19) leads to

$$R_{nl} = \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} (-1)^{\frac{1}{2}n} \cos\tfrac{1}{2}n\pi \quad r_<^{n-l} r_>^l \times F\left(l-\tfrac{1}{2}n; -\tfrac{1}{2}-\tfrac{1}{2}n; \tfrac{3}{2}+l; \frac{r_>^2}{r_<^2}\right) + \frac{\Gamma(l+\frac{3}{2})\pi^{\frac{1}{2}}}{\Gamma(-\frac{1}{2}n)\Gamma(2+l+\frac{1}{2}n)} (-1)^{\frac{1}{2}(n+1)} r_<^{n+l+1} r_>^{-1-l} \times F\left(-1-l-\tfrac{1}{2}n, -\tfrac{1}{2}-\tfrac{1}{2}n; \tfrac{1}{2}-l; \frac{r_>^2}{r_<^2}\right), \tag{24}$$

the constant factor of the first series having been simplified by means of the relation (B 1.2.6)

$$\Gamma(z)\Gamma(1-z) = \pi/\sin\pi z. \tag{25}$$

Equation (24) shows the nature of the analytic continuation of R_{nl} from $r_1 < r_2$ to $r_1 > r_2$, or conversely. As expected, this agrees with the true expression (19) for $r_1 > r_2$ only if n is a nonnegative even integer; in this case, the second series in (24) has zero coefficient. For the nonterminating series R_{nl} in the case of negative even n, the second term in (24) has a purely imaginary coefficient of indeterminate sign; the true function (19) for $r_1 > r_2$ corresponds to the first term in (24) only, and is therefore not the analytic continuation of R_{nl} for

$r_1 < r_2$, but its Cauchy principal value with respect to the logarithmic singularity at $r_1 = r_2$.

The relations between the three parameters occurring in the hypergeometric function in (19) allow additional, quadratic transformations to be applied to the R_{nl}. Thus, application of (B2.11.34,36),

$$F(\alpha, \beta; \alpha - \beta + 1; x) = (1 + x)^{-\alpha} \times F[\tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha + \tfrac{1}{2}; \alpha - \beta + 1; 4x(1 + x)^{-2}] \quad (26a)$$

$$= (1 + x^{\frac{1}{2}})^{-2\alpha} F[\alpha, \alpha - \beta + \tfrac{1}{2}; 2\alpha - 2\beta + 1; \times 4x^{\frac{1}{2}}(1 + x^{\frac{1}{2}})^{-2}], \quad (26b)$$

to (19) leads to

$$R_{nl}(r_1, r_2) = \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} \frac{(r_1 r_2)^l}{(r_1^2 + r_2^2)^{l-\frac{1}{2}n}} \times F\left[\tfrac{1}{2}l - \tfrac{1}{2}n, \tfrac{1}{2}l - \tfrac{1}{4}n + \tfrac{1}{2}; \tfrac{3}{2} + l; \frac{4r_1^2 r_2^2}{(r_1^2 + r_2^2)^2}\right], \quad (27a)$$

$$= \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} \frac{(r_1 r_2)^l}{(r_1 + r_2)^{2l-n}} \times F\left[l - \tfrac{1}{2}n, 1 + l; 2 + 2l; \frac{4r_1 r_2}{(r_1 + r_2)^2}\right]. \quad (27b)$$

These expressions are completely symmetric in r_1 and r_2, the asymmetry in (19) in the two variables is related to the transformations inverse to (26) and (27) [cf. (B 2.11.6, 31)], which involve square roots which must be taken with a fixed sign. This leads to variables of the form

$$\frac{r_1^2 + r_2^2 - |r_1^2 - r_2^2|}{r_1^2 + r_2^2 + |r_1^2 - r_2^2|} \text{ and } \left[\frac{r_1 + r_2 - |r_1 - r_2|}{r_1 + r_2 + |r_1 - r_2|}\right]^2, \quad (28)$$

both of which equal $r_<^2/r_>^2$ of (2). Similar considerations apply to the factor outside the hypergeometric function. Fontana[4] has derived a formula equivalent to (27a) by group-theoretical methods, and given explicit expressions for $R_{-1,l}$ and $R_{-2,l}$ in terms of double factorials; a number of numerical results given in Fontana's paper thus appear as special cases of (26). For positive even n, the functions F in (27) reduce to polynomials; but for odd n, they are infinite series, so that the main advantage of (19) and (21) is lost by this transformation.

Hypergeometric functions which admit of quadratic transformations such as (26) are related to Legendre functions. Comparison of (27a) with (B 3.2.41) shows that the $R_{nl}(r_1, r_2)$ can be expressed in terms of associated Legendre functions of the second kind $Q_l^\mu[(r_1^2+r_2^2)/(2r_1r_2)]$, where $\mu = -1-\frac{1}{2}n$. Since, however, the various definitions of Q_l^μ for fractional μ involve differing phase angles, this approach is not studied further.

4. RECURRENCE RELATIONS

Any three contiguous hypergeometric functions, i.e., whose parameters differ by an integer only, satisfy a linear recurrence relation; hence there exists a linear relation between any three radial functions $R_{nl}(r_1, r_2)$, provided the values of l differ by integers and those of n, by even integers. Thus application of (B 2.8.31) to (27b) yields

$$(4 + 2l + n)(2l - 2 - n)R_{n+2,l} + 2(2 + n)^2(r_1^2 + r_2^2)R_{nl} - n(n + 2)(r_1^2 - r_2^2)^2 R_{n-2,l} = 0, \quad (29)$$

of (B 2.9.3) and (B 2.8.45) to (19)

$$\frac{r_1^2 + r_2^2}{r_1 r_2} R_{nl} - \frac{l + 2 + \frac{1}{2}n}{l + \frac{3}{2}} R_{n,l+1} - \frac{l - 1 - \frac{1}{2}n}{l - \frac{1}{2}} R_{n,l-1} = 0, \quad (30)$$

and of (B 2.8.35) to (27a)

$$(r_1^2 + r_2^2)R_{nl} - \frac{2l + 1}{l - \frac{1}{2}} r_1 r_2 R_{n,l-1} - \frac{2 + l + \frac{1}{2}n}{1 + \frac{1}{2}n} R_{n+2,l}. \quad (31a)$$

Elimination of $R_{n,l-1}$ or R_{nl} from (30) and (31a) leads to

$$(r_1^2 + r_2^2)R_{nl} - \frac{2l + 1}{l + \frac{3}{2}} r_1 r_2 R_{n,l+1} + \frac{l - 1 - \frac{1}{2}n}{1 + \frac{1}{2}n} R_{n+2,l} = 0, \quad (31b)$$

and

$$r_1 r_2 \left[\frac{R_{n,l+1}}{l + \frac{3}{2}} - \frac{R_{n,l-1}}{l - \frac{1}{2}}\right] - \frac{R_{n+2,l}}{1 + \frac{1}{2}n} = 0, \quad (31c)$$

respectively, and application of (29) to (31a) and (31b) yields

$$n(r_1^2 - r_2^2)^2 R_{n-2,l} = (2l + 2 + n)(r_1^2 + r_2^2)R_{nl} - (2l + 1)(2l - 2 - n)r_1 r_2 R_{n,l-1}/(l - \tfrac{1}{2}), \quad (32a)$$

$$= -(2l - n)(r_1^2 + r_2^2)R_{nl} + (2l + 1)(4 + 2l + n)r_1 r_2 R_{n,l+1}/(l + \tfrac{3}{2}); \quad (32b)$$

with a renewed application of (30), this leads to

$$\frac{n(r_1^2 - r_2^2)^2 R_{n-2,l}}{r_1 r_2} = 2\frac{(l + 1 + \frac{1}{2}n)_2}{l + \frac{3}{2}} R_{n,l+1} - 2\frac{(l - 1 - \frac{1}{2}n)_2}{l - \frac{1}{2}} R_{n,l-1}. \quad (32c)$$

All these formulas are three-term recurrence relations, independent of the relative magnitudes of r_1 and r_2. As mentioned in the introduction, use has previously been made of (6) to express $R_{n+2,l}$ in terms of $R_{nl}, R_{n,l-1}$ and $R_{n,l+1}$; such formulas are, of necessity, four-term recurrence relations.

5. EXPLICIT FORMULAS FOR THE LOGARITHM AND THE INVERSE SQUARE

The expansion for $\log r$ corresponding to (5),

$$\log r = \sum R_{\log,l}(r_1, r_2)P_l(\cos\vartheta_{12}), \tag{34}$$

is most easily deduced from the limiting process

$$\log r = \lim \partial(r^n)/\partial n, \qquad \text{as} \quad n \to 0. \tag{35}$$

The factor $(-\frac{1}{2}n)_l$, which occurs in the expressions for R_{nl}, vanishes for $n = 0$, $l > 0$, but gives a nonzero derivative; hence for all $l > 0$ we obtain from (19), (21), and (27),

$$R_{\log,l} = -\frac{(l-1)!}{(\frac{3}{2})_{l-1}}\left(\frac{r_<}{r_>}\right)^l F\left(l, -\tfrac{1}{2}; l+\tfrac{3}{2}; \frac{r_<^2}{r_>^2}\right), \tag{36a}$$

$$= -\frac{(l-1)!}{(\frac{3}{2})_{l-1}}\frac{r_<^l(r_>^2 - r_<^2)}{r_>^{l+2}}$$

$$\times F\left(l+2, \tfrac{3}{2}; l+\tfrac{3}{2}; \frac{r_<^2}{r_>^2}\right), \tag{36b}$$

$$= -\frac{(l-1)!}{(\frac{3}{2})_{l-1}}\left(\frac{r_1 r_2}{r_1^2 + r_2^2}\right)^l$$

$$\times F\left[\tfrac{1}{2}l, \tfrac{1}{2}l + \tfrac{1}{2}; l+\tfrac{3}{2}; \frac{4r_1^2 r_2^2}{(r_1^2+r_2^2)^2}\right], \tag{36c}$$

$$= -\frac{(l-1)!}{(\frac{3}{2})_{l-1}}\frac{(r_1 r_2)^l}{(r_1+r_2)^{2l}}$$

$$\times F\left[l, l+1; 2l+2; \frac{4r_1 r_2}{(r_1+r_2)^2}\right]. \tag{36d}$$

For $l = 0$, the differentiation must be applied to the other factors; (19) and (27) yield

$$R_{\log,0} = \log r_> + \sum \frac{(r_</r_>)^{2s}}{2s(2s-1)(2s+1)}, \tag{37a}$$

$$= \log(r_1 + r_2) - \tfrac{1}{2}\sum \frac{1}{s(s+1)}\frac{(4r_1 r_2)^s}{(r_1+r_2)^{2s}}, \tag{37b}$$

$$= \tfrac{1}{2}\log(r_1^2 + r_2^2) - \tfrac{1}{8}\sum \frac{1}{s(s+\frac{1}{2})}\left(\frac{2r_1 r_2}{r_1^2+r_2^2}\right)^{2s}, \tag{37c}$$

the index of summation running from 1 to ∞ in all cases. These series can be summed, leading to

$$R_{\log,0} = \log|r_1 - r_2| + \frac{(r_1+r_2)^2}{4r_1 r_2}\log\frac{r_1+r_2}{|r_1-r_2|} - \tfrac{1}{2}. \tag{38a}$$

Similarly, (36) can be summed for $l = 1$, with the result

$$R_{\log,1} = \frac{3}{16}\left(\frac{r_1^2 - r_2^2}{r_1 r_2}\right)^2 \times \log\frac{r_1+r_2}{|r_1-r_2|} - \frac{3}{8}\left(\frac{r_1^2+r_2^2}{r_1 r_2}\right). \tag{38b}$$

Differentiation of (30) yields, with (35), for $l > 0$,

$$\frac{r_1^2+r_2^2}{r_1 r_2}R_{\log,l} - \frac{2l+4}{2l+3}R_{\log,l+1} - \frac{2l-2}{2l-1}R_{\log,l-1} + \delta_{l,1} = 0, \tag{39}$$

$\delta_{l,m}$ being the Kronecker symbol. Similarly, (19) can be easily summed for $n = -2$ leading to

$$R_{-2,0} = \log[(r_1+r_2)/|r_1-r_2|](2r_1 r_2)^{-1}, \tag{40a}$$

$$R_{-2,1} = \tfrac{3}{4}(r_1^{-2} + r_2^{-2}) \times \log[(r_1+r_2)/|r_1-r_2|] - \tfrac{3}{2}(r_1 r_2)^{-1}. \tag{40b}$$

The recurrence relations (30) remain valid for $n = -2$, but in (31) the limiting ratio $R_{n+2,l}(1+\frac{1}{2}n)^{-1}$ is to be interpreted as $2R_{\log,l}$ $(l > 0)$; similarly in (32), R_{nl}/n tends to $R_{\log,l}$ as n tends to zero and $l > 0$.

6. EXPANSION FORMULAS FOR ARBITRARY FUNCTIONS OF r

The expansion (19) has the advantage that n occurs, as an exponent, for $r_>$ only, and, within each gamma product, only in the numerator. This allows the algebraic products to be expressed as products of the operator $(\partial/\partial r_>)$. In fact, we can equate

$$(-\tfrac{1}{2}n)_{l+s}(-\tfrac{1}{2}-\tfrac{1}{2}n)_s \; r_>^{n-l-2s} = \frac{(-)^l}{2^{2+2s}}r_>^l\left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l\left[\frac{1}{r_>}\left(\frac{\partial}{\partial r_>}\right)^{2s} r_>^{n+1}\right] \tag{41}$$

so that (19) can be written as

$$R_{nl} = (-r_< r_>)^l(2l+1)\sum_{s=0}^{\infty}\frac{r_<^{2s}}{(2s)!!\,(2s+2l+1)!!} \times \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l\left[\frac{1}{r_>}\left(\frac{\partial}{\partial r_>}\right)^{2s} r_>^{n+1}\right], \tag{42}$$

where

$$(2k)!! = 2\cdot 4\cdots 2k = 2^k k!, \; 0!! = (-1)!! = 1, \quad (2k+1)!! = 1\cdot 3\cdots(2k+1) = 2^{k+1}(\tfrac{1}{2})_{k+1}. \tag{43}$$

This suggests, for any function $f(r)$ which can be represented as a finite or infinite sum of powers,

not necessarily integer,

$$f(r) = \sum c_n r^n, \tag{44}$$

i.e., for essentially all well-behaved functions $f(r)$, that

$$f(r) = \sum_{l=0}^{\infty} f_l(r_>, r_<) P_l(\cos \vartheta_{12}), \tag{45}$$

where

$$f_l = (2l+1)(-r_< r_>)^l \sum_{s=0}^{\infty} \frac{r_<^{2s}}{(2s)!!\,(2s+2l+1)!!} \times \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l \left\{\frac{1}{r_>}\left(\frac{\partial}{\partial r_>}\right)^{2s}[r_> f(r_>)]\right\}. \tag{46}$$

This formula can be written symbolically by means of the modified spherical Bessel functions

$$i_l(z) = \sum_{s=0}^{\infty} \frac{z^{l+2s}}{(2s)!!\,(2l+2s+1)!!} = \left(\frac{\pi}{2z}\right)^{\frac{1}{2}} I_{l+\frac{1}{2}}(z) \tag{47}$$

[this is not the notation given in (B 7.2.6)] as

$$f_l = (2l+1)(-r_< r_>)^l \left(\frac{1}{r_>}\frac{\partial}{\partial r_>}\right)^l \times \left\{\frac{1}{r_>}\frac{i_l(r_<\partial/\partial r_>)}{(r_<\partial/\partial r_>)^l}[r_> f(r_>)]\right\}. \tag{48}$$

Similarly, (27) can be turned into an operational expansion if we introduce the new variables $\rho = (r_1^2 + r_2^2)^{\frac{1}{2}}$ and $r_+ = r_1 + r_2$. Thus (27a) leads to

$$f_l = \sum_s \frac{(-r_1 r_2)^{l+2s}(2l+1)}{(2s)!!\,(2s+2^l+1)!!}\left(\frac{1}{\rho}\frac{\partial}{\partial\rho}\right)^{l+2s} f(\rho) = (2l+1) i_l\left[-\frac{r_1 r_2}{\rho}\frac{\partial}{\partial\rho}\right] f(\rho). \tag{49}$$

Similarly, (27b) yields

$$f_l = \frac{1}{(2l-1)!!} \times \sum \frac{2^s(-r_1 r_2)^{l+s}(1+l)_s}{s!\,(2+2l)_s}\left(\frac{1}{r_+}\frac{\partial}{\partial r_+}\right)^{l+s} f(r_+) = \frac{1}{(2l-1)!!}\left(-\frac{r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right)^l \times \Phi\left(1+l; 2+l; \frac{-2r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right) f(r_+), \tag{50}$$

where Φ is the confluent hypergeometric function (B 6). In both (49) and (50), the product $r_1 r_2$ is to be treated as a constant on differentiation. The equivalence of (49) and (50) follows from the connection of $\Phi(\alpha; 2\alpha; 2z)$ and the Bessel functions (B 6.9.10),

$$I_\nu(z) = [(\tfrac{1}{2}z)^\nu/\Gamma(\nu+1)]e^{-z}\Phi(\tfrac{1}{2}+\nu; 1+2\nu; 2z), \tag{51}$$

which with (47) turns (50) into

$$f_l = (2l+1) i_l\left(-\frac{r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right) \exp\left(-\frac{r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right) f(r_+). \tag{52}$$

Taylor's expansion, which can be written operationally

$$\exp(h\partial/\partial z) f(z) = f(z+h), \tag{53}$$

and the identity

$$(z^{-1}\partial/\partial z) = 2\partial/\partial(z^2), \tag{54}$$

show that (52) is equivalent to

$$f_l = (2l+1) i_l\left(-\frac{r_1 r_2}{r_+}\frac{\partial}{\partial r_+}\right) f[(r_+^2 - 2r_1 r_2)^{\frac{1}{2}}], \tag{55}$$

which is another way of writing (49).

The convergence of the expansions (42), (49), and (50) are not discussed in detail. Qualitively we can say that, for any function $f(r)$ which is analytic for $|r| < M$, the expansions converges as long as $|r_1| + |r_2| < M$. If $f(r)\cdot r^{-n} (n \neq 0)$ tends to a finite nonzero limit as r tends to zero, this does not affect the convergence for $r_1 \neq r_2$, and even when $r_1 = r_2$, (22) shows that we can expect convergence as long as $n > -2$.

For two types of functions $f(r)$, the expansions (42), (49), and (50) factorize. Let $f(r)$ be a spherically symmetric solution of the wave equation,

$$\nabla^2 f = r^{-1}\partial^2(rf)/\partial r^2 = -k^2 f, \tag{56}$$

i.e., a spherical Bessel function of order zero of the first, second, or third kind (B 7.2.6),

$$j_0(kr) = \sin(kr)/(kr), \quad y_0(kr) = -\cos(kr)/(kr), \quad h_0^{(1)}(kr) = -ie^{ikr}/(kr), \quad h_0^{(2)}(kr) = ie^{-ikr}/(kr), \tag{57}$$

where the same relation as (47) holds between the pairs of functions j_l and $J_{l+\frac{1}{2}}$, y_l and $Y_{l+\frac{1}{2}}$, and h_l and $H_{l+\frac{1}{2}}$. Then in view of (56), the recurrence relations (B 7.11.7–10),

$$w_l(z) = (-z)^l(z^{-1}d/dz)^l w_0(z), \quad w = j, y, h^{(1)}, h^{(2)}, \tag{58}$$

and the series expansion for $j_l(z)$ which differs from (47) only by the factor $(-)^s$, (45), and (46), lead to

$$w_0(kr) = \sum_l (2l+1) j_l(kr_<) w_l(kr_>) P_l(\cos\vartheta_{12}), \quad w = j, y, h^{(1)}, h^{(2)}, \tag{59}$$

which is Gegenbauer's addition theorem[1] (B 7.15.28, 30) particularized to spherical Bessel functions. For the modified Bessel functions i_l and $k_l = (\pi/2z)^{\frac{1}{2}} K_{l+\frac{1}{2}}$, the corresponding results are, in view of (B 7.2.43)

and (B 7.11.20),

$$i_0(kr) = \sum (-)^l(2l+1)i_l(kr_<)i_l(kr_>)P_l(\cos\vartheta_{12}), \quad (60)$$
$$k_0(kr) = \sum (2l+1)i_l(kr_<)k_l(kr_>)P_l(\cos\vartheta_{12})$$

(cf. B 7.6.3); the latter serves as the basis of the zeta-function expansion about a common center in the method by Barnett and Coulson[5] for evaluating molecular integrals.

If $f(r)$ is a Gaussian function,

$$f(r) = \exp(-kr^2), \quad (r^{-1}\partial/\partial r)f(r) = -2kf(r), \quad (61)$$

the expansions (49) and (50) factorize, with the result

$$\exp(-kr^2) = \sum (2l+1)i_l(2kr_1r_2) \times \exp[-k(r_1^2+r_2^2)]P_l(\cos\vartheta_{12}), \quad (62)$$

or, on dividing by the common exponential,

$$\exp(2kr_1r_2\cos\vartheta_{12}) = \sum (2l+1)i_l(2kr_1r_2)P_l(\cos\vartheta_{12}). \quad (63a)$$

For imaginary values of k, this becomes

$$\exp(2ikr_1r_2\cos\vartheta_{12}) = \sum i^l(2l+1)j_l(2kr_1r_2)P_l(\cos\vartheta_{12}); \quad (63b)$$

these two formulas are equivalent to Sonine's expansion (B 7.10.5) for $\nu = \frac{1}{2}$; (63b) is equivalent to the well-known expansion for a three-dimensional plane wave in terms of spherical harmonics.

ACKNOWLEDGMENT

The writer wishes to thank Dr. A. W. Weiss, Dr. P. R. Fontana, and Professor E. Hylleraas for stimulating discussions and advice.

[5] M. P. Barnett and C. A. Coulson, Phil. Trans. Roy. Soc. **A243**, 221 (1951).

Three-Dimensional Addition Theorem for Arbitrary Functions Involving Expansions in Spherical Harmonics*

R. A. SACK
Department of Mathematics, Royal College of Advanced Technology, Salford, England†
and
Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin
(Received 16 August 1963)

For any vector $\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2$ an expansion is derived for the product of a power r^N of its magnitude and a surface spherical harmonic $Y_L^M(\vartheta, \varphi)$ of its polar angles in terms of spherical harmonics of the angles (ϑ_1, φ_1) and (ϑ_2, φ_2). The radial factors satisfy simple differential equations; their solutions can be expressed in terms of hypergeometric functions of the variable $(r_</r_>)^2$, and the leading coefficients by means of Gaunt's coefficients or $3j$ symbols. A number of linear transformations and three-term recurrence relations between the radial function are derived; but in contrast to the case $L = 0$, no generally valid expressions symmetric in r_1 and r_2 could be found. By interpreting the terms operationally, an expansion is derived for the product of $Y_L^M(\vartheta, \varphi)$ and an arbitrary function $f(r)$. The radial factors are expansions in derivatives of $f(r_>)$; for spherical waves, they factorize into Bessel functions of r_1 and r_2 in agreement with the expansion by Friedman and Russek. The $3j$ symbols are briefly discussed in an unnormalized form; the new coefficients are integers, satisfying a simple recurrence relation through which they can be arranged on a five-dimensional generalization of Pascal's triangle.

1. INTRODUCTION

In the preceding paper,[1] a generalization was derived of Laplace's expansion for the inverse distance between two points Q_1 and Q_2, specified by the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ or the spherical polar coordinates $(r_1, \vartheta_1, \varphi_1)$ and $(r_2, \vartheta_2, \varphi_2)$. It was shown that in the expansion for an arbitrary power of the distance in terms of Legendre polynomials of $(\cos \vartheta_{12})$,

$$|r_2 - r_1|^n = \sum R_{nl}(r_1, r_2) P_l(\cos \vartheta_{12}); \qquad (1)$$

the radial functions R_{nl} can be expressed in terms of hypergeometric functions of the argument $(r_</r_>)^2$, and by giving the expressions an operational interpretation, an addition theorem was obtained, valid for arbitrary analytic functions of $|\mathbf{r}_2 - \mathbf{r}_1|$.

A more general addition theorem would apply to functions $H(\mathbf{r}_2 - \mathbf{r}_1)$ or $H(\mathbf{r}_2 + \mathbf{r}_1)$, depending on the direction as well as on the magnitude of the vector argument. In Cartesian coordinates, such an expansion is given by Taylor's theorem in three variables; in many physical applications, however, it is of advantage to specify the dependence on the angles in terms of spherical harmonics. These harmonics can be defined in several ways in terms of the associated Legendre functions $P_l^m(x)$,

$$P_l^{|m|}(x) = (-)^m (1 - x^2)^{\frac{1}{2}|m|} [d^{|m|} P_l(x)/dx^{|m|}];$$
$$P_l^{-m}(x) = (-)^m P_l^m(x)[(l - m)!/(l + m)!]. \qquad (2)$$

The most useful definitions are for the unnormalized harmonics

$$\Theta_l^m(\vartheta, \varphi) = e^{im\varphi} P_l^{|m|}(\cos \vartheta),$$
$$\Omega_l^m(\vartheta, \varphi) = e^{im\varphi} P_l^m(\cos \vartheta), \qquad (3a, b)$$

and the normalized form

$$Y_l^m(\vartheta, \varphi) = [(2l + 1)(l - m)!/4\pi(l + m)!]^{\frac{1}{2}} \times e^{im\varphi} P_l^m(\cos \vartheta). \qquad (3c)$$

The functions $P_l(\cos \vartheta_{12})$ in (1) can be written as

$$P_l(\cos \vartheta_{12}) = \sum_{m=-l}^{l} (-)^m \Omega_l^{-m}(\vartheta_1, \varphi_1) \Omega_l^m(\vartheta_2, \varphi_2), \qquad (4)$$

with corresponding expressions in terms of Θ or Y (cf. B 3.11.2).[2]

The purpose of the present paper is to derive the expansion for the product of a spherical harmonic and a power of the radius

$$V_{NML} = r^N \Omega_L^M(\vartheta, \varphi)$$
$$= \sum R(N, L, l_1, l_2, M, m_1, m_2; r_1, r_2) \times \Omega_{l_1}^{m_1}(\vartheta_1, \varphi_1) \Omega_{l_2}^{m_2}(\vartheta_2, \varphi_2), \qquad (5)$$

and its generalization for functions of the type $f(r)\Omega_L^M(\vartheta, \varphi)$. In contrast to I, the vector $\mathbf{r} = (r, \vartheta, \varphi)$ denotes the *sum* of $\mathbf{r}_1$ and $\mathbf{r}_2$; the corresponding expressions for the difference $(\mathbf{r}_2 - \mathbf{r}_1)$ differ from those in (5) at most by a sign, corresponding to the parity of l_1. The spherical harmonics in (5) could equally well be expressed in terms of Θ or Y; the

* Supported in part by National Aeronautics and Space Administration Grant NsG-275-62(4180). This work was begun at the Laboratory of Molecular Structure and Spectra, University of Chicago, Chicago, Illinois, Supported by Office of Naval Research Contract Nonr-2121(01).
† Permanent address.
[1] R. A. Sack, J. Math. Physics 5, 246 (1964). (Hereafter referred to as I).
[2] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953). Formulas in this work are directly referenced by the prefix B.

corresponding radial functions R_Θ and R_Y differ from $R \equiv R_\Omega$ only by a factor which is easily calculated from (2) and (3). In view of the transformation properties of the normalized functions Y, their use would have the advantage that the azimuthal quantum numbers $\mathbf{m} = M, m_1, m_2$ can affect the expressions R_Y only through the Wigner coefficients or $3j$-symbols.[3-5] The writer's personal preference is for the functions Ω, as they do not necessitate the use of square roots; the place of the $3j$-symbols is then taken by unnormalized $3j$-coefficients which have the advantage of being integers; as shown in the Appendix, they can be arranged on a five-dimensional generalization of Pascal's triangle.

For some specific cases, expansions of the type (5) have been given before; an addition theorem for solid spherical harmonics ($N = L$ or $N = -L - 1$) have been given by Rose,[6] and for spherical waves by Friedman and Russek;[7] more recently similar results have been rederived by Seaton.[8] The radial functions in the expansion (5) for the general case could be obtained by combining these results with those of I, i.e., by considering the product

$$V_{NLM} = r^n \cdot r^L \Omega_L^M(\vartheta, \varphi), \qquad n = N - L, \tag{6}$$

but this would involve the summation of multiple infinite series. Instead, the derivation of the functions R for arbitrary values of N are based, as in I, on the solution of the set of differential equations

$$\nabla_1^2 V_{NLM} = \nabla_2^2 V_{NLM}, \tag{7a}$$

$$\nabla^2 V_{NLM} = (N - L)(N + L + 1) V_{N-2,LM}. \tag{7b}$$

These solutions are again expressible in terms of hypergeometric functions, and leading coefficients are determined by comparison with special known cases; it is found that these constants can always be expressed in terms of integrals of products of three harmonics which may be given in their normalized or unnormalized forms. An alternative method of deriving these coefficients could be based on the transformation properties of the spherical harmonics, but neither this approach, nor any other group-theoretical arguments are employed in this paper. The only use made of the extensive theory of normalized harmonics[3-5] is of the relation between the integrals over triple products (Gaunt's coefficients)[9,10] and the $3j$-symbols, and the results obtained in terms of the functions Ω is reformulated in terms of the normalized harmonics Y.

The solutions of the Eqs. (7) satisfying the appropriate continuity conditions will be derived in Sec. 2, and the results discussed in Sec. 3. A selected number of recurrence relations are given in Sec. 4, and in Sec. 5 the formulas are given an operational form, applicable to arbitrary functions of r. The special case that one of the vectors points in the direction of the polar axis is considered in a later paper.

2. MATHEMATICAL DERIVATION

To avoid an excessive use of subscripts, formulas in this section are derived for the range $r_2 > r_1$ only. The dimensionality of (5) requires that the functions R be of the form

$$R(N, \mathbf{l}, \mathbf{m}; r_1, r_2) = r_1^{l_1} r_2^{N-l_1} \sum c_{Ns} (r_1/r_2)^s. \tag{8}$$

The differential equation (7a) substituted in (5) leads to

$$\frac{\partial^2 R}{\partial r_1^2} + \frac{2}{r_1}\frac{\partial^2 R}{\partial r_1} - l_1(l_1 + 1)\frac{R}{r_1^2} = \frac{\partial^2 R}{\partial r_2^2} + \frac{2}{r_2}\frac{\partial R}{\partial r_2} - l_2(l_2 + 1)\frac{R}{r_2^2}, \tag{9}$$

which together with (8) yields the recurrence relations

$$(s + 2)(2l_1 + s + 3)c_{N,s+2} = (N - l_1 - l_2 - s)(N - l_1 + l_2 + 1 - s)c_{Ns}. \tag{10}$$

The leading term in the power series (8) is of degree $s = 0$, since the other possible solution, beginning with $s = -2l - 1$, would lead to a singularity as $r_1 \to 0$. As in I, the solution is best expressed in terms of Gauss' hypergeometric function

$$F(\alpha, \beta; \gamma; z) = \sum_0^\infty (\alpha)_w (\beta)_w z^w / [(\gamma)_w w!], \tag{11}$$

where[11]

$$(\alpha)_0 = 1; \quad (\alpha)_w = (\alpha; w) = \alpha(\alpha + 1) \cdots \times (\alpha + w - 1) = \Gamma(\alpha + w)/\Gamma(\alpha). \tag{12}$$

[3] E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).
[4] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957).
[5] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1961).
[6] M. E. Rose, J. Math. and Phys. **37**, 215 (1958).
[7] B. Friedman and J. Russek, Quart. Appl. Math. **12**, 13 (1954).
[8] M. J. Seaton, Proc. Phys. Soc. **77**, 184 (1961).
[9] J. A. Gaunt, Phil. Trans. Roy. Soc. **A228**, 151 (1929).
[10] M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, *The 3-j and 6-j Symbols* (Technology Press, Cambridge, Massachusetts, 1959).
[11] The archaic form $(\alpha; w)$ is employed mainly when w carries a subscript.

If we abbreviate

$$\Lambda = \tfrac{1}{2}(L + l_1 + l_2), \qquad \lambda = \Lambda - L, \qquad \lambda_1 = \Lambda - l_1, \qquad \lambda_2 = \Lambda - l_2, \tag{13}$$

and use n as defined in (6), the solutions (8) and (10) can be expressed in the form

$$R(N, \mathbf{l}, \mathbf{m}; r_1, r_2) = K(N, \mathbf{l}, \mathbf{m}) r_1^{l_1} r_2^{N-l_1} F[\tfrac{1}{2}(l_1 + l_2 - N), \tfrac{1}{2}(l_1 - l_2 - 1 - N); l_1 + \tfrac{3}{2}; r_1^2/r_2^2], \tag{14a}$$

$$= K(N, \mathbf{l}, \mathbf{m}) r_1^{l_1} r_2^{n+L-l_1} \times F(\lambda - \tfrac{1}{2}n, -\tfrac{1}{2} - \tfrac{1}{2}n - \lambda_1; l_1 + \tfrac{3}{2}; r_1^2/r_2^2). \tag{14b}$$

As the functions Ω_l^m have the parity of l on inversion, $(\vartheta, \varphi) \to (\pi - \vartheta, \pi + \varphi)$, $K(n, \mathbf{l}, \mathbf{m})$ can take nonzero values only if

$$L - l_1 - l_2 = \text{even}, \tag{15}$$

and hence all the quantities defined in (13) are integers. The leading coefficients K in (14 a,b) satisfy the recurrence relation, in view of (7b):

$$n(n + 1 + 2L)K(N - 2, \mathbf{l}, \mathbf{m}) = (n - 2\lambda)(n + 1 + 2\lambda_1)K(N, \mathbf{l}, \mathbf{m}). \tag{16}$$

This means that K depends on n through the factors

$$(-\tfrac{1}{2}n; \lambda)(-\tfrac{1}{2} - \tfrac{1}{2}n - L; \lambda_2); \tag{17}$$

the only other way K could depend on n would be through an additional, periodic, factor of period 2; but, according to the results of I, the factor r^n in (6) does not show any such periodicity and the solid harmonics are independent of n; hence (17) describes the full dependence of K on n. To find the absolute value of $K(N, \mathbf{l}, \mathbf{m})$ we first consider the case $N = L$ or $n = 0$. Making use of (B 3.7.25) and its converse

$$P_l^m(\cos\vartheta)e^{im\varphi} = \frac{i^m(l + m)!}{2\pi l!} \times \int_0^{2\pi} [\cos\vartheta + i\sin\vartheta\cos(\varphi - \psi)]^l e^{im\psi}\, d\psi, \tag{18a}$$

$$[\cos\vartheta + i\sin\vartheta\cos(\varphi - \psi)]^l = \sum_m \frac{l!}{i^m(l + m)!} P_l^m(\cos\vartheta)e^{im(\varphi - \psi)}, \tag{18b}$$

we obtain for the solid harmonics, by means of the binomial theorem,

$$r^L\Omega_L^M(\vartheta, \varphi) = \frac{i^M(L + M)!}{2\pi L!} \int_0^{2\pi} (z_1 + ix_1\cos\psi + iy_1\sin\psi + z_2 + ix_2\cos\psi + iy_2\sin\psi)^L e^{iM\psi}\, d\psi = \sum (L + M)! [(l_1 + m_1)!\,(l_2 + m_2)!]^{-1} \times r_1^{l_1} r_2^{l_2} \Omega_{l_1}^{m_1}(\vartheta_1, \varphi_1)\Omega_{l_2}^{m_2}(\vartheta_2, \varphi_2), \tag{19}$$

the sum to be taken over all

$$m_1 + m_2 = M; \qquad l_1 + l_2 = L. \tag{20 a,b}$$

This is the unnormalized form of Rose's addition theorem.[6] Multiplication by r^n gives rise to terms for which (20b) is no longer satisfied. For positive even n, Eq. (19) of I shows that in the expansion (1) for $|\mathbf{r}_1 + \mathbf{r}_2|^n$, the radial coefficient of P_λ is $\lambda!(r_1r_2)^\lambda/(\frac{1}{2})_\lambda$ for $\lambda = \frac{1}{2}n$, and vanishes for $\lambda > \frac{1}{2}n$. Hence the leading term in $R(l_1 + l_2, \mathbf{l}, \mathbf{m})$, in view of (6), (13), and (19), is made up of terms

$$\frac{\lambda!\, r_1^{l_1} r_2^{l_2}}{(\frac{1}{2})_\lambda} \sum_\mu (-)^\mu \frac{(L + M)!}{(\lambda_2 + m_1 + \mu)!\,(\lambda_1 + m_2 - \mu)!} \times \Omega_{\lambda_2}^{m_1+\mu}(1)\Omega_{\lambda_1}^{m_2-\mu}(2)\Omega_\lambda^{-\mu}(1)\Omega_\lambda^\mu(2). \tag{21}$$

The product of two surface harmonics of the same coordinates (ϑ, φ) can be expressed as a sum of spherical harmonics,[3-5,9,12,13]

$$\Omega_l^m\Omega_\lambda^\mu = \frac{(2l)!\,(2\lambda)!\,(l + \lambda)!\,(l + \lambda - m - \mu)!}{(2l + 2\lambda)!\, l!\, \lambda!\,(l - m)!\,(\lambda - \mu)!} \times \Omega_{l+\lambda}^{m+\mu} + \cdots \Omega_{l+\lambda-2}^{m+\mu} + \cdots. \tag{22}$$

This leading term can be found most easily by a comparison of the leading coefficients of $P_l^m(x)$, which in view of (2) and Rodrigues' formula (B 3.6.16), are

$$P_l^m(x) = (-)^m(1 - x^2)^{m/2} \times [(2l)!/2^l l!\,(l - m)!]x^{l-m} + \cdots. \tag{23}$$

The leading coefficient K in (14) for $N = l_1 + l_2$ thus becomes, in view of (21) and (22),

$$K(l_1 + l_2, \mathbf{l}, \mathbf{m}) = (-)^{\Lambda+M} \times \frac{(L + M)!\,(l_1 - m_1)!\,(l_2 - m_2)!\,(2\lambda)!\, l_1!\, l_2!}{\lambda_1!\,\lambda_2!\,\lambda!\,(\frac{1}{2})_\lambda(2l_1)!\,(2l_2)!} \times U\begin{pmatrix} L & l_1 & l_2 \\ -M & m_1 & m_2 \end{pmatrix}, \tag{24}$$

where the symbols U represent the sums

$$U\begin{pmatrix} L & l_1 & l_2 \\ -M & m_1 & m_2 \end{pmatrix} = \sum_\mu (-)^{-\Lambda+\mu+M} \times \binom{2\lambda}{\lambda + \mu}\binom{2\lambda_1}{\lambda_1 - m_2 + \mu}\binom{2\lambda_2}{\lambda_2 + m_1 + \mu}, \tag{25}$$

provided (20a) holds. They are related to the Wigner $3j$-symbols[3-5]

$$U\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times \left[(2\Lambda + 1)! \prod_{s=1}^{3} \frac{(2\lambda_s)!}{(j_s - m_s)!\,(j_s + m_s)!}\right]^{\frac{1}{2}}, \tag{26}$$

[12] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[13] E. A. Hylleraas, Math. Scand. **10**, 189 (1962).

where, for this equation only, we have put

$$\Lambda = \tfrac{1}{2}(j_1 + j_2 + j_3); \qquad \lambda_s = \Lambda - j_s \geq 0 \quad (s = 1, 2, 3). \tag{27}$$

In the present context these unnormalized $3j$-symbols are required for integral values of $\mathbf{l}$, $\mathbf{m}$, and $\boldsymbol{\lambda}$ only, but, as shown in the Appendix, their definition (25) also covers the case of half-integer parameters. For integer Λ they are invariant under a permutation of (1, 2, 3) in (26), and under a simultaneous change of sign of all the m_s.

The expression (24) can be simplified by the explicit use of Gaunt's coefficients[9,10] for the integral over the product of three associated Legendre functions. If we put

$$I_\Omega\begin{pmatrix} L & l & l' \\ M & -m & -m' \end{pmatrix} = \int_{-1}^{1} P_L^M(x) P_l^{-m}(x) P_{l'}^{-m'}(x)\, dx, \tag{28}$$

where the azimuthal numbers add up to zero, these integrals can be expressed in terms of the U's as[9]

$$I_\Omega\begin{pmatrix} L & l_1 & l_2 \\ M & -m_1 & -m_2 \end{pmatrix} = \frac{(-)^\Lambda 2(l_1 - m_1)!\,(l_2 - m_2)!\,(L + M)!\,\Lambda!}{(2\Lambda + 1)!\,\lambda_1!\,\lambda_2!\,\lambda!} \times U\begin{pmatrix} L & l_1 & l_2 \\ M & -m_1 & -m_2 \end{pmatrix}, \tag{29}$$

so that (24) becomes

$$K(l_1 + l_2, \mathbf{l}, \mathbf{m}) = \frac{(-)^M \lambda!\,(\tfrac{1}{2}; \Lambda + 1)}{(\tfrac{1}{2}; l_1)(\tfrac{1}{2}; l_2)}\, I_\Omega\begin{pmatrix} L & l_1 & l_2 \\ M & -m_1 & -m_2 \end{pmatrix}. \tag{30}$$

Using (17), we find, for the leading coefficient for arbitrary N,

$$K(N, \mathbf{l}, \mathbf{m}) = (-)^\lambda \frac{(-\tfrac{1}{2}n; \lambda)(-\tfrac{1}{2} - \tfrac{1}{2}n - L; \lambda_2)}{\lambda!\,(-\tfrac{1}{2} - \Lambda; \lambda_2)} \times K(l_1 + l_2, \mathbf{l}, \mathbf{m}) = (-)^{\lambda+M} \frac{(l_2 + \tfrac{1}{2})(-\tfrac{1}{2}n; \lambda)(\tfrac{3}{2} + \tfrac{1}{2}n; L)}{(\tfrac{1}{2}; l_1)(\tfrac{3}{2} + \tfrac{1}{2}n; \lambda_1)} \times I_\Omega\begin{pmatrix} L & l_1 & l_2 \\ M & -m_1 & -m_2 \end{pmatrix}. \tag{31}$$

3. DISCUSSION OF THE RADIAL FUNCTIONS R

According to (14) and (31), the radial functions R in the expansion (5) are given, for $r_2 > r_1$, by

$$R(N, \mathbf{l}, \mathbf{m}; r_1, r_2) = K'(\mathbf{l}, \mathbf{m}) R'(N, \mathbf{l}; r_1, r_2), \tag{32}$$

where

$$K'(\mathbf{l}, \mathbf{m}) = (-)^{\lambda+M}(l_1 + \tfrac{1}{2})(l_2 + \tfrac{1}{2}) \times I_\Omega\begin{pmatrix} L & l_1 & l_2 \\ M & -m_1 & -m_2 \end{pmatrix}, \tag{33}$$

$$R'(N, \mathbf{l}; r_1, r_2) = \frac{(-\tfrac{1}{2}n; \lambda)(\tfrac{3}{2} + \tfrac{1}{2}n; L)}{(\tfrac{1}{2}; l_1 + 1)(\tfrac{3}{2} + \tfrac{1}{2}n; \lambda_1)} r_1^{l_1} r_2^{N-l_1} \times F(\lambda - \tfrac{1}{2}n, -\tfrac{1}{2} - \tfrac{1}{2}n - \lambda_1; l_1 + \tfrac{3}{2}; r_1^2/r_2^2), \tag{34}$$

and the symbols are explained in (6), (11)–(13), (25), (28), and (29); for $r_1 > r_2$, the subscripts 1 and 2 should be interchanged. Equation (32) factorizes the functions $R(N, \mathbf{l}, \mathbf{m})$ into a constant K', independent of N or n, and a function R', independent of the azimuthal quantum numbers $\mathbf{m}$. The precise separation is, to some extent, arbitrary, since any dependence on $\mathbf{l}$ or $\boldsymbol{\lambda}$ only can be drawn into either factor; the selection (33), (34) was chosen primarily to give the recurrence relations of Sec. 4 their simplest form. In the case of spherical symmetry,

$$L = M = \lambda_1 = \lambda_2 = 0, \quad m_1 = -m_2 = m, \quad l_1 = l_2 = \lambda = l, \quad K' = (-)^{l+m}(l + \tfrac{1}{2}), \tag{35}$$

and the R' differ from the functions R_{nl} of (1) and I only by a factor $(l + \frac{1}{2})^{-1}$.

If the spherical harmonics in (5) are given in their normalized form Y_l^m, (3c), the analogous radial functions R_Y can be factorized as in (32),

$$R_Y(N, \mathbf{l}, \mathbf{m}; r_1, r_2) = K'_Y(\mathbf{l}, \mathbf{m}) R'(N, \mathbf{l}; r_1, r_2), \tag{36}$$

where R' remains unaltered as in (34), whereas in view of (3c) and (26),

$$K'_Y(\mathbf{l}, \mathbf{m}) = 2\pi(-)^\lambda (LM\ |Y_{l_1}^{m_1}|\ l_2 m_2). \tag{37}$$

Here

$$(LM\ |Y_l^m|\ l'm') = (-)^M[(2l + 1)(2l' + 1)(2L + 1)/4\pi]^{\frac{1}{2}} \times \begin{pmatrix} L & l & l' \\ -M & m & m' \end{pmatrix}\begin{pmatrix} L & l & l' \\ 0 & 0 & 0 \end{pmatrix} = \iint [Y_L^M(\vartheta, \varphi)]^* Y_l^m(\vartheta, \varphi) Y_{l'}^{m'}(\vartheta, \varphi) \times \sin\vartheta\, d\vartheta\, d\varphi \tag{38}$$

is the integral of the product of three normalized harmonics taken over the whole unit sphere.[10] In view of the properties of the $3j$-symbols, the coefficients K', and hence the radial functions R, is nonzero only if the conditions

$$|l_1 - l_2| \leq L \leq l_1 + l_2 \tag{39}$$

are satisfied, as well as (15) and (20a). The functions R also vanish, in view of (13), (32), and (34), if

$$L \leq N < l_1 + l_2, \quad n \text{ even}, \tag{40a}$$

or

$$-L \leq N + 1 < l_1 - l_2, \quad n \text{ odd}, \quad r_2 > r_1. \tag{40b}$$

The hypergeometric series are polynomials if

$$N \geq l_1 + l_2, \quad n \text{ even},$$

or

$$N \geq l_1 - l_2 - 1, \quad n \text{ odd}; \tag{41}$$

if either equality holds, they reduce to the leading term unity. The particular case $N = L$ has been discussed in (19) and (20); if $N = -L - 1$, the only nonvanishing functions in (5) for $r_2 > r_1$ are those for which $l_2 = L + l_1$, and for these we have, in view of (22) or from Refs. 9 and 11,

$$I_\Omega\begin{pmatrix} L, & l_1, & L + l_1 \\ M, & -m_1, & m_1 - M \end{pmatrix} = (-)^{M-m_1} \times \frac{2(2L)!\,(2l_1)!\,(l_1 + L)!\,(l_1 + L + m_1 - M)!}{(2l_1 + 2L + 1)!\,L!\,l_1!\,(L - M)!\,(l_1 + m_1)!}, \tag{42}$$

so that (5) and (32)–(34) yield

$$r^{-L-1}\Omega_L^M(\vartheta, \varphi) = \sum_{l,m} (-)^{l+m} \frac{(l + L + m - M)!}{(l + m)!\,(L - M)!} r_1^l r_2^{-L-l-1} \times \Omega_l^m(\vartheta_1, \varphi_1)\Omega_{L+l}^{M-m}(\vartheta_2, \varphi_2). \tag{43}$$

This corresponds to the expansion for normalized "irregular" solid harmonics given by Rose[6] and recently by Chiu.[14]

As in I, the transformation theory of the hypergeometric functions can be applied to the expression (34) for the functions R'. Thus (B 2.9.1, 2) or Eq. (20a) of I leads to

$$R'(N, 1; r_1, r_2) = \frac{(-\tfrac{1}{2}n; \lambda)(\tfrac{3}{2} + \tfrac{1}{2}n; L)}{(\tfrac{1}{2}; l_1 + 1)(\tfrac{3}{2} + \tfrac{1}{2}n; \lambda_1)} \frac{r_1^{l_1}(r_2^2 - r_1^2)^{N+2}}{r_2^{N+4+l_1}} \times F(\Lambda + 2 + \tfrac{1}{2}n, \tfrac{3}{2} + \tfrac{1}{2}n + \lambda_2; l_1 + \tfrac{3}{2}; r_1^2/r_2^2), \tag{44}$$

which shows that the radial functions are also rational in r_1 and r_2 if $\frac{1}{2}(l_1 + l_2 + N) + 1$ or $\frac{1}{2}(l_1 - l_2 + N + 1)$ are negative integers. Similarly (B 2.10.1) or (20b) of I yield

$$R'(N, 1; r_1, r_2) = \frac{2^{n+1}(-\tfrac{1}{2}n; \lambda)(\tfrac{3}{2} + \tfrac{1}{2}n; L)(2 + n; L)}{(1 + \tfrac{1}{2}n; \Lambda + 1)(\tfrac{3}{2} + \tfrac{1}{2}n; \lambda_1)(\tfrac{3}{2} + \tfrac{1}{2}n; \lambda_2)} r_1^{l_1} r_2^{N-l_1} \times F(\lambda - \tfrac{1}{2}n, -\tfrac{1}{2} - \tfrac{1}{2}n - \lambda_1; -n - 1 - L; (r_2^2 - r_1^2)/r_2^2) + \left[-\frac{(-)^{\lambda_2}(\tfrac{3}{2} + \tfrac{1}{2}n; L)}{(n + 2; L + 1)2^{n+2}} \frac{r_1^{l_1}(r_2^2 - r_1^2)^{N+2}}{r_2^{N+4+l_1}}\right] \times F\left(\Lambda + 2 + \tfrac{1}{2}n, \tfrac{3}{2} + \tfrac{1}{2}n + \lambda_2; n + 3 + L; \frac{r_2^2 - r_1^2}{r_2^2}\right), \tag{45}$$

where the coefficients have been simplified in view of the properties of the gamma function (B 1.2.6) and (B 1.2.15), or (23) and (25) of I. This equation shows the nature of the branch point as r_1 approaches r_2; the difficulties arising for integer values of n have been discussed in I, following Eq. (22); the result is either a polynomial or a series involving logarithmic terms.

In the case $L = 0$, it was shown in I that, by means of quadratic transformations applied to the hypergeometric functions, the radial functions R_{nl} could be expressed in several forms symmetric in r_1 and r_2, involving power series in $r_1r_2/(r_1 + r_2)^2$ or in $r_1r_2/(r_1^2 + r_2^2)$. The same transformations can be applied whenever $l_1 = l_2$, regardless of the value of L; for general values of l_1 and l_2, (34) shows that even the leading coefficients are different as $r_1 < r_2$ or $r_1 > r_2$. In consequence, it is unlikely that analogous simple symmetric expansions exist in the general case. On the other hand, the leading coefficients in (45) are invariant for $r_1 \gtrless r_2$, and together with the symmetry of the recurrence relations derived below, this suggests the existence of symmetric expansions involving power series in the two arguments $r_1r_2/(r_1^2 + r_2^2)$ and $(r_1^2 - r_2^2)/(r_1^2 + r_2^2)$ or similar variables, though presumably involving the one variable only to a finite power depending on $|l_1 - l_2|$. So far the writer has been unable to derive such expansions.

Quadratic transformations for arbitrary hypergeometric functions have recently been derived by Kuipers and Meulenbeld[15] in terms of generalized hypergeometric functions or MacRobert's E functions, [cf. (B 4) and (B 5)]. This generalization, however, is not quite relevant to the problem at this stage, as it corresponds t ‹a generalization of the transformation from (27a) to (27b) of I, and not of the transformation from (19) to (27).

[14] Y. N. Chiu, J. Math. Phys. 5, 283 (1964).

[15] L. Kuipers and B. Meulenbeld, J. London Math. Soc. 35, 221 (1960).

It might be considered that the expansion (5) would simplify if one of the vectors, say $\mathbf{r}_1$, points in the direction of the polar axis; for this choice all the Legendre functions of $\cos\vartheta_1$ are 1 or 0, according as $m_1 = 0$ or $m_1 \neq 0$, and hence for all nonvanishing terms, $m_2 = M$. The individual terms in (5) are therefore considerably simpler than in the general case; on the other hand, because of the restrictions imposed on $\mathbf{r}_1$, the rotational quantum number l_1 ceases to be meaningful and any consistent expansion making use of this restriction should reasonably involve an implicit summation over l_1, i.e., over products involving $3j$-symbols. From an analytic point of view, these symbols are generalized hypergeometric series[9,16] [cf. also (B 4)] of unit argument and all integer parameters, and any expansion involving such functions is likely to lead back to functions of at least the same, and possibly higher, complexity. This has indeed been found to be the case, and in order not to complicate any further the mathematical apparatus required for the present paper, the case $\vartheta_1 = 0$ is to be considered separately in a later publication.

4. RECURRENCE RELATIONS

The relations between contiguous hypergeometric functions (B 2.8.28–45) can be used, as in I, to derive linear recurrence relations between any three radial functions R' for which L, l_1, l_2 and $\frac{1}{2}N$ differ by integers only; the recurrence formulas between the coefficients K' of (33) or (37) are known from the theory of angular momentum.[3-5,9,10] Equation (14a) shows that the functions F depend on n and L only through their sum N; according to (34),

$$\frac{R'(N, L+2, l_1, l_2)}{R'(N, L, l_1, l_2)} = \frac{3+N+L}{L-N} = -\frac{3+n+2L}{n}. \tag{46}$$

It is therefore sufficient to derive any further relations for varying values of the angular quantum numbers L, l_1, and l_2 only, leaving

$$n = N - L = \text{const}; \tag{47}$$

the value of N can then be increased or decreased in steps of 2 by means of (46). In view of the larger number of independent parameters, the number of recurrence relations for even small changes in 1 are considerable; we therefore confine our attention to the following special cases:

(i) Between any two of the three functions R', none of the numbers L, l_1, and l_2 differ by more than unity.

(ii) One of the angular quantum numbers remains constant, the second varies by at most unity, and the third by at most two units.

There are eight inequivalent three-term recurrence relations of type (i) and 12 of type (ii); for the sake of brevity, only those parameters are indicated which differ from L, l_1, l_2, e.g., $R'(L+, l_1-) = R'(L+1, l_1-1, l_2)$ [cf. (B 2.9)], and N is understood to vary according to (47). The formulas are

$$(\tfrac{3}{2}+\tfrac{1}{2}n+L)(r_2^2-r_1^2)R' = (\lambda_1+\tfrac{3}{2}+\tfrac{1}{2}n)r_2R'(L+, l_2+) - (\lambda_2+\tfrac{3}{2}+\tfrac{1}{2}n)r_1R'(L+, l_1+), \tag{48a}$$

$$= (\lambda-1-\tfrac{1}{2}n)r_2R'(L+, l_2-) - (\Lambda+2+\tfrac{1}{2}n)r_1R'(L+, l_1+), \tag{48b}$$

$$= (\Lambda+2+\tfrac{1}{2}n)r_2R'(L+, l_2+) - (\lambda-1-\tfrac{1}{2}n)r_1R'(L+, l_1-), \tag{48c}$$

$$= -(\tfrac{3}{2}+\lambda_2+\tfrac{1}{2}n)r_2R'(L+, l_2-) + (\lambda_1+\tfrac{3}{2}+\tfrac{1}{2}n)r_1R'(L+, l_1-); \tag{48d}$$

$$(\tfrac{1}{2}+\tfrac{1}{2}n+L)^{-1}R' = (\lambda-\tfrac{1}{2}n)^{-1}[r_2R'(L-, l_2+) + r_1R'(L-, l_1+)], \tag{49a}$$

$$= (\tfrac{1}{2}+\tfrac{1}{2}n+\lambda_1)^{-1}[r_2R'(L-, l_2-) - r_1R'(L-, l_1+)], \tag{49b}$$

$$= (\tfrac{1}{2}+\tfrac{1}{2}n+\lambda_2)^{-1}[-r_2R'(L-, l_2+) + r_1R'(L-, l_1-)], \tag{49c}$$

$$= (\Lambda+1+\tfrac{1}{2}n)^{-1}[r_2R'(L-, l_2-) + r_1R'(L-, l_1-)]; \tag{49d}$$

$$(l_2+\tfrac{1}{2})R' = (\tfrac{1}{2}+\tfrac{1}{2}n+L)r_2[R'(L-, l_2+)+R'(L-, l_2-)], \tag{50a}$$

$$= r_2[(\Lambda+2+\tfrac{1}{2}n)(\lambda_1+\tfrac{3}{2}+\tfrac{1}{2}n)R'(L+, l_2+) - (\lambda-1-\tfrac{1}{2}n)(\lambda_2+\tfrac{3}{2}+\tfrac{1}{2}n) \times R'(L+, l_2-)]/[(\tfrac{3}{2}+\tfrac{1}{2}n+L)(r_2^2-r_1^2)], \tag{50b}$$

$$= (r_2/r_1)[-(\lambda_1+\tfrac{3}{2}+\tfrac{1}{2}n)R'(l_1-, l_2+) + (\lambda-1-\tfrac{1}{2}n)R'(l_1-, l_2-)], \tag{51a}$$

$$= (r_2/r_1)[(\Lambda+2+\tfrac{1}{2}n)R'(l_1+, l_2+) + (\tfrac{3}{2}+\tfrac{1}{2}n+\lambda_2)R'(l_1+, l_2-)]; \tag{51b}$$

[16] P. E. Bryant, *Tables of Wigner 3j-Symbols* (Res. Rept. 60-1, published by University of Southampton, Southampton, England, 1960).

$$(L+n+2)(\tfrac{3}{2}+\tfrac{1}{2}n+L)r_2R'$$
$$=(\tfrac{1}{2}+\tfrac{1}{2}n+L)_2(r_2^2-r_1^2)R'(L-,\,l_2-)$$
$$+(\lambda-1-\tfrac{1}{2}n)(\lambda_2+\tfrac{1}{2}n+\tfrac{3}{2})R'(L+,\,l_2-), \quad (52a)$$
$$=-(\tfrac{1}{2}+\tfrac{1}{2}n+L)_2(r_2^2-r_1^2)R'(L-,\,l_2+)$$
$$+(2+\Lambda+\tfrac{1}{2}n)(\lambda_1+\tfrac{1}{2}n+\tfrac{3}{2})R'(L+,\,l_2+). \quad (52b)$$

The other six relations of the type (ii) are obtained by an interchange of the subscripts 1 and 2 in (50)–(52). Although the resulting equations are invariant on interchanging (l_1, r_1) and (l_2, r_2), their derivation is not symmetrical; thus (50) follows from (B 2.8.32, 37), but the corresponding equations for varying l_1 follow from

$$\gamma(\gamma-1)[F(\gamma-)-F]=\alpha\beta zF(\alpha+,\,\beta+,\,\gamma+) \quad (53)$$

and from (B 2.9.1, 2). Equations (49 a,b) follow from (B 2.8.38, 43), and (48c, d) from (B 2.8.35, 42); the remaining relations are derived from these by linear elimination, though to prove (51), the values of L in (48) and (49) must be lowered or raised.

It should be remembered in applying the recurrence relations (48)–(52), that they do not apply to the full radial functions R of (32); these latter vanish whenever the triangular condition (39) is violated because of the factor K' in (33), whereas the factors R' have perfectly well defined, usually nonzero, values in accordance with (46) regardless of the relative values of L, l_1, and l_2, provided only (15) is satisfied.

5. AN OPERATIONAL EXPANSION FOR ARBITRARY FUNCTIONS

As in I, the way in which the power N enters into the expressions (32)–(34) allows the functions $R'(N, 1; r_1, r_2)$ to be expressed in operational form. For $r_2 > r_1$, the expressions differ according to the relative magnitudes of L and l_2. For the factor in the general term in (34), which depends on N, we have, using (11)–(14),

$$(-)^{l_1}2^{l_1+2s}(\tfrac{1}{2}L-\tfrac{1}{2}N;\,\lambda+s)$$
$$\times(-\tfrac{1}{2}-\tfrac{1}{2}N-\tfrac{1}{2}L;\,\lambda_2+s)r_2^{N-l_1-2s}$$
$$=r_2^{-1-l_s}\left(\frac{1}{r_2}\frac{\partial}{\partial r_2}\right)^{L-l_s}r_2^L$$
$$\times\left[\frac{\partial^2}{\partial r_2^2}-\frac{L(L+1)}{r_2^2}\right]^{\lambda+s}r^{N+1}, \quad L\geq l_2, \quad (54a)$$
$$=r_2^{l_s}\left(\frac{1}{r_2}\frac{\partial}{\partial r_2}\right)^{l_s-L}r_2^{-1-L}$$
$$\times\left[\frac{\partial^2}{\partial r_2^2}-\frac{L(L+1)}{r_2^2}\right]^{\lambda_s+s}r^{N+1}, \quad L\leq l_2. \quad (54b)$$

Hence any function $f(r)$, which can be represented as a power series in r, we can expand, in analogy to (5),

$$f(r)\Omega_L^M(\vartheta,\varphi)=\sum K'(1,\mathrm{m})f'(1;r_1,r_2)$$
$$\times\,\Omega_{l_1}^{m_1}(\vartheta_1,\varphi_1)\Omega_{l_2}^{m_2}(\vartheta_2,\varphi_2), \quad (55)$$

where K' is given by (33), or by (37) if normalized surface harmonics are used. For the radial functions we obtain from (34) and (54)

$$f'(1;r_1,r_2)$$
$$=2(-)^\lambda\sum_s\frac{r_1^{l_1+2s}g_s(1;r_2)}{(2l_1+2s+1)!!\,(2s)!!}, \quad r_2>r_1 \quad (56)$$

(for the double factorials see (43) of I), where

$$g_s(1;r_2)=r_2^{-1-l_s}\left(\frac{1}{r_2}\frac{d}{dr_2}\right)^{L-l_s}r_2^L$$
$$\times\left[\frac{d^2}{dr_2^2}-\frac{L(L+1)}{dr_2^2}\right]^{\lambda+s}[r_2f(r_2)], \quad L\geq l_2, \quad (57a)$$
$$=r_2^{l_s}\left(\frac{1}{r_2}\frac{d}{dr_2}\right)^{l_s-L}r_2^{-1-L}$$
$$\times\left[\frac{d^2}{dr_2^2}-\frac{L(L+1)}{r_2^2}\right]^{\lambda_s+s}[r_2f(r_2)], \quad L\leq l_2. \quad (57b)$$

Alternatively, the powers of the operator $(r_2^{-1}d/dr_2)$ can be put last, with the result

$$g_s(1,r_2)=\frac{1}{r_2}\left[\frac{d^2}{dr_2^2}-\frac{l_2(l_2+1)}{r_2^2}\right]^{\lambda+s}\frac{1}{r_2^{l_s}}$$
$$\times\left[\frac{1}{r_2}\frac{d}{dr_2}\right]^{L-l_s}[r_2^{L+1}f(r_2)], \quad L\geq l_2, \quad (58a)$$
$$=\frac{1}{r_2}\left[\frac{d^2}{dr_2^2}-\frac{l_2(l_2+1)}{r_2^2}\right]^{\lambda_s+s}r_2^{l_s+1}$$
$$\times\left[\frac{1}{r_2}\frac{d}{dr_2}\right]^{l_s-L}\left[\frac{f(r_2)}{r_2^L}\right], \quad L\leq l_2. \quad (58b)$$

The quadratic operators occurring in (57) and (58) can be factorized, but not expressed as squares; hence the operational factorization of f_l in (48) of I, in terms of Bessel functions of a differential operator, does not appear to have a simple analog in the general case.

The expressions (56)–(58) factorize analytically if f is a spherical Bessel function,

$$f(r)=w_L(kr), \qquad w_L=i_L,\,y_L,\,h_L^{(1)},\,h_L^{(2)} \quad (59)$$

in the usual notation, satisfying

$$[d^2/dr^2-L(L+1)/r^2][rf(r)]=-k^2rf(r). \quad (60)$$

In view of (B 7.2.44–46, 52, 53) and (B 7.11.5–13),

we have

$$(z^{-1}d/dz)^s[z^{-l}w_l(z)] = (-)^s z^{-l-s}w_{l+s}(z),$$
$$(z^{-1}d/dz)^s[z^{1+l}w_l(z)] = z^{1+l-s}w_{l-s}(z), \tag{61}$$

so that (56) and (57) or (58) yield

$$f'(1; r_1, r_2) = 2j_{l_1}(kr_1)w_{l_2}(kr_2), \qquad r_2 \geq r_1. \tag{62}$$

Substituting this into (55) and making use of (33) or (37), we find an expansion equivalent to the expansion theorem for spherical waves derived by Friedman and Russek[7]; apparent discrepancies are due to the differing definitions of the spherical harmonics. For modified spherical Bessel functions, the expressions corresponding to (62) become, in view of (B 2.7.19–22),

$$f = i_l(kr); \quad f' = 2(-)^{\lambda}i_{l_1}(kr_1)i_{l_2}(kr_2),$$
$$f = k_l(kr); \quad f' = 2(-)^{\lambda_3}i_{l_1}(kr_1)k_{l_2}(kr_2), \quad r_2 \geq r_1. \tag{63}$$

It should be borne in mind that the actual signs in the expansion (55) are not necessarily those given in (62) or (63) in view of the changes in sign occurring in (33) and (37).

The algebraic recurrence relations (48)–(52) are not directly applicable to the operational expansion terms (55)–(58); it should, nevertheless, be possible to derive recurrence relations for the functions $f'(1)$, if necessary involving more than three terms. Such relations might lead to a considerable simplification in the evaluation of the radial functions.

APPENDIX: THE UNNORMALIZED 3j-SYMBOLS

The theory of the Wigner $3j$-symbols is well established[3–5] and their values have been extensively tabulated[10,16]; it may therefore appear futile to return to the use of unnormalized harmonics and $3j$-symbols associated with these. However, the use of integers has its advantages, compared with expressions involving square roots, and from this point of view, the symbols U introduced in (25) may be found useful. Their definition is easily generalized to any set of integral or half-integral parameters (j_s, m_s), provided $m_1 + m_2 + m_3 = 0$, all the $(j_s + m_s)$ as well as $2\Lambda = j_1 + j_2 + j_3$ are integers, and the triangular relation (39) holds for the j's. Using the abbreviations (27), we define

$$U\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \sum_{\mu} (-)^{\sigma-\Lambda+\mu}\binom{2\lambda_1}{\lambda_1+\mu} \times \binom{2\lambda_2}{\lambda_2 - m_3 + \mu}\binom{2\lambda_3}{\lambda_3 + m_2 + \mu}, \tag{A1}$$

where

$$\sigma \equiv 2j_1 + m_2 - m_3 \equiv m_1 + 2m_3 \equiv -m_1 - 2m_2 \pmod 2, \tag{A2}$$

and the sum is to be taken over all integral or half-integral values of μ (depending on λ_1) for which all the binomial coefficients are nonzero. The relation of these quantities to Wigner's normalized $3j$-symbols[3–5] is given in (26); like the latter they are invariant under a cyclic permutation of (1, 2, 3), and are multiplied by $(-)^{2\Lambda}$ for a noncyclic permutation or for the transformation $\mathbf{m} \to -\mathbf{m}$. On the other hand, the constant numerator in the sum (A1) destroys the Regge symmetries[17] of the symbols under permutation of the triples $2\lambda_s$, j_s+m_s, j_s-m_s.

Against this loss of symmetry, the definition (A1) has the advantage that all the terms in the sum are integers which, even for $\Lambda = 16$, never exceed 10^9. For $j_1 = j_2 + j_3$, i.e., $\lambda_1 = 0$, the sum reduces to a single term,

$$U\begin{pmatrix} j_2 + j_3, & j_2, & j_3 \\ -m_2 - m_3, & m_2, & m_3 \end{pmatrix} = (-)^{j_2+j_3-m_2+m_3}\binom{2j_2}{j_2+m_2}\binom{2j_3}{j_3+m_3}. \tag{A3}$$

In view of the property of the binomial coefficients

$$\binom{N}{M} = \binom{N-1}{M-1} + \binom{N-1}{M}, \tag{A4}$$

the definition (A1) entails the recurrence formula

$$U\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = U\begin{pmatrix} j_1, & j_2 - \frac{1}{2}, & j_3 - \frac{1}{2} \\ m_1, & m_2 - \frac{1}{2}, & m_3 + \frac{1}{2} \end{pmatrix} - U\begin{pmatrix} j_1, & j_2 - \frac{1}{2}, & j_3 - \frac{1}{2} \\ m_1, & m_2 + \frac{1}{2}, & m_3 - \frac{1}{2} \end{pmatrix}; \tag{A5}$$

the equivalent formula for the normalized $3j$-symbols has been given by Edmonds.[4] Apart from signs, the relation (A5) is similar to that obtaining in Pascal's triangle; and since for $j_3 = 0$ the absolute values of U are binomial coefficients, the whole set of coefficients U can be regarded as a five-dimensional generalization of Pascal's triangle. The numbers can thus be generated by means of (A3) and (A5); for work with electronic computers, this would appear more convenient than the more usual representation of the squares of the normalized symbols as products and ratios of powers of primes.[10,16] A more detailed discussion of the symbols U is given elsewhere.

[17] T. Regge, Nuovo Cimento **10**, 545 (1958).

Two-Center Expansion for the Powers of the Distance Between Two Points*

R. A. SACK
Laboratory of Molecular Structure and Spectra, University of Chicago, Chicago, Illinois
and
Department of Mathematics, Royal College of Advanced Technology, Salford, England†
(Received 16 August 1963)

The powers r^n of the distance between two points specified by spherical polar coordinates relating to two different origins, or of the modulus of the sum of three vectors, are expanded in spherical harmonics of the angles. The radial factors satisfy simple partial differential equations, and can be expressed in terms of Appell functions F_4, and Wigner or Gaunt's coefficients. In the overlap region, first discussed by Buehler and Hirschfelder, the expressions are valid for integer values of $n \geq -1$, but in the other regions, for arbitrary n. For high orders of the harmonics, individually large terms in the overlap region may have small resulting sums; as a consequence the two-center expansion is of limited usefulness for the evaluation of molecular integrals. Expansions are also derived for the three-dimensional delta function within the overlap region, and for arbitrary functions $f(r)$, valid outside that region.

1. INTRODUCTION

THE inverse distance between two points Q_1 and Q_2 specified by the polar coordinates $(r_1, \vartheta_1, \varphi_1)$ and $(r_2, \vartheta_2, \varphi_2)$ with respect to a common origin O, is given by the well-known Laplace expansion in powers of $r_</r_>$ and in terms of Legendre polynomials of the mutual direction cosine $(\cos \vartheta_{12})$. For powers other than the inverse first, analogous expansions exist *either* in powers of $r_</r_>$ *or* in $P_l(\cos \vartheta_{12})$; in the former case the angular dependence is given by Gegenbauer polynomials of $(\cos \vartheta_{12})$[1]; in the latter case, the writer has shown in two recent papers that the radial dependence can be expressed by means of Gauss' hypergeometric function[2]

$$F(\alpha, \beta; \gamma; z) = \sum \frac{(\alpha)_w(\beta)_w}{(\gamma)_w w!} z^w; \tag{1a}$$

$$(\alpha)_w = (\alpha; w) = \alpha(\alpha + 1) \cdots (\alpha + w - 1) = \Gamma(\alpha + w)/\Gamma(\alpha). \tag{1b}$$

In many physical problems, it is more convenient to express the positions of Q_1 and Q_2 in spherical polars about two different origins O_1 and O_2 in such a way that the polar axes and the planes defining $\varphi = O$ are kept parallel. If the coordinates of O_2 with respect to O_1 are given by $(r_3 = a, \vartheta_3, \varphi_3)$, expansions for the inverse distance in terms of spherical harmonics of the angles have been given by Carlson and Rushbrooke, by Rose, and by Buehler and Hirschfelder.[3-6] The precise form of the expressions depends on the specific definition of the spherical harmonics; in the present context, the most useful are the unnormalized forms

$$\Theta_l^m(\vartheta, \varphi) = e^{im\varphi} P_l^{|m|}(\cos \vartheta), \tag{2a}$$

$$\Omega_l^m(\vartheta, \varphi) = e^{im\varphi} P_l^m(\cos \vartheta), \tag{2b}$$

and the normalized form

$$Y_l^m(\vartheta, \varphi) = [(2l + 1)(l - m)!/4\pi(l + m)!]^{\frac{1}{2}} \times e^{im\varphi} P_l^m(\cos \vartheta). \tag{2c}$$

Buehler and Hirschfelder[5] consider in detail the case $\vartheta_3 = 0$ and put

$$|Q_1Q_2|^{-1} = \sum B(l_1, l_2, |m|; r_1, r_2, a) \times \Theta_{l_1}^{-m}(\vartheta_1, \varphi_1)\Theta_{l_2}^{m}(\vartheta_2, \varphi_2)$$
$$[l_1, l_2 = 0, 1, \cdots; \quad -l_< \leq m \leq l_<; \quad l_< = \min(l_1, l_2)]. \tag{3}$$

They have shown that the form of the radial func-

* Supported in part by Office of Naval Research Contract Nonr-2121(01). This work was completed at the Theoretical Chemistry Institute, University of Wisconsin, Madison, Wisconsin, supported by National Aeronautics and Space Administration Grant NsG-275-62(4180).

† Permanent address.

[1] L. Gegenbauer, Wien. Sitzung. **70**, 6, 434 (1874); **75**, 891 (1877).

[2] R. A. Sack, J. Math. Phys. **5**, 245 (1964); **5**, 252 (1964). (Hereafter referred to as I and II, respectively.

[3] B. C. Carlson and G. S. Rushbrooke, Proc. Cambridge Phil. Soc. **46**, 215 (1950).

[4] M. E. Rose, J. Math. and Physics **37**, 215 (1958).

[5] R. J. Buehler and J. O. Hirschfelder, Phys. Rev., **83**, 628 (1951); **85**, 149 (1952).

[6] J. O. Hirschfelder, C. F. Curtiss and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954).

tions B differs according to the relative values of r_1, r_2 and $r_3 = a$; there are, in fact, four distinct regions defined by the following inequalities [see Fig. 1(a)]:

$$S_0: |r_1 - r_2| \leq r_3 \leq r_1 + r_2; \quad S_1: r_1 \geq r_2 + r_3, \quad S_2: r_2 \geq r_1 + r_3; \quad S_3: r_3 \geq r_1 + r_2. \tag{4}$$

The same arguments apply to the more general expansions for arbitrary values of the angle ϑ_3 and of the power n of r.

$$r^n = \sum_{l,m} [{}_2R(n; l_1, l_2, l_3, m_1, m_2, m_3; r_1, r_2, r_3) \times \prod_1^3 \Omega_{l_s}^{m_s}(\vartheta_s, \varphi_s)]; \tag{5}$$

if either of the definitions (2a) or (2c) is used for the spherical harmonics, the corresponding radial functions differ from those in (5) by constants only; the subscripts Θ or Y are added to ${}_2R$ in such cases.

For the inverse first power $n = -1$, the functions ${}_2R \equiv {}_2R_i$ vanish in each of the "outer" regions S_i $(i = 1, 2, 3)$ unless

$$l_i = l_j + l_k; \quad l_s \geq 0 \qquad (s = 1, 2, 3), \tag{6a}$$

and

$$m_i + m_j + m_k = 0, \; |m_s| \leq l_s \quad (s = 1, 2, 3); \tag{6b}$$

throughout this paper, (i, j, k) denote permutations of (1, 2, 3). If (6) is satisfied, ${}_2R_i$ consists of a single term,

$${}_2R_i = {}_2K_i(-1; l_i, l_j, l_k, m_i, m_j, m_k) r_j^{l_j} r_k^{l_k} / r_i^{l_i+1}, \tag{7}$$

where the coefficients ${}_2K_i$ can be expressed in terms of Wigner coefficients[3,4] or as ratios of factorials.[5,6]

For the overlap region S_0, Buehler and Hirschfelder[5,6] found an expression for B_0 as a double power series in r_1/a and r_2/a for which they tabulated the coefficients as ratios of integers for $0 \leq m \leq l_1 \leq l_2 \leq 3$. They could not derive a generally valid formula in this region, though in their later paper (second paper of Ref. 5) they gave a (rather cumbersome) generating function for the function B_0.

The aim of the present paper is to derive generally valid expressions for B_0 or ${}_2R_i$ in all the regions; but for the sake of greater symmetry, the vector $\mathbf{r} = (r, \vartheta, \varphi)$ in (5) is to be understood to mean, not the vector Q_1Q_2, i.e. $\mathbf{r}_2 + \mathbf{r}_3 - \mathbf{r}_1$, but the vector sum $\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3$; the corresponding radial functions ${}_2R$ differ only by the factor $(-)^{l_1}$.[3] As in I and II, the functions are derived as solutions of sets of

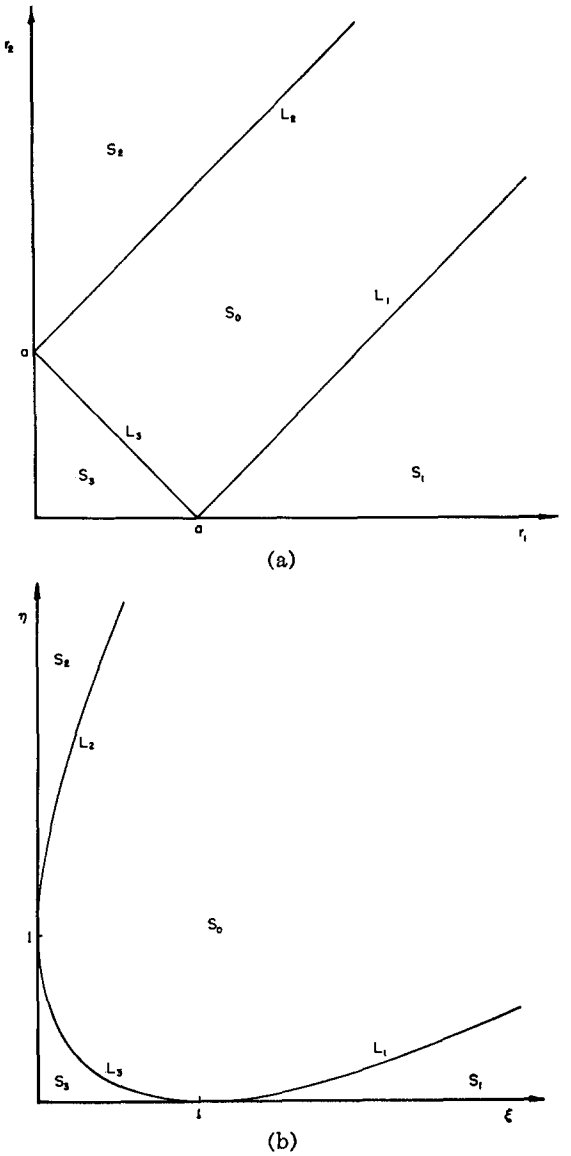


FIG. 1. The four regions S_0, S_1, S_2, S_3 and their boundaries. (a) as functions of r_1 and r_2; (b) as functions of ξ and η.

partial differential equations; they can be expressed in terms of the Appell functions F_4, which form a generalization to two variables of the hypergeometric function (1):

$$F_4(\alpha, \beta; \gamma, \gamma'; \xi, \eta) = \sum \frac{(\alpha)_{u+v}(\beta)_{u+v}}{(\gamma)_u(\gamma')_v u!\, v!} \xi^u \eta^v, \tag{8}$$

summed over all nonnegative values of u and v. The theory of these functions is given in detail in the monographs by Appell and Kampé de Fériet[7,8];

[7] P. Appell, "Sur les fonctions hypergéometriques de plusieurs variables" in *Mémorial des Sciences Mathématiques*, Fasc. 3, (Gauthier-Villars, Paris, 1925).

[8] P. Appell and J. Kampé de Fériet, *Fonctions hypergéometriques et hypersphériques, Polynômes d' Hermite* (Gauthier-Villars, Paris, 1926).

most of the relevant formulas are to be found in Chap. 5 of the Bateman Manuscript Project,[9] but for the benefit of the reader, all the formulas utilized in the present paper will be collected in the Appendix. The differential equations do not involve the azimuthal quantum numbers $\mathbf{m}$, and hence the nature of the functions ${}_2R$ does not depend on these numbers; they can only affect the leading coefficients. In the outer regions, these constants can be determined from the results of I and II, and in the inner region S_0, indirectly, by means of certain linear relations between Appell functions along critical lines. They can be expressed, as in II, by means of $3j$-symbols, Wigner coefficients, or integrals of triple products of spherical harmonics (Gaunt's coefficients).[10-11]

It is found that for $n = -1$ the functions ${}_2R \equiv {}_2R_0$ appear in the region S_0 with nonzero coefficients whenever

$$|l_1 - l_2| \leq l_3 \leq l_1 + l_2, \quad l_1 + l_2 + l_3 = \text{even}, \tag{9}$$

so that by confining attention to this case, it would not be possible to determine the leading coefficients in this region, unless at least one of Eqs. (6a) is satisfied. In consequence, the general case (5) is considered from the start; the resulting formulas are valid for arbitrary values of n in the outer regions, but in S_0 only for integer $n \geq -1$. The formulas obtained for S_1, S_2, and S_3 can be put into an operational form which permit generally valid expansions to be derived in these regions for any function of r; this is done in Sec. 4. Within S_0, on the other hand, the Laplacean operator applied to the expansion for r^{-1} does not vanish; this gives an analogous expansion for the three-dimensional Dirac delta function and its derivatives.

2. MATHEMATICAL DERIVATION

The functions r^n satisfy the differential equation

$$\nabla_1^2(r^n) = \nabla_2^2(r^n) = \nabla_3^2(r^n) = n(n+1)r^{n-2}, \tag{10}$$

which, when substituted into (5), yields

$$\left[\frac{\partial^2}{\partial r_s^2} + \frac{2}{r_s}\frac{\partial}{\partial r_s} - \frac{l_s(l_s+1)}{r_s^2}\right] {}_2R(n, 1, \mathbf{m})$$

$$= \text{invariant} \quad (s = 1, 2, 3), \tag{11a}$$

$$= n(n+1)\, {}_2R(n-2, 1, \mathbf{m}). \tag{11b}$$

Furthermore, all the ${}_2R$ are homogeneous functions of the variables r_s of degree n, and in the region S_i they are regular as r_j and r_k tend to zero; hence they must be of the form

$${}_2R_i = r_j^{l_j} r_k^{l_k} r_i^{n-l_j-l_k} G_i(n, 1, \mathbf{m}; r_j/r_i, r_k/r_i), \tag{12}$$

where

$$G_i(n, 1, \mathbf{m}) = \sum C_{\mu\nu}(n, 1, \mathbf{m})(r_j/r_i)^\mu (r_k/r_i)^\nu. \tag{13}$$

Substitution of (11) into (12) and (13) leads to the recurrence relations

$$(\mu+2)(2l_j+\mu+3)C_{\mu+2,\nu}(n)$$

$$= (\nu+2)(2l_k+\nu+3)C_{\mu,\nu+2}(n), \tag{14a}$$

$$= (n+1+l_i-l_j-l_k-\mu-\nu)$$

$$\times (n-l_i-l_j-l_k-\mu-\nu)C_{\mu\nu}(n), \tag{14b}$$

$$= n(n+1)C_{\mu\nu}(n-2). \tag{14c}$$

This defines G_i as an Appel function F_4 in the variables

$$\xi = r_j^2/r_i^2, \qquad \eta = r_k^2/r_i^2: \tag{15}$$

$$G_i(n, 1, \mathbf{m}) = K_i(n, 1, \mathbf{m})F_4(\Lambda - \tfrac{1}{2}n,$$

$$\lambda_i - \tfrac{1}{2} - \tfrac{1}{2}n;\ l_j + \tfrac{3}{2},\ l_k + \tfrac{3}{2};\ \xi,\ \eta), \tag{16}$$

where we abbreviate

$$\Lambda = \tfrac{1}{2}(l_1 + l_2 + l_3), \quad \lambda_s = \Lambda - l_s$$

$$(s = 1, 2, 3). \tag{17}$$

In view of (11a) and (12), it is easily shown that the function G_i satisfies the set of differential equations of Appell's function (A2) (see Appendix), with the variables and parameters defined in (15)–(17). Hence, according to (A3), the complete set of solutions satisfying the differential equations for ${}_2R(n, 1, \mathbf{m})$ becomes

$$\Psi_{i0} = r_i^n (r_j/r_i)^{-l_j-1}(r_k/r_i)^{-l_k-1} F_4(-\Lambda - \tfrac{1}{2}n - \tfrac{3}{2},$$

$$-\lambda_i - 1 - \tfrac{1}{2}n;\ \tfrac{1}{2} - l_j,\ \tfrac{1}{2} - l_k;\ \xi,\ \eta), \tag{18a}$$

$$\Psi_{ii} = r_i^n (r_j/r_i)^{l_j}(r_k/r_i)^{l_k} F_4(\Lambda - \tfrac{1}{2}n,\ \lambda_i - \tfrac{1}{2}n - \tfrac{1}{2};$$

$$\tfrac{3}{2} + l_j,\ \tfrac{3}{2} + l_k;\ \xi,\ \eta), \tag{18b}$$

$$\Psi_{ij} = r_i^n (r_j/r_i)^{-l_j-1}(r_k/r_i)^{l_k} F_4(\lambda_j - \tfrac{1}{2}n - \tfrac{1}{2},$$

$$-\lambda_k - 1 - \tfrac{1}{2}n;\ \tfrac{1}{2} - l_j,\ \tfrac{3}{2} + l_k;\ \xi,\ \eta), \tag{18c}$$

$$\Psi_{ik} = r_i^n (r_j/r_i)^{l_j}(r_k/r_i)^{-l_k-1} F_4(\lambda_k - \tfrac{1}{2}n - \tfrac{1}{2},$$

$$-\lambda_j - 1 - \tfrac{1}{2}n;\ \tfrac{3}{2} + l_j,\ \tfrac{1}{2} - l_k;\ \xi,\ \eta). \tag{18d}$$

[9] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953). Formulas in this work are directly referenced by the prefix B.

[10] See Refs. 3–5 and 10 of II.

[11] J. A. Gaunt, Phil. Trans. Roy. Soc. **A228**, 157 (1929).

Here the first subscript in notation Ψ_{it} indicates which radius r_i occurs in the denominator of the definitions (15) for ξ and η, and the second subscript shows that Ψ_{it}/r_t^n becomes singular as $r_t \to 0$; if $t = 0$, this ratio becomes singular, whichever radius tends to zero. Further we denote the function ${}_2R$ in the region S_w by ${}_2R_w$ as in (12), and the coefficients of (18) in the expression for ${}_2R_w$ by K_{wit}:

$$ {}_2R_w(n, \mathbf{l}, \mathbf{m}) = \sum_t K_{wit}(n, \mathbf{l}, \mathbf{m})\Psi_{it}(n, \mathbf{l}). \tag{19} $$

In view of (A1) the Appell functions in the outer regions are convergent only if $i = w$, and the regularity of ${}_2R_i$ for small values of r_j and r_k requires that the solution is of the form given by (12) and (16),

$$ K_{iit} \equiv 0, \qquad t \neq i; \qquad K_{iii} = K_i. \tag{20} $$

In the region S_0, the series are always divergent unless they terminate; in those cases in which (18) leads to useful expansions, the choice of i is somewhat arbitrary. The nature of the functions Ψ of (19) being known from (18), it remains to calculate the coefficients K_i and K_{0it}.

To determine K_i in S_i we provisionally combine $\mathbf{r}_j + \mathbf{r}_k$ to a vector $(r_{jk}, \vartheta_{jk}, \varphi_{jk})$; then according to (19) of I and the addition theorem for the $P_l(\cos\vartheta_{12})$ [(B 3.11.2)], we have

$$ r^n = \sum_{l,m} \left[r_i^{n-l} r_{jk}^l (-)^{l+m} \Omega_l^{-m}(\vartheta_i, \varphi_i)\Omega_l^m(\vartheta_{jk}, \varphi_{jk}) \right. $$
$$ \left. \times \frac{(-\frac{1}{2}n)_l}{(\frac{1}{2})_l} F\left(l - \tfrac{1}{2}n, -\tfrac{1}{2} - \tfrac{1}{2}n; \tfrac{3}{2} + l; \frac{r_{jk}^2}{r_i^2}\right) \right]. \tag{21} $$

This expression involves r_{jk} only through the solid harmonics $r_{jk}^l\Omega_l^m$ and through positive even powers $r_{jk}^{2\nu}$; both factors are regular functions of r_j and r_k. If these products, in turn, are expanded in spherical harmonics of (ϑ_j, φ_j) and (ϑ_k, φ_k), we see from (5) and (32)–(34) of II that the lowest power 2ν which contributes to terms for which $l_j + l_k - l = 2\lambda$ occurs for $\nu = \lambda$, irrespective of n. We have, from (5) and (30) of II,

$$ r_{jk}^{l+2\lambda}\Omega_l^m(\vartheta_{jk}, \varphi_{jk}) = (-)^m \sum r_j^{l_j} r_k^{l_k} \Omega_{l_j}^{m_j}(\vartheta_j, \varphi_j)\Omega_{l_k}^{m_k}(\vartheta_k, \varphi_k) $$
$$ \times \frac{(\frac{1}{2}; l_j + l_k + 1)\lambda!}{(\frac{1}{2}; l_j)(\frac{1}{2}; l_k)} I_\Omega \begin{bmatrix} l & l_j & l_k \\ m & -m_j & -m_k \end{bmatrix} + \cdots, \tag{22} $$

where I_Ω is Gaunt's coefficient[11]

$$ I_\Omega \begin{bmatrix} l & l' & l'' \\ m & m' & m'' \end{bmatrix} = \int_{-1}^{1} P_l^m(x) P_{l'}^{m'}(x) P_{l''}^{m''}(x)\, dx. \tag{23} $$

If we put $l = l_i$, $m = -m_i$, $\lambda = \lambda_i$, and use the abbreviations (17), the constant $K_i(n, \mathbf{l}, \mathbf{m})$ in (16) is found, from (21)–(23) and (1),

$$ K_i(n, \mathbf{l}, \mathbf{m}) = (-)^{l_i} \frac{(l_i + \frac{1}{2})(-\frac{1}{2}n; \Lambda)(-\frac{1}{2} - \frac{1}{2}n; \lambda_i)}{(\frac{1}{2}; l_i)(\frac{1}{2}; l_k)} $$
$$ \times I_\Omega \begin{bmatrix} l_i & l_j & l_k \\ -m_i & -m_j & -m_k \end{bmatrix}. \tag{24} $$

The radial functions ${}_2R_i$ in the outer regions are thus completely determined by (12), (16), and (24). The corresponding expressions in the overlap region S_0 can be obtained by means of the linear relation (A8) between the four solutions (A3) of the differential equations (A2) on the critical lines (A4) (see Fig. 1). These lines correspond exactly to the boundaries L_i separating the regions S_i from S_0, but the ${}_2R_w$ must be brought to a common set of variables before (A8) can be applied. In (18) we can transform Ψ_{ij}, on which ${}_2R_i$ solely depends, into a linear combination of Ψ_{ii} and Ψ_{ij} by means of (A6), but the resulting series are in general divergent. The only cases in which (A6) leads to an expression which can be usefully interpreted without recourse to contour integration, are those in which the initial series terminates, i.e., where α or β is a nonpositive integer; then (A6) shows that one of the series in the new variables has zero coefficient and the other terminates. Applying this argument to the set (18), we find we can deal with two cases:

(A) n is a nonnegative even integer; then r^n is analytic throughout and can be represented by a finite expansion common to all regions; the value of i in (18) and (19) is immaterial, and the result can be expressed in a form involving only positive powers of the r_s.

(B) n is an odd integer ≥ -1; then

$$ \Psi_{ij} = T_{ji}\Psi_{ij}; $$
$$ T_{ji} = (-)^{\frac{1}{2}+\frac{1}{2}n-\lambda_j} \frac{\Gamma(\frac{3}{2} + l_i)\Gamma(\frac{1}{2} + l_j)}{\Gamma(\Lambda - \frac{1}{2}n)\Gamma(2 + \frac{1}{2}n + \lambda_k)}. \tag{25} $$

Now since ${}_2R_0 = {}_2R_s$ on $L_s (s = 1, 2, 3)$, the coefficients K_{0it} in (19) can be determined from (24) and (A8) leading to

$$ K_{0ii} = \tfrac{1}{2}K_i, \quad K_{0ij} = \tfrac{1}{2}K_{iij} = \tfrac{1}{2}K_j/T_{ji}, $$
$$ K_{0ik} = \tfrac{1}{2}K_{kik} = \tfrac{1}{2}K_k/T_{ki}, \tag{26} $$
$$ K_{0i0} = -\tfrac{1}{2}K_i $$
$$ \times \frac{\Gamma(\frac{3}{2} + l_j)\Gamma(\frac{3}{2} + l_k)\Gamma(1 + \frac{1}{2}n - \Lambda)\Gamma(\frac{3}{2} + \frac{1}{2}n - \lambda_i)}{\Gamma(\frac{1}{2} - l_j)\Gamma(\frac{1}{2} - l_k)\Gamma(2 + \frac{1}{2}n + \lambda_i)\Gamma(\frac{5}{2} + \frac{1}{2}n + \Lambda)}, \tag{27a} $$

$$= (-)^{l_i+1}\tfrac{1}{2}(l_i + \tfrac{1}{2}) \times \frac{(\frac{1}{2}; l_j + 1)(\frac{1}{2}; l_k + 1)}{(1 + \frac{1}{2}n; \lambda_i + 1)\left(\frac{3}{2} + \frac{n}{2}; \Lambda + 1\right)} \times I_\Omega\begin{bmatrix} l_i & l_j & l_k \\ -m_i & -m_j & -m_k \end{bmatrix}. \quad (27b)$$

The expression (27a) is meaningless $(0 \cdot \infty)$ if $\lambda_i > \frac{1}{2} + \frac{1}{2}n$; nevertheless (27b) is valid for all values of $\mathbf{l}$ and $\mathbf{m}$ satisfying (6b) and (9); the result could first be derived for n raised by a sufficiently large even number for this difficulty to disappear, and then be extended by repeated application of (11b).

3. DISCUSSION OF RESULTS

As in II, it is convenient to factorize the expression for the functions ${}_2R$ in (5) in the form

$${}_2R(n, \mathbf{l}, \mathbf{m}; r_1, r_2, r_3) = {}_2K'(\mathbf{l}, \mathbf{m})\ {}_2R'(n, \mathbf{l}; r_1, r_2, r_3), \quad (28)$$

where the constant ${}_2K'$ is independent of n and the values of r_s and comprises the complete dependence on $\mathbf{m}$. The selection preferred by the writer is

$${}_2K'(\mathbf{l}, \mathbf{m}) = (l_1 + \tfrac{1}{2})(l_2 + \tfrac{1}{2})(l_3 + \tfrac{1}{2}) \times I_\Omega\begin{bmatrix} l_1 & l_2 & l_3 \\ -m_1 & -m_2 & -m_3 \end{bmatrix}, \quad (29)$$

where the I_Ω are Gaunt's coefficients[11] defined in (23), or if the unnormalized $3j$ symbols defined in (25) and (29) of II, and the abbreviations (17) are used,

$${}_2K'(\mathbf{l}, \mathbf{m}) = 2(-)^\Lambda \frac{\Lambda!}{(2\Lambda + 1)!} \times \prod_{s=1}^{3}\left[\frac{(l_s - m_s)!}{\lambda_s!}(l_s + \tfrac{1}{2})\right] \times U\begin{bmatrix} l_1 & l_2 & l_3 \\ -m_1 & -m_2 & -m_3 \end{bmatrix}. \quad (30)$$

The second factor ${}_2R'$ in (28) differs according to the region S_w; in the "outer" regions S_i we obtain, from (12), (16), (17), and (24),

$${}_2R'_i(n, \mathbf{l}; r_i, r_j, r_k) = (-)^{l_i} \times \frac{(-\frac{1}{2}n; \Lambda)(-\frac{1}{2} - \frac{1}{2}n; \lambda_i)}{(\frac{1}{2}; l_j + 1)(\frac{1}{2}; l_k + 1)}\left(\frac{r_j}{r_i}\right)^{l_j}\left(\frac{r_k}{r_i}\right)^{l_k} r_i^n \times F_4(\Lambda - \tfrac{1}{2}n, \lambda_i - \tfrac{1}{2} - \tfrac{1}{2}n; l_j + \tfrac{3}{2}, l_k + \tfrac{3}{2}; r_j^2/r_i^2, r_k^2/r_i^2). \quad (31)$$

In the overlap region S_0, the expressions for ${}_2R'$ are valid, according to the discussion of the previous chapter, only if n is an integer ≥ 1; two cases are to be distinguished:

(A) n even ≥ 0; then ${}_2R$ is the same expression in all four regions,

$${}_2R'_1 = {}_2R'_2 = {}_2R'_3 = {}_2R'_0. \quad (32)$$

(B) n odd ≥ -1; then according to (18), (19), (26), and (27),

$${}_2R'_0 = \tfrac{1}{2}({}_2R'_1 + {}_2R'_2 + {}_2R'_3) - \tfrac{1}{2}(-)^{l_i} \times \frac{(\frac{1}{2}; l_j)(\frac{1}{2}; l_k)}{(1 + \frac{1}{2}n; \lambda_i + 1)(\frac{3}{2} + \frac{1}{2}n; \Lambda + 1)} \times r_i^n\left(\frac{r_i}{r_j}\right)^{l_j+1}\left(\frac{r_i}{r_k}\right)^{l_k+1} F_4\left(-\tfrac{3}{2} - \tfrac{1}{2}n - \Lambda, -1 - \tfrac{1}{2}n - \lambda_i; \tfrac{1}{2} - l_j, \tfrac{1}{2} - l_k; \frac{r_j^2}{r_i^2}, \frac{r_k^2}{r_i^2}\right). \quad (33)$$

Equations (31) and (33) show that for odd n those functions which represent ${}_2R'$ in the outer regions appear with half their coefficient in the overlap region and vanish in the other outer regions; the last term in (33), which is specific to S_0, could equally well be expressed in terms of Appell functions with r_j or r_k in the denominator of the arguments, the results being a polynomial in each case.

If the spherical harmonics in (5) are given in their normalized form (2c), the corresponding radial functions ${}_2R_Y$ factorize, in analogy to (28), into functions ${}_2R'$, which are the same as in (31) and (33), and constants ${}_2K'_Y$ given, in view of (2) and (29), (30),

$${}_2K'_Y = (-)^{m_1}4\pi^2(l_1, -m_1\ |Y^{m_3}_{l_3}|\ l_3, m_3), \quad (34a)$$

$$= 2\pi^{\frac{3}{2}}[(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)]^{\frac{1}{2}} \times \begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{bmatrix}\begin{bmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{bmatrix}, \quad (34b)$$

in terms of integrals of products of three harmonics over the unit sphere or of (normalized) $3j$-symbols [cf. (37), (38) and Ref. 10 of II].

The main application of the expressions derived in this paper is likely to be the evaluation of integrals for the interaction between two "charge" distributions referred to different origins and interacting with a negative power of the distance,

$$\iint \rho_1(\mathbf{r}_1)\rho_2(\mathbf{r}_2)\ |Q_1Q_2|^n\ d^3\mathbf{r}_1\ d^3\mathbf{r}_2. \quad (35)$$

If the functions ρ are expanded in spherical harmonics,

$$\rho_u = \sum W_u(l_u, m_u; r_u)\Omega^{m_u}_{l_u}(\vartheta_u, \varphi_u), \quad u = 1, 2, \quad (36)$$

the expansion for $|Q_1Q_2|^n$ is given by (5) and (28)–(34), except for a factor $(-)^{l_1}$ in each term, as discussed in the introduction and by Carlson and Rushbrooke.[3] The orthogonality of the functions leads to straightforward integrations over the angles for common values of l_1, m_1, l_2, m_2, and the results have to be summed over all compatible values of l_3. The spherical harmonics of ϑ_3 and φ_3 are best left unnormalized, even if the expansion (36) is given in normalized harmonics; in particular, for the case considered by Buehler and Hirschfelder,[5] $\vartheta_3 = 0$, we have $\Omega = 1$ $(m_3 = 0)$, $\Omega = 0$ $(m_3 \neq 0)$.

For $n = -1$, the results in the outer regions have been known from previous work[3-6]; for other negative integer values of n and $\vartheta_3 = 0$, an expansion has been derived by Prigogine[12] by means of the appropriate Gegenbauer polynomials,[1] which were then reexpanded in terms of spherical harmonics; the resulting expressions are valid in the region S_3 only, though this limitation has not been noticed by Prigogine. The complete analysis of the expansions for general values of n and $\vartheta_3 = 0$, which implies a summation over l_3 in (5), leads to expressions for the radial factors which involve more complicated functions than the Appell polynomials used in the present paper, and for this reason is not discussed here.

The most important case considered is $n = -1$, for which we obtain, for ${}_2R'$ in (31) and (33),

$$ {}_2R'_i = (-)^{l_i} \times \frac{4(2l_i - 1)!!}{(2l_j + 1)!!\,(2l_k + 1)!!} \frac{r_j^{l_j} r_k^{l_k}}{r_i^{l_i+1}} \delta_{l_i, l_j+l_k}, \tag{37} $$

in agreement with previous work,[3-6] and for the overlap region S_0,

$$ \begin{aligned} {}_2R'_0 &= \tfrac{1}{2}({}_2R'_1 + {}_2R'_2 + {}_2R'_3) - (-)^{l_3} \\ &\times \frac{(2l_1 - 1)!!\,(2l_2 - 1)!!}{(2\lambda_3 + 1)!!\,2^{\Lambda}(\Lambda + 1)!} \left(\frac{a}{r_1}\right)^{l_1+1} \left(\frac{a}{r_2}\right)^{l_2+1} a^{-1} \\ &\times F_4(-1 - \Lambda, -\tfrac{1}{2} - \lambda_3; \tfrac{1}{2} - l_1, \\ &\times \tfrac{1}{2} - l_2; r_1^2/a^2, r_2^2/a^2), \end{aligned} \tag{38} $$

where $(2k)!! = 2^k k!$, $(2k - 1)!! = 2^k(\tfrac{1}{2})_k$ [cf. (43) of I]; the function F_4 in (38) represents a polynomial of degree $\Lambda + 1$ in ξ and η, or $2(\Lambda + 1)$ in (r_1/a) and (r_2/a). By substituting (29), (30), and (38) into (5) with the special value $\vartheta_3 = 0$, using the harmonics Θ of (2) instead of Ω, and summing over l_3, the writer has been able to reproduce and extend the list of coefficients tabulated by Buehler and Hirschfelder.[5,6]

The lack of an expansion for $n < -1$ valid within S_0 is a serious limitation to the applicability of the method to molecular problems; it precludes its use for the evaluation of relativistic corrections to Coulomb energies, for which $n = -2$, or of van der Waals energies $(n = -6)$ for interpenetrating or even closely approaching elongated distributions. The existence of expansions valid in S_0 for fractional n appears doubtful because of the highly complicated branch points of the function ${}_2R'_0$ corresponding to the physical singularity at $\mathbf{r} = 0$. On the other hand, if the relation

$$ \nabla^2(r^{-1}) = -4\pi\delta^3(r), \tag{39} $$

where δ^3 is the three-dimensional Dirac delta function, is applied to (37) and (38), an expansion for δ^3 is obtained, analogous to (5) and (28)–(33), with

$$ {}_2R'_i(\delta, 1) = 0, \qquad i = 1, 2, 3, \tag{40a} $$

$$ \begin{aligned} &{}_2R'_0(\delta, l) \\ &= \frac{(-)^{l_3+1}(2l_1 - 1)!!\,(2l_2 - 1)!!}{\pi(2\lambda_3 - 1)!!\,2^{\Lambda-2}\Lambda!} \left(\frac{a}{r_1}\right)^{l_1+1} \left(\frac{a}{r_2}\right)^{l_2+1} a^{-3} \\ &\times F_4(-\Lambda, \tfrac{1}{2} - \lambda_3; \tfrac{1}{2} - l_1, \tfrac{1}{2} - l_2; r_1^2/a^2, r_2^2/a^2). \end{aligned} \tag{40b} $$

In contrast to (31) and (33), the functions ${}_2R'(\delta)$ are discontinuous along the boundaries L_i; hence, although the Laplacean operator could, in turn, be applied to δ^3, any integral making use of such an expansion for $\nabla^2(\delta^3)$ would have to be supplemented by line integrals taken along the L_i, and correspondingly for higher derivatives.

Even in such cases, where the complete expansion is known in S_0, its use for the numerical evaluation of integrals may give rise to considerable difficulties. The joint degree in r_1 and r_2 of the terms in ${}_2R'_0$,

$$ w = -2 - l_1 - l_2 + \mu + \nu, \tag{41} $$

may be positive as well as negative; on the other hand, for large values of $(r_1 + r_2)$, the functions cannot increase faster than with this sum raised to the nth power. Hence for $n = -1$, all those terms in a given ${}_2R'_0$ with a constant value of $w \geq 0$ must contain the factor $(r_1 - r_2)^{w+1}$, which, in view of (4), remains bounded. If, therefore, an attempt is made to evaluate the integrals in (35) and (36) term by term over the expansions for $1/r$ in (12) and (13), we obtain repeated integrals of

[12] I. Prigogine, *The Molecular Theory of Solutions* (North-Holland Publishing Company, Amsterdam, 1957).

the form

$$a^{l_1+l_2+1-\mu-\nu}C_{\mu\nu}\int r_1^{1-l_1+\mu}W_1(l_1, m_1; r_1)\,dr_1 \times \int r_2^{1-l_2+\nu}W_2(l_2, m_2; r_2)\,dr_2, \qquad (42)$$

with limits corresponding to the boundaries of S_0. These terms are likely to be largest for large μ and ν, but add up to a small sum when summed over constant values of w, thereby reducing the accuracy of any numerical method employed. To avoid this difficulty, we could first calculate ${}_2R_0'(-1, 1)$ over a grid in S_0, and evaluate the integrals by a suitable two-dimensional quadrature formula. This is bound to be more cumbersome than the repeated integration in (42), and also necessitates knowledge of recurrence formulas by which R_0 can be computed for large l from values with small l without loss of accuracy; the writer has been unable to derive such recurrence formulas, not only those involving three functions as suggested by Appell,[7,8] but even numerically useful formulas involving four or more terms.

The usefulness of the two-center expansion for molecular integrals would thus appear limited to the following special cases:

(a) The expansion for ρ_1 and ρ_2 only extend to small values of l, i.e., the change distributions are atomic (Coulomb integrals). For this case, other methods are available, but the present approach seems to be competitive in simplicity and efficiency.

(b) Compared with the distance $r_3 = a$, ρ_1 and ρ_2 are sufficiently concentrated so that the integrand becomes negligible outside the region S_3. In this case, the two-center expansion is the most convenient method for the evaluation of the integrals; its usefulness could be increased considerably by numerical methods for the approximate evaluation of small, but not negligible contributions from the region S_0.

(c) The functions ρ_1 and ρ_2 are of such a nature that the integrals over S_0 of their products with the ${}_2R_0'$ can be evaluated analytically; this approach again necessitates the establishment of recurrence relations, in this case for the integrals. For exponential functions ρ, this method is to be treated in a separate paper.

In a recent paper, Fontana[13] has sketched a two-center expansion analogous to (27a) of I, which is independent of the region S_i, but introduces powers of $(r_1^2 + r_2^2 + r_3^2)$ in the denominator. The explicit formulas are not given by Fontana, and for the reasons discussed at the end of Sec. 3 of II, the writer considers that the expansion involves functions of greater complexity than those considered in the present series of papers.

More recently, Chiu[14] has derived some of the results of this paper by means of irreducible tensor algebra. Chiu also considers cases for which the functions depend on the angles of $\mathbf{r}_2 + \mathbf{r}_3 - \mathbf{r}_1$ provided $\vartheta_3 = 0$; a complete analysis of such cases would require the use of $6j$-symbols and has been purposely postponed by the writer.

4. AN EXPANSION THEOREM FOR ARBITRARY FUNCTIONS $f(r)$

As in I and II, the formula (31) for the functions ${}_2R_i'$ in the outer regions, though not (33) for ${}_2R_0'$, can be put in an operational form. For the factor in the general term of (31) and (8) which depends on n, we obtain

$$(-\tfrac{1}{2}n; \Lambda+u+v)(-\tfrac{1}{2}-\tfrac{1}{2}n; \lambda_i+u+v) r_i^{n-l_i-l_k-2u-2v} = \frac{(-r_i)^{l_i}}{2^{l_i+l_k+2u+2v}}\left(\frac{1}{r_i}\frac{\partial}{\partial r_i}\right)^{l_i}\frac{1}{r_i}\left(\frac{\partial}{\partial r_i}\right)^{2\lambda_i+2u+2v} r_i^{n+1}. \qquad (43)$$

Hence if we expand any function $f(r)$ which can be represented as a power series in r, we obtain in S_i in analogy to (5), (28), and (29),

$$f(r) = \sum [{}_2f' \cdot {}_2K'(1, \mathbf{m}) \cdot \prod \Omega_{l_s}^{m_s}(\vartheta_s, \varphi_s)], \qquad (44)$$

where

$${}_2f_i'(1; r_s) = \sum_{u,v} \frac{4r_i^{l_i} r_j^{l_j+2u} r_k^{l_k+2v}}{(2u)!!\,(2v)!!\,(2l_i+2u+1)!!\,(2l_k+2v+1)!!} \times \left(\frac{1}{r_i}\frac{\partial}{\partial r_i}\right)^{l_i}\frac{1}{r_i}\left(\frac{\partial}{\partial r_i}\right)^{l_j+l_k-l_i+2u+2v} [r_i f(r_i)], \qquad (45)$$

or using modified spherical Bessel functions $i_l(x)$,

$${}_2f_i' = 4r_i^{l_i}\left(\frac{1}{r_i}\frac{\partial}{\partial r_i}\right)^{l_i}\frac{1}{r_i}\frac{i_{l_j}(r_j\partial/\partial r_i) i_{l_k}(r_k\partial/\partial r_i)}{(\partial/\partial r_i)^{l_i}} \times [r_i f(r_i)]. \qquad (46)$$

As in I and II, this expression factorizes if $f(r)$ is a spherical Bessel function,

$$f(r) = w_0(Kr), \qquad w = j, y, h^{(1)}, h^{(2)}; \qquad (47)$$

then in view of (56)–(60) of I,

$${}_2f_i' = (-)^{\Lambda} j_{l_j}(Kr_j) j_{l_k}(Kr_k) w_{l_i}(Kr_i), \qquad (48)$$

[13] P. R. Fontana, J. Math. Phys. 2, 825 (1961).

[14] Y. N. Chiu, J. Math. Phys. (to be published).

and for the modified Bessel function $f(r) = i_0(Kr)$,

$$ {}_2f_i' = \prod_{s=1}^{3} i_{l_s}(Kr_s), \tag{49a} $$

and for the modified Bessel function of the second kind $f(r) = k_0(Kr)$,

$$ {}_2f_i' = (-)^{l_i} i_{l_i}(Kr_i) i_{l_k}(Kr_k) k_{l_i}(Kr_i). \tag{49b} $$

For j_0 and i_0, which are even functions of the argument, the expansion is invariant on permuting (i, j, k) and is therefore also valid in S_0; for the other Bessel functions, the writer has been unable to find the expression appropriate to S_0.

ACKNOWLEDGMENTS

The writer wishes to thank Professor J. O. Hirschfelder, Dr. M. J. M. Bernal, and Dr. Y. N. Chiu for stimulating discussions and advice.

APPENDIX: PROPERTIES OF THE APPELL FUNCTIONS F_4

Appell's function F_4 as defined in (8) represents a polynomial in ξ and η of degree $|\alpha|$ or $|\beta|$ if α or β is a nonpositive integer. In all other cases, F_4 is an infinite series which converges for values ξ and η such that

$$ |\xi|^{\frac{1}{2}} + |\eta|^{\frac{1}{2}} < 1; \tag{A1} $$

for other values of the variables, the function can be defined in terms of contour integrals (cf. B 5.7.44 and B 5.8.9, 13). They satisfy the pair of differential equations (B 5.9.12),

$$ \begin{aligned} &\xi^2 \frac{\partial^2 Z}{\partial \xi^2} + 2\xi\eta \frac{\partial^2 Z}{\partial \xi\, \partial \eta} + \eta^2 \frac{\partial^2 Z}{\partial \eta^2} \\ &\quad + (\alpha + \beta + 1)\left(\xi \frac{\partial Z}{\partial \xi} + \eta \frac{\partial Z}{\partial \eta}\right) + \alpha\beta Z \\ &\quad = \xi \frac{\partial^2 Z}{\partial \xi^2} + \gamma \frac{\partial Z}{\partial \xi} = \eta \frac{\partial^2 Z}{\partial \eta^2} + \gamma' \frac{\partial Z}{\partial \eta}. \end{aligned} \tag{A2} $$

This set has, in general, four linearly independent solutions (p. 52 of Ref. 8),

$$ \begin{aligned} Z_0 &= \xi^{1-\gamma}\eta^{1-\gamma'} F_4(\alpha + 2 - \gamma - \gamma', \\ &\qquad \beta + 2 - \gamma - \gamma'; 2 - \gamma, 2 - \gamma'; \xi, \eta), \\ Z_i &= F_4(\alpha, \beta; \gamma, \gamma'; \xi, \eta), \\ Z_j &= \xi^{1-\gamma} F_4(\alpha + 1 - \gamma, \beta + 1 - \gamma; 2 - \gamma, \gamma'; \xi, \eta), \\ Z_k &= \eta^{1-\gamma'} F_4(\alpha + 1 - \gamma', \beta + 1 - \gamma'; \gamma, 2 - \gamma'; \xi, \eta); \end{aligned} \tag{A3} $$

but the four independent solutions of systems such as (A3) become linearly dependent with constant coefficients on certain critical lines (Sec. 12 of reference 8). For any function F_4 there exist at least three critical lines[15]

$$ L_i: \xi^{\frac{1}{2}} + \eta^{\frac{1}{2}} = 1; \quad L_j: \xi^{\frac{1}{2}} - \eta^{\frac{1}{2}} = 1; \quad L_k: \eta^{\frac{1}{2}} - \xi^{\frac{1}{2}} = 1, \tag{A4} $$

which form sections of a single parabola

$$ \xi^2 - 2\xi\eta + \eta^2 - 2\xi - 2\eta + 1 = 0. \tag{A5} $$

For variations of ξ and η along L_i, Appell has shown that $Z = Z[\xi, \eta(\xi)]$ taken as a function of ξ satisfies a third-order ordinary differential equation, instead of a fourth-order one, as along an arbitrary line; hence (A2) has only three linearly independent solutions on L_i. For the other lines, this dependence follows from the transformation (B 6.11.9),

$$ \begin{aligned} F_4(\alpha, \beta; \gamma, \gamma'; \xi, \eta) &= \frac{\Gamma(\gamma')\Gamma(\beta - \alpha)}{\Gamma(\gamma' - \alpha)\Gamma(\beta)} (-\eta)^{-\alpha} \\ &\times F_4\left(\alpha, \alpha + 1 - \gamma'; \gamma, \alpha + 1 - \beta; \frac{\xi}{\eta}, \frac{1}{\eta}\right) \\ &+ \frac{\Gamma(\gamma')\Gamma(\alpha - \beta)}{\Gamma(\gamma' - \beta)\Gamma(\alpha)} (-\eta)^{-\beta} \\ &\times F_4\left(\beta, \beta + 1 - \gamma'; \gamma, \beta + 1 - \alpha; \frac{\xi}{\eta}, \frac{1}{\eta}\right), \end{aligned} \tag{A6} $$

and a corresponding transformation to $(1/\xi, \eta/\xi)$. Appell has not explicitly stated the coefficients relating the functions (A3); the writer has been able to deduce them for restricted values of the parameters only. Considering their behavior near (1, 0) and (0, 1), we see that two of the functions are singular in the vanishing variable, and two analytic (for fractional values of γ and γ'); regarded as functions of the other variable, they are essentially hypergeometric series, and since (B 2.1.14)

$$ F(\alpha, \beta; \gamma; 1) = \Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)[\Gamma(\gamma - \alpha) \times \Gamma(\gamma - \beta)]^{-1}, \quad \mathrm{Re}\,(\gamma) > \mathrm{Re}\,(\alpha + \beta), \tag{A7} $$

the only relation with constant coefficients which can hold on the line L_s of (4) is

$$ \begin{aligned} &\frac{\Gamma(\gamma)\Gamma(\gamma')}{\Gamma(\gamma + \gamma' - \alpha - 1)\Gamma(\gamma + \gamma' - \beta - 1)} Z_0 \\ &\quad + \epsilon_{si} \frac{\Gamma(2 - \gamma)\Gamma(2 - \gamma')}{\Gamma(1 - \alpha)\Gamma(1 - \beta)} Z_i \\ &\quad + \epsilon_{sj} \frac{\Gamma(\gamma)\Gamma(2 - \gamma')}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} Z_j \\ &\quad + \epsilon_{sk} \frac{\Gamma(2 - \gamma)\Gamma(\gamma')}{\Gamma(\gamma' - \alpha)\Gamma(\gamma' - \beta)} Z_k = 0, \end{aligned} \tag{A8} $$

[15] P. Appell, J. Maths. Pures Appls., Ser. 3 10, 407 (1884).

where

$$\epsilon_{ss} = 1; \qquad \epsilon_{st} = -1, \qquad s \neq t. \tag{A9}$$

The precise form of (A8) for the lines L_j and L_k follows from that for L_i and (A6). On the other hand, a more careful investigation of the behavior of $F(\alpha, \beta; \gamma; x)$ near $x = 1$ [(B 2.10.1)], shows that (A8) is correct only if all the series terminate which appear with nonvanishing coefficients; otherwise terms of the form $(1 - \xi)^{\gamma-\alpha-\beta}$ enter into (A8) which do not add up to zero.

Appell has also stated (p. 19 of Ref. 8) that any three contiguous functions F_4 satisfy a linear recurrence relation, a total of 28 equations, if only one parameter at a time changes by unity; but the writer has been unable to find the complete set of such relations in the literature or to derive it, and he doubts the validity of Appell's statement.[16]

[16] Professor A. Erdélyi (private communication) has concurred with this opinion.

Explicit Formal Construction of Nonlinear Quantum Fields*

I. E. SEGAL
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 21 June 1963)

It is shown that there can be associated with any given nonlinear relativistic partial differential equation an operator field satisfying the canonical commutation relations, transforming appropriately under the action of the Lorentz group, and propagated in accordance with the given differential equation. This quantization procedure is unique, apart from the scale of the commutators. The treatment is intuitive, but is capable of rigorization in terms of the mathematical theory of analysis in function space (functional integration). The results are in harmony with crucial conventional ones, so far as has been determined, and provide in principle a possible means of computing the collision matrix for particular systems without recourse to perturbation theory.

INTRODUCTION

THE original idea of Dirac for field quantization may be paraphrased, in its application to an equation of the simplest nontrivial relativistic form

$$\Box\phi = \rho(\phi) \quad (\rho \text{ a given function}), \tag{1}$$

as follows. (We preserve the spirit of Dirac but utilize implicitly the later relativistic developments due to Heisenberg and Pauli, which, as first shown by Rosenfeld, are essentially equivalent; this favors succinctness, and shows the germ of the connection between the relativistic and nonrelativistic approaches.)

The "classical field" $\phi(x)$, whose value at each point x is a number, is to be replaced by an operator-valued field $\tilde{\phi}(x)$, the so-called "quantized field." The $\tilde{\phi}(x)$ are determined in the following way. At a particular time, the classical field, or more exactly the Cauchy data thereof, are expanded into a complete orthonormal set of functions $\{f_k\}$ in the Hilbert space of functions of the space variables:

$$\begin{aligned}\phi(\mathbf{x}, t) &\sim \sum_k p_k f_k(\mathbf{x}),\\ (\partial\phi(\mathbf{x}, t)/\partial t) &\sim \sum_k q_k f_k(\mathbf{x})\end{aligned} \qquad (k = 1, 2, \cdots). \tag{2}$$

The coefficients p_k and q_k are then replaced by canonical variables P_k and Q_k:

$$P_k Q_{k'} - Q_{k'} P_k = -i\delta_{kk'} I \ (I = \text{unit variable}).$$

Having done this at some particular time t', the quantized field $\tilde{\phi}(\mathbf{x}, t)$ at other times is determined by the solution of the Cauchy problem

$$\Box\tilde{\phi}(x) = \rho(\tilde{\phi}(x)), \tag{3a}$$

$$\begin{aligned}\tilde{\phi}(\mathbf{x}, t') &= \sum_k P_k f_k(\mathbf{x}),\\ (\partial\tilde{\phi}(x, t)/\partial t)_{t=t'} &= \sum_k Q_k f_k(\mathbf{x}).\end{aligned} \tag{3b}$$

Which time t' was employed was effectively immaterial (in fact, as Dirac later indicated, any spacelike surface could be used in place of the surface $t = t'$); the use of any other time in place of t' amounts merely to a similarity transformation on the field by a fixed (canonical) transformation.

The basis of the idea is basically analogous to the formulation of nonrelativistic quantum mechanics by Heisenberg and Schödinger. Having attained this construction for the quantized field, it could be seen that alternatively the quantized field could be defined by the differential equation (3a) in combination with the "canonical" commutation relations

$$\begin{aligned}[\tilde{\phi}(\mathbf{x}, t), \tilde{\phi}(x', t)] &= 0,\\ [\phi(\mathbf{x}, t), (\partial/\partial t)\phi(\mathbf{x}', t)] &= i\delta(\mathbf{x} - \mathbf{x}').\end{aligned} \tag{4}$$

The field could then be defined quasiaxiomatically, as the presumably unique solution of the differential equation (3a) in conjunction with the commutation relations (4).

While this beautifully simple theory has worked out extremely well in important respects, and is widely regarded as embodying some variety of fundamental physical truth, it has resisted all attempts at being put on a solid foundation. In a mathematical way, the doubtfulness of the theory is revealed in many more or less closely related ways, of which the simplest is probably the *prima facie* meaninglessness of such expressions as $\rho(\tilde{\phi})$, for a nonlinear function [e.g. the cube, $\rho(l) = l^3$], even in the explicitly accessible case in which $\tilde{\phi}$ is a "free field." In other words, it has not been possible

* Research supported in part by the Office of Scientific Research.

to give the nonlinear differential equation (3a) for $\tilde{\phi}$ a definite, effective, mathematical meaning. Even after some 35 years, this might not be conclusive grounds for suspecting the essential meaninglessness of such expressions if there were a physical basis for the definition of $\rho(\tilde{\phi})$; but there is not even a conceptual experiment which might measure $\rho(\tilde{\phi})$, so far as is known. (The point is, such an experiment could measure a linear average of $\tilde{\phi}$ over a space–time region, as Bohr and Rosenfeld showed in detail, but there is no known way to extract a linear average of nonlinear functions of $\tilde{\phi}$ from linear averages of $\tilde{\phi}$ itself.

In an earlier paper,[1] an approach was indicated to the construction of a particular quantum field (without attempting to deal with the axiomatic description of quantum fields in general) which made no use of the notion of local nonlinear function of a quantized field, or of equivalent undefined objects. However, the approach did not appear formally quite as simple as the one indicated above, nor was it mathematically rigorous. The purpose of the present article is to remedy the former deficiency; the latter is to be dealt with elsewhere. Even in its specialization to the linear case, the present method is in some respects simpler and more general than the conventional one, applying quite directly to Maxwell's equations with the Lorentz side condition, for example.

The present article is thus frankly heuristic, and free use will be made of formal measures in infinite-dimensional Euclidean spaces, which have actually no *prima facie* mathematical existence; but we note that a mathematically unexceptionable way of dealing with such matters, on the basis of suitably sophisticated reformulations, is now available (indications of the type of work involved may be seen[2,3]).

1. THE PHASE SPACE OF A CONTINUOUS SYSTEM

The following treatment for scalar relativistic equations is readily adaptable to arbitrary temporally homogeneous equations for fields of even spin. (In the present article only Bose–Einstein fields are considered. There is no fundamental difficulty in treating Fermi–Dirac fields, but the treatment can not be developed intuitively from classical theoretical physics, and the relevant notion of integration is materially different from that appropriate to Bose–Einstein fields [cf. Ref. 4]; the main ideas are well represented by the conjunction of the present ones with those of this reference).

A system described by Eq. (1) has its state determined by a particular solution of (1), so that as a set, its phase space is merely the aggregate of all solutions of (1). (It is physically reasonable and mathematically appropriate to restrict to those solutions which, at any particular time, do not grow rapidly in the space variables near infinity; this can be done, but such aspects play only a minor role in the present intuitive treatment.) Now this phase space is quite analogous to an infinite-dimensional manifold in which there are local linear coordinates, such as may be supplied for example by coordinatizing the manifold by the Cauchy data for any particular solution at a particular time t. This manifold is much more similar to a classical mechanical phase space than might appear, for there is defined on it a fundamental nondegenerate second-order differential form analogous to the "fundamental bilinear covariant" $\sum_i dp_i\, dq_1$ in classical mechanics. Since care should presumably be used in dealing with differential forms in infinite-dimensional spaces, it is useful to be rather explicit, and use such a current mathematical definition of a differential form as a function assigning to each point of the manifold M a skew-symmetric multilinear functional in the "tangent vectors" to the manifold M at the point.

The term "tangent vector" here signifies a vector in the infinite-dimensional linear space providing the best (linear) approximation to M in the vicinity of the given point ϕ. In the present context, this is identifiable as a solution η of the first-order variational equation

$$\Box\eta = \rho'(\phi)\eta, \tag{5}$$

for small displacements from ϕ. Thus a second-order differential form Ω on M depends on the point ϕ, and for each ϕ gives a function $\Omega_\phi(\eta, \eta')$ of the two solutions η, η' of (5) which is bilinear and skew-symmetric.

Such a form Ω may be defined as follows:

$$\Omega_\phi(\eta, \eta') = \int [\eta(\partial\eta'/\partial t) - \eta'(\partial\eta/\partial t)]\, d_3x,$$

the integration being over the space variables for fixed time. The result would appear to depend on the time t, but on differentiation of the right side and making use of (5), it is found actually to be independent of t.

Now a basic property of the fundamental bilinear covariant $\sum_i dp_i\, dq_i$ is that its covariant differential

[1] I. E. Segal, J. Math. Phys. **1**, 468 (1960).
[2] I. E. Segal, Trans. Am. Math. Soc. **81**, 160 (1956).
[3] I. E. Segal, Trans. Am. Math. Soc. **88**, 12 (1958).
[4] I. E. Segal, Ann. Math. **63**, 160 (1956).

vanishes (it is a classical theorem in fact that this property together with nondegeneracy characterizes those forms on an $2n$-dimensional manifold which may be locally expressed as $\sum_i dp_i\, dq_i$). Therefore the question arises as to the vanishing of the covariant differential of the form just defined. Rather than engage in a distracting discussion of this mathematical concept as it applies in the infinite-dimensional case, it may be said that it is virtually obvious that this is the case on expansion of η and η' into a suitable orthonormal basis.

More specifically, let $\{f_k\}$ be a complete orthonormal set in the space $L_2(E_3)$ of square-integrable functions on three-dimensional Euclidean space, with the usual inner product, and for fixed t expand

$$\phi(\mathbf{x}, t) = \sum_k p_k f_k(\mathbf{x}),\ (\partial\phi(\mathbf{x}, t)/\partial t) = \sum_k q_k f_k(\mathbf{x}),$$

$$[p_k = p_k(t);\ q_k = q_k(t)].$$

The p_k and q_k thus form a system of coordinates in M. Now if

$$\eta(\mathbf{x}, t) = \sum_k a_k f_k(\mathbf{x}),$$

$$\partial\eta(\mathbf{x}, t)/\partial t = \sum_k b_k f_k(\mathbf{x})$$

gives the corresponding expansions of two tangent vectors at ϕ, then

$$\int [\eta(\partial\eta'/\partial t) - \eta'(\partial\eta/\partial t)]\, d_3x = \sum_k (a_k b_k' - a_k' b_k).$$

This means precisely that

$$\Omega = \sum_i dp_i\, dq_i. \tag{6}$$

This shows also that Ω is nondegenerate. Furthermore, the independence of t of the expression (6) for Ω shows incidentally that the transformation from the $\{p_k(t);\ q_k(t);\ k = 1, 2, \cdots\}$ to similar variables at another time determines a *contact transformation* on M, i.e., one preserving the form Ω, as is clear otherwise from the fact that this transformation is a time displacement, under which Ω is invariant.

The parallelism with the classical mechanical situation extends in a formal way to the existence of an invariant volume element in phase space analogous to that of Liouville. In a $2n$-dimensional manifold with fundamental form $\Omega = \sum_{i=1}^{n} dp_i\, dq_i$, the invariant volume element is $\Omega^n = \prod_i dp_i\, dq_i$, using the notion of the exterior product of differential forms. In the present infinite-dimensional situation, $\lim_{n\to\infty} \Omega^n = \Omega^\infty$ has no rigorous meaning, but is formally $\prod_i dp_i\, dq_i$. This shows that this Euclidean volume element is left invariant by any contact transformation, for leaving Ω invariant, it leaves also any power of it invariant, by the invariance of the notion of exterior product.

Of course, a major point of formal difference from classical mechanical theory is that the time variable is already effectively included in the present phase space. This is necessarily the case in a relativistic theory, and does not appear in classical mechanics because of the absence of any such nontrivial relativistic system, which is related in turn to the nonexistence of nontrivial finite-dimensional unitary representations of the Lorentz group. The only valid concern need be with the Lorentz invariance of the theory. It is clear that any Lorentz transformation transforms the manifold M into itself. To show that Ω is invariant under any Lorentz transformation, it suffices to check infinitesimal invariance under the Lorentz rotations, since from the formulation of Ω, it is obviously invariant under Euclidean transformations. By a suitable choice of Lorentz frame, the infinitesimal Lorentz transformation may be taken in the form $x(\partial/\partial t) + t(\partial/\partial x)$, and an explicit computation then shows the invariance of Ω.

More specifically, let η_1 and η_2 be the solutions of the first-order variational equations in the vicinity of the solution ϕ of Eq. (1), and let $A = x(\partial/\partial t) + t(\partial/\partial x)$. Under the action of the Lorentz transformation $e^{A\theta}$ generated by A $(-\infty < \theta < \infty)$, given by the equations

$$x' = x \cosh\theta + t \sinh\theta,$$

$$y' = x \sinh\theta + t \cosh\theta,$$

the form Ω is transformed into a new form $\Omega^{(\theta)}$, whose value on the tangent vectors η_1 and η_2 at ϕ is obtained by evaluating $\Omega(\eta_1^{(\theta)}, \eta_2^{(\theta)})$ at the point $e^{A\theta}\phi$ of M, where the $\eta_i^{(\theta)}$ are the transforms of the η_i under the induced action on the space of tangent vectors at ϕ of $e^{A\theta}$. Now the η_i satisfy the equation (5), as do the $\eta_i^{(\theta)}$, with ϕ replaced by $e^{A\theta}\phi$. It follows that $\eta_i^{(\theta)} = e^{A\theta}\eta_i$, which means that

$$\Omega_\phi(\theta)(\eta_1^{(\theta)}, \eta_2^{(\theta)}) = \int_{t'=\text{const}} \left(\eta_1 \frac{\partial\eta_2}{\partial t} - \eta_2 \frac{\partial\eta_1}{\partial t}\right) d_3x'.$$

It suffices to show that the derivative with respect to θ of the latter expression, evaluated at $\theta = 0$, vanishes. This derivative is the sum of three terms, one each involving the action of the infinitesimal Lorentz transformation A on η_1 and η_2, and one arising from the transformation of the volume

element. To take the last term first, for $t' = \text{const}$, $dx \sinh\theta + dt\cosh\theta = 0$, while $dx' = dx\cosh\theta + dt\sinh\theta = (\cosh\theta)^{-1}dx$. Now $d_3x' = dx'\,dy'\,dz'$, while $y' = y$ and $z' = z$, so that the volume element is multiplied by $(\cosh\theta)^{-1}$. Since the derivative of this expression vanishes for $\theta = 0$, the entire term in question vanishes, i.e., there is no infinitesimal contribution from the change in the volume element.

The remaining terms are

$$\Omega(A\eta_1, \eta_2) + \Omega(\eta_1, A\eta_2) = \Omega(A\eta_1, \eta_2) - \Omega(A\eta_2, \eta_1).$$

By direct computation,

$$\Omega(A\eta_1, \eta_2) = \int \left[x\frac{\partial\eta_1}{\partial t}\frac{\partial\eta_2}{\partial t} + t\frac{\partial\eta_1}{\partial x}\frac{\partial\eta_2}{\partial t} - x\eta_2\frac{\partial^2\eta_2}{\partial t} - \eta_2\frac{\partial\eta_1}{\partial x} - t\eta_2\frac{\partial^2\eta_1}{\partial x\,\partial t}\right] d_3x.$$

Using the skew-adjointness of the operator $\partial/\partial x$ in the Hilbert space of functions of the space variables, together with the first-order variational Eq. (5), the expression on the right may also be expressed as

$$\int \left[x\frac{\partial\eta_1}{\partial t}\frac{\partial\eta_2}{\partial t} + t\frac{\partial\eta_1}{\partial x}\frac{\partial\eta_2}{\partial t} + t\frac{\partial\eta_2}{\partial x}\frac{\partial\eta_1}{\partial t} - x\eta_2\rho'(\phi)\eta_1 - \eta_2\frac{\partial\eta_1}{\partial x} - x\eta_2\Delta\eta_1\right] d_3x.$$

All but the last two terms in the foregoing integrand are symmetric between the indices 1 and 2, and thus drop out on subtraction of $\Omega(A\eta_2, \eta_1)$. The remaining two terms give, after this subtraction, all together

$$\int \left[-\eta_2\frac{\partial\eta_1}{\partial x} + \eta_1\frac{\partial\eta_2}{\partial x}(x\eta_2)(\Delta\eta_1) + (x\eta_1)(\Delta\eta_2)\right] d_3x.$$

Using the self-adjointness of Δ and evaluating the commutator of x and Δ as $-2(\partial/\partial x)$, the contribution from the last two terms in this integrand is $-2\int \partial\eta_1/\partial x\ \eta_2\, d_3x$. Using now the skew-adjointness of $(\partial/\partial x)$ again in connection with the first two terms, they provide a contribution to the total integral precisely canceling that of the last two terms.

2. THE QUANTUM FIELD VARIABLES

In ordinary quantum mechanics, one way of constructing variables satisfying the Heisenberg commutation relations is, as indicated in reference 1, to take operators in phase space rather than in physical space. For any classical observable such as a position coordinate, a linear or angular momentum represented by the function f on phase space, the corresponding quantum mechanical variable may be taken as the operator $-i\hbar X_f + \omega(X_f)$, where X_f is the infinitesimal contact transformation with Hamiltonian f, and ω is the action form $\sum_i p_i\, dq_i$. Without going into the details of this construction, we note its presently relevant advantage; the canonical P's and Q's appear on essentially the same footing, so that the formalism is potentially better adapted to the relativistic case, in which there is no Lorentz-invariant separation between the P's and the Q's. (The representation is not irreducible, but this can be taken care of; see below.)

To define the quantum field variables in a related way, choose a time t and an arbitrary real function $f(\mathbf{x})$ of the real space variables, and define a corresponding one-parameter group of transformations $T^s_{f,t}(-\infty < s < \infty)$ on the manifold M as follows: a given element ϕ of M is transformed into that element ϕ' whose Cauchy data at the time t differ from those of ϕ as follows:

$$\phi'(\mathbf{x}, t) = \phi(\mathbf{x}, t); \quad \partial\phi'(\mathbf{x}, t)/\partial t = \partial\phi(\mathbf{x}, t)/\partial t + sf(\mathbf{x}).$$

Let $X_{f,t}$ denote the infinitesimal generator of this one-parameter group; it is the vector field on M assigning to any element ϕ the tangent vector represented by the solution η of the first-order variational equation $\Box\eta = \rho'(\varphi)\eta$, with the Cauchy data $\eta(\mathbf{x}, t) = 0$, $\partial\eta(\mathbf{x}, t)/\partial t = f(\mathbf{x})$. Let $M_{f,t}$ denote the operation of multiplication (of functionals on M) by $\int f(\mathbf{x})\phi(\mathbf{x}, t)\, d_3x$. Now let g be an arbitrary real constant and define

$$R_{f,t} = -i(\tfrac{1}{2}g)X_{f,t} + M_{f,t}.$$

Thus $R_{f,t}$ is a first-order linear differential operator on M. If $f'(\mathbf{x})$ is any other real function of the space variables, it may be verified that

$$[R_{f,t}, R_{f',t}] = 0; \qquad [R_{f,t}, (\partial/\partial t)R_{f',t}] = -ig\int f(\mathbf{x})f'(\mathbf{x})\, d_3x.$$

More specifically, $[X_{f,t}, X_{f',t}] = 0$ since vector translations in the space of Cauchy data (at a fixed time) commute. Also, $[M_{f,t}, M_{f',t}] = 0$ since any two multiplication operators by functions commute. Now $[X_{f,t}, M_{f',t}]$, the commutator of a vector field and a multiplication operator, is the multiplication operator by the result of applying the vector field to the function associated with the original multiplication operator, i.e., by

$$X_{f,t}\left[\int f(\mathbf{x})\phi(\mathbf{x}, t)\, d_3x\right].$$

This vanishes since $X_{f,t}$ does not affect $\phi(\mathbf{x},t)$, but only $\partial\phi(\mathbf{x},t)/\partial t$. From these results follow

directly the vanishing of the commutator $[R_{f,t}, R_{f',t}]$.

To compute $(\partial/\partial t)R_{f,t}$, write $(\partial/\partial t)X_{f,t} = \lim_{h\to 0} h^{-1}[X_{f,t+h} - X_{f,t}]$, and consider that $X_{f,t+h}$ displaces the Cauchy data at time $t + h$ as follows.

$$sX_{f,t+h}\colon \phi(\mathbf{x}, t + h) \to \phi(\mathbf{x}, t + h),$$

$$(\partial\phi/\partial t)(\mathbf{x}, t + h) \to (\partial\phi/\partial t)(\mathbf{x}, t + h) + sf(\mathbf{x}).$$

The corresponding displacement in the Cauchy data at time t may be obtained as follows. Neglecting terms of second and higher order in h, which are irrelevant for the determination of the derivative in question,

$$\phi(\mathbf{x}, t) = \phi(\mathbf{x}, t + h) - h(\partial\phi(\mathbf{x}, t + h)/\partial t)$$
$$\to \phi(\mathbf{x}, t + h) + h(\partial\phi(\mathbf{x}, t + h)/\partial t + sf(\mathbf{x}))$$

or

$$\phi(\mathbf{x}, t) \to \phi(\mathbf{x}, t) - hsf(\mathbf{x}).$$

Similarly,

$$\partial\phi(\mathbf{x}, t)/\partial t) = \partial\phi(\mathbf{x}, t + h)/\partial t - h\partial^2\phi(\mathbf{x}, t + h)/\partial t^2$$
$$\to \partial\phi(\mathbf{x}, t + h)/\partial t + sf(\mathbf{x}) - h\partial^2\phi(\mathbf{x}, t + h)/\partial t^2,$$

since $\partial^2\phi(\mathbf{x}, t + h)/\partial t$ is not displaced by $X_{f,t+h}$, in view of the satisfaction by the displacement η of Eq. (5), and the vanishing at $t + h$ of η. Thus

$$\partial\phi(\mathbf{x}, t)/\partial t \to \partial\phi(\mathbf{x}, t) + sf(\mathbf{x}).$$

It follows that $(\partial/\partial t)X_{f,t}$ may be characterized as the vector field which generates the one-parameter group of transformations in the Cauchy data at time t:

$$\phi(\mathbf{x}, t) \to \phi(\mathbf{x}, t) - sf(\mathbf{x}), \quad \partial\phi(\mathbf{x}, t)/\partial t \to \partial\phi(\mathbf{x}, t)/\partial t.$$

To evaluate $[R_{f,t}, (\partial/\partial t)R_{f',t}]$, note that $X_{f,t}$ and $(\partial/\partial t)X_{f',t}$ commute since they generate vector translations in the space of Cauchy data at time t. Also, the multiplication operators $M_{f,t}$ and $(\partial/\partial t)M_{f',t}$ commute. The only terms contributing to the commutator are then the two involving the brackets of vector fields with multiplication operators, which brackets are multiplication operators by the functions obtained by application of the respective vector fields to the functions associated with the original multiplication operators. The functions involved in the commutator are consequently

$$X_{f,t}\Big[(\partial/\partial t)\int \phi(\mathbf{x}, t)f'(\mathbf{x})\, d_3x\Big]$$

and

$$(\partial/\partial t)X_{f',t}\int \phi(\mathbf{x}, t)f(\mathbf{x})\, d_3x.$$

The first of these is computed, by displacing $\partial\phi(\mathbf{x}, t)/\partial t$ in $\int \partial\phi(\mathbf{x}, t)f'(\mathbf{x})\, d_3x$, in accordance with $X_{f,t}$, as $\int f(\mathbf{x})f'(\mathbf{x})\, d_3x$. The second of these is similarly computed by displacing $\phi(\mathbf{x}, t)$ in $\int \phi(\mathbf{x}, t)f(\mathbf{x})\, d_3x$, in accordance with $(\partial/\partial t)X_{f',t}$, as $-\int f(\mathbf{x})f'(\mathbf{x})\, d_3x$. These contributions occur with opposite sign, and introducing finally the coefficient $-i(\frac{1}{2}g)$, it results that $[R_{f,t}, (\partial/\partial t)R_{f',t}]$ is as stated above.

Now defining $\tilde{\phi}(x) = R_{f,t}$ with f taken as the delta function at the point $\mathbf{x}$, the foregoing commutation relations give the canonical relations at equal times $t = t'$:

$$[\tilde{\phi}(x), \tilde{\phi}(x')] = 0,$$
$$[\tilde{\phi}(x), (\partial/\partial t)\tilde{\phi}(x')] = -ig\delta(\mathbf{x} - \mathbf{x}'),$$
$$[(\partial/\partial t)\tilde{\phi}(x), (\partial/\partial t)\tilde{\phi}(x')] = 0.$$

These relations are not in manifestly relativistic form. The Lorentz-invariant character of the field $\tilde{\phi}$, and incidentally the action of the, Lorentz group on the present state vectors, is shown as follows. If L is any Lorentz transformation in physical space, there is a corresponding action $U(L)$ on M: $\phi(x) \to \phi(L^{-1}x)$ (in terms of the coordinates provided by the Cauchy data at any fixed time, this action is of course nonlinear). Any such transformation $U(L)$ on M gives a transformation $V(U(L))$ of the linear space of functionals on M, as follows: if T is any transformation on M, $V(T)$ carries the functional $F(\phi)$ into $F(T(\phi))$. Putting $V(U(L)) = \Gamma(L)$, the map $L \to \Gamma(L)$ is then a linear representation of the Lorentz group on the space of functionals over M.

To say that the field is Lorentz-invariant is to say that

$$\tilde{\phi}(Lx) = \Gamma(L)\tilde{\phi}(x)\Gamma(L).$$

It suffices to show that the two constituents $X_{\delta_x,t}$ and $\phi(\mathbf{x}, t)$, where δ_x denotes the delta function at x, separately enjoy the same transformation properties. In case L is an Euclidean motion or translation in time, this is clear from the construction; it is the invariance under Lorentz rotations which requires examination. It is, therefore, no essential loss of generality to take $x = Lx = 0$, and to take L as the transformation

$$x' = x\cosh\theta + t\sinh\theta,$$
$$y' = y, \qquad z' = z,$$
$$t' = x\sinh\theta + t\cosh\theta,$$

and as above, to establish infinitesimal invariance. In the case of the multiplication operator $M_{\delta_x,t}$, which is here the same as multiplication by $\phi(x)$, the invariance is clear.

Consider then the action of the given transformations on the vector field $X_{\delta,0}$. The tangent vectors corresponding to this field are represented by solutions η of Eq. (5) with the initial conditions

$$\eta(\mathbf{x}, 0) = 0, \qquad \partial\eta(\mathbf{x}, t)/\partial t]_{t=0} = \delta(\mathbf{x}).$$

Up to second order in t, η has then the form

$$\eta(\mathbf{x}, t) = t\delta(\mathbf{x}).$$

In the transformed Lorentz frame, it is clear that

$$\eta(\mathbf{x}, t)]_{t'=0} = 0,$$

while $\partial\eta/\partial t' = (\partial\eta/\partial x)(\partial x/\partial t') + (\partial\eta/\partial t)(\partial t/\partial t')$, so that for $t' = 0$, when $x = -\tanh\theta\, t$,

$$\partial\eta/\partial t']_{t'=0} = (\cosh\theta)^{-1}\,(\partial\eta/\partial t)]_{t'=0}$$
$$= (\cosh\theta)^{-1}\,\delta(x\cosh\theta)\delta(y)\delta(z) = \delta(\mathbf{x}') + O(\theta^2).$$

This shows that the vector field $X_{\delta_x,t}$ depends only on the point $(\mathbf{x}, t)$ and not on the Lorentz frame at this point.

So far, $\tilde{\phi}(x)$ has been defined as a formal operator on arbitrary functionals on M [actually, $\int \tilde{\phi}(x)f(x)d_3x$ has mathematical existence, as a well-defined operator on sufficiently smooth functionals over M, only if f is small at infinity and sufficiently smooth, but this is of no significance for the remainder of this section]. For quantum mechanical purposes, a Hilbert space of functionals over M must be set up in which the field operators are Hermitian, the action of the Lorentz group unitary, etc. This Hilbert space is naturally that with inner product

$$\langle F \mid G\rangle = \int \overline{F(\phi)}G(\phi)\, d\phi,$$

with the heuristic volume element indicated above,

$$d\phi = \prod_i dp_i\, dq_i.$$

[It should perhaps be mentioned that this is the main mathematically heuristic element in our treatment; involved is the suitable conjunction of the theory of analysis of functionals over Hilbert space, (including "functional integration") in essence already established, with the theory of propagation of nonlinear hyperbolic differential equations with a given probability distribution, of the suitably generalized type relevant for the cited analysis of functionals, in the space of Cauchy data.] The fact that the $U(L)$ are, for any Lorentz transformation L, contact transformations on M implies, as already indicated, the invariance of the volume element Ω^∞, from which the unitary character of $\Gamma(L)$ follows directly. The total field momenta (linear, angular, and the energy) are therefore well-determined Hermitian operators in this Hilbert space $\mathcal{K}$, say, of square-integrable functionals over M, as the infinitesimal generators, apart from factors of $\pm\hbar/i$, of the corresponding one-parameter groups of the $\Gamma(L)$.

The multiplication operator constituent $M_{f,t}$ of $R_{f,t}$ is Hermitian since the functional $\int \phi(\mathbf{x}, t)f(\mathbf{x})\, d_3x$ is real. To see that the vector field constituent $-i(\frac{1}{2}g)X_{f,t}$ is also Hermitian, observe that the transformations $T^s_{f,t}$ defined earlier preserve the fundamental form Ω, and so induce by direct transformation on the functionals over M, unitary transformations. As the generator of the one-parameter group $\{T^s_{f,t};\ -\infty < s < \infty\}$, $X_{f,t}$ is consequently skew-adjoint.

The quantum field variables defined above need not however act irreducibly on $\mathcal{K}$, even if taken in conjunction with the action of the Lorentz group. This is evidenced in the free-field case, in which the representation given above reduces to the 'holomorphic functional' representation previously described,[5-7] so-called because the holomorphic functionals on $\mathcal{K}$ define an invariant subspace of $\mathcal{K}$ under the actions of the quantum field variables and the Lorentz group. Upon restriction to this subspace, representations of the quantum field variables and of the Lorentz group which are rigorously unitarily equivalent to those in the rigorous formulation of the Fock representation are obtained. To obtain an appropriate Hilbert space for labeling the physical states in a unique fashion, rather than with infinite multiplicity as in the case of $\mathcal{K}$, it is assumed that there exists a Lorentz-invariant vector in $\mathcal{K}$ of minimal energy, say v. [This is of course a quite material restriction on the theory. It is plausible physically for Eq. (1) if $\rho'(\phi) \geq$ for all ϕ, and if $\rho(\phi)$ does not grow rapidly with ϕ, but is not actually needed for the determination of the collision matrix, as indicated later. It is introduced here because it is a commonly made assumption, and facilitates the correlation of the present formalism with more conventional ones.] The space $\mathcal{K}_{\text{phys}}$ spanned by successive application of the field variables to the vacuum representative v is then invariant under the Lorentz group as well as the field operators, and its vectors give the usual type of multiplicity-free representation for the physical states.

In the free-field case cited above, the representation of the field variables and the vector v may be

[5] I. E. Segal, Illinois J. Math., **6**, 500 (1962).
[6] I. E. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, Rhode Island, 1963).
[7] D. Shale, J. Math. Phys. 3, 915 (1962).

chosen in such a fashion that the subspace $\mathcal{K}_{\text{phys}}$ consists precisely of the holomorphic elements of $\mathcal{K}$. In the soluble cases in which the interaction Hamiltonian is at most quadratic in the canonical variables, a state of minimal energy may be obtained by transformation of the free physical vacuum by a type of linear transformation (viz. the symplectic transformations considered in references 6–8. Although Lorentz-invariance is generally lost in these cases, the same general formalism applies, and $\mathcal{K}_{\text{phys}}$ together with the action of the field variables and the relevant symmetry group is explicitly identifiable with those of conventional formalisms.

On the other hand, a physically more natural if mathematically seemingly more sophisticated way of treating the propagation of states, is to deal not with state vectors or wavefunctions, but with the expectation value forms, which may arise from the state vectors, along lines due originally to von Neumann and to Weyl. If ψ is such a vector, the expectation value $E(A)$ of the observable A in the state represented by ψ, $E(A) = \langle\psi\,|A|\,\psi\rangle$, has the following properties.

$$E(aA + bB) = aE(A) + bE(B), \quad E(A^2) \geq 0,$$

$$E(I) = 1 \ (I = \text{unit observable}).$$

Here A and B are arbitrary observables and a and b arbitrary real numbers. Conversely, these properties formally fully characterize expectation values in (possibly mixed) states; more specifically, any such functional is a limit of linear combinations with positive coefficients of sum unity, of functionals of the type described arising from wavefunctions. When the observables are represented by an irreducible set of operators, the states represented by a single wavefunction, i.e., the "pure" states may be identified formally from the expectation value form as those whose form can not be represented as a linear combination with positive coefficients of two other forms with the cited properties. The use of the expectation value form is independent of the possible existence of a physical vacuum, and is mathematically more broadly applicable than the use of wavefunctions, applying in cases where there is ambiguity in the representation of the canonical variables, or where a continuous spectrum might otherwise introduce vectors of infinite norm.

For present purposes the forms are used in the following way. The set of all operators on $\mathcal{K}$ which are functions of the field operators $\tilde{\phi}(x)$ form a subset $\mathfrak{A}$ (actually, subalgebra) of the collection of all linear operators on $\mathcal{K}$. The physical states are then represented in a one-to-one fashion by the forms on $\mathfrak{A}$ with the cited properties, the pure states by the forms which are "extreme points" of the set of all the expectation value forms in the sense described. The action $L \to \Gamma(L)$ of the Lorentz group on $\mathcal{K}$ induces a corresponding action on the field variables, $\tilde{\phi}(x) \to \tilde{\phi}(Lx)$, and this in turn gives the action on $\mathfrak{A}$. The action on the expectation value forms is that contragredient to the action on $\mathfrak{A}$. Thereby the dynamics and kinematics of the expectation value forms are well-determined. This identification of the physical states and their motions does not label them by the vectors in a linear space, but is nevertheless entirely adequate for conceptual empirical purposes. There is a well known mathematical procedure for setting up a Hilbert space in which any given set of expectation-value forms on a given algebra of operators may be represented as forms arising from vectors in the space, giving a theoretically more familiar and in some ways more convenient formalism, but the empirically relevant results of the theory would not be affected.

3. THE COLLISION MATRIX

The problem of the determination of the collision matrix is a fundamental one for the correlation of the present theory with the empirical situation. The theory proceeds directly from the partial differential equations of motion; no *ad hoc* interaction Lagrangian, or Hamiltonian, or incoming field, or particle description, have been postulated. (A total Hamiltonian has been determined, but no way has been given to split it into a "free" and an "interaction" part.) This is generally advantageous, since a postulate concerning the structure or even the existence of an incoming field, in particular any particle description, would in principle overdetermine the theory, introducing, in general, material theoretical inconsistencies. The use of Lagrangians or of the concept of interaction Hamiltonian involves a conceptually somewhat unempirical element. It appears that a useful notion of collision matrix can not be introduced without the use of additional structure in the direction of those cited. However, a more deliberate approach to the matter should illuminate the logical role of such assumptions.

In principle, the collision theory is determined by the dynamics of states, which is well-determined formally in the manner indicated above. If the quantum field associated with the Eq. (1) is in the state σ at the time t, σ may be represented by a functional on the space of Cauchy data at this time,

[8] D. Shale, Trans. Am. Math. Soc. **103**, 149 (1962).

say by $f(u)$ (u = Cauchy datum). The state σ' at a later time t' is then represented by $f(T_{t't}u)$, where $T_{t't}$ is the nonlinear transformation in the space of Cauchy data defined by the dynamical Eq. (1). That is to say, $T_{t't}$ gives the Cauchy datum at the time t' for the solution of the dynamical Eq. (1) with datum u at time t. If now τ is any given state at the later time t', the transition probability amplitude from σ at t to τ at t' is given by the inner product $\langle g(u) \mid f(T_{t't}u)\rangle$, where g is the functional on the space of Cauchy data representing the state τ at time t'.

The crucial theoretical problem which however arises here is that of correlating the states σ, σ', τ with the empirically observed "free" states. That is to say, the specific analytic forms of the observed particle and scattering states are required, a set of labels, or a physico-mathematical dictionary of the free states. To arrive at a suitable scheme, note that in the preceding paragraph, the states in question were specified at particular times, and the only way of propagating them to other times was by the transformations induced from the dynamical Eq. (1). Labels for these states would then propagate equivalently, and the presumably much greater simplicity of the observed "incoming" field over the actual physical "interacting" field is not made use of. The matrix elements in question in the preceding paragraph are simply those of the exponential of the total field Hamiltonian, with a time displacement of $t' - t$, whose limiting behavior as $t' \to +\infty$ and $t \to -\infty$ is evidently trivial and irrelevant.

To make use of the observed simplicity of such systems for very early or late times ("asymptotic" simplicity), a different, "free" manner of propagation for the states σ and τ must be employed. To be rather conservative, this free motion, or kinematics, is postulated as given by an equation of the same form as (1), but with a different function ρ_0:

$$\Box\phi = \rho_0(\phi). \tag{7}$$

[As indicated in reference 1, the "reference" Eq. (7) is mathematically naturally chosen as the first-order variation to the "physical" Eq. (1), in the vicinity of the (generally unique) Lorentz-invariant element of M (which is $\phi \equiv 0$ in the case of Eq. (1); and this mathematically simplest choice represents theoretical physical practice.]

To develop the basic theory of the collision matrix, it does not matter whether ρ_0 is linear or nonlinear (in fact it may even be time-dependent or have variable space coefficients for most purposes). If $\rho_0 = \rho$, the collision matrix is the unit matrix and no new structure or label has been added. If however $\rho_0(\phi) = m^2\phi$ for some constant m (i.e., has the unique relativistic linear form for a scalar field), in conformity with much conventional practice, no assumption as to the value of m need be made at this point. (It may however be helpful to think of it as the empirical mass of the particles whose physics the theoretical model is intended to describe.) To be quite specific, it is not assumed that Eq. (7) is theoretically derivable from (5), either as a classical equation or as a basis for the construction of a quantum field as indicated above. [It is in fact mathematically quite dubious whether any such effective derivation is possible. The notion of "incoming field" is also physically not altogether unambiguous, and it is particularly open to question whether in situations thought to be describable by equations such as (1)—e.g., even in quantum electrodynamics—complex states not admitting particle descriptions of a simple nature, as in the Fock representation, can be physically definitively excluded.]

The collision and wave matrices relative to the given kinematics (7) may be set up as follows. Let M_0 be the manifold of all solutions of Eq. (7). For any time t there is a map $w(t)$ of M_0 onto M, defined as that taking any element of M_0 into that of M having the same Cauchy data at time t. This map induces a corresponding map $W(t)$ of the functionals on M_0 to the functionals on M. Specifically, if f_0 is a given functional on M_0, set $f(\phi) = f_0(w(t)^{-1}\phi)$; then $W(t)$ is the map $f_0 \to f$.

This map $W(t)$ is evidently linear, and is in addition unitary, from the space $L_2(M_0)$ of square integrable functionals defined on M_0 onto the space $L_2(M)$ of square-integrable functionals defined on M. To see this, recall that a point transformation between two sets with given volume elements induces a unitary transformation between the corresponding spaces of square-integrable functionals provided it is volume-preserving. Now the transformation $w(t)$ of M_0 onto M is symplectic with respect to the fundamental forms defined above, i.e., carries by its induced action the fundamental form Ω_0 on M_0 into the fundamental form Ω on M. For in terms of the parametrizations of M_0 and M provided by the Cauchy data at time t, Ω_0 and Ω are expressible in the same forms. Since the volume elements on M_0 and M are derived from the respective forms, any symplectic transformation of M_0 onto M will be volume-preserving. $W(t)$ has addi-

tionally the important property of transforming the reference into the physical field at time t: $W(t)^{-1}\tilde{\phi}(\mathbf{x}, t)W(t) = \phi_0(\mathbf{x}, t)$, and similarly with $\tilde{\phi}(x)$ replaced by $\partial\tilde{\phi}/\partial t$.

Defining now

$$W_- = \lim_{t\to-\infty} W(t), \qquad W_+ = \lim_{t\to\infty} W(t),$$

W_- and W_+ are formally unitary operators from $L_2(M_0)$ to $L_2(M)$. They are designated the *gross* incoming or outgoing *wave operators* (the reduced wave operators, to be introduced later, are essentially the same except that the multiplicity in the description of the physical states has been eliminated; the reduced operators correspond directly to the conventional physical concepts, but are analytically less convenient than as well as directly determined by the gross operators). The operators

$$C_0 = W_+^{-1}W_-, \qquad C = W_-W_+^{-1}$$

are then formally unitary operators on $L_2(M_0)$ and $L_2(M)$, respectively, and define the gross collision operators, as represented on the respective spaces. Noting that

$$C = W_-C_0W_-^{-1} = W_+C_0W_+^{-1},$$

it is evident that the distinction between C_0 and C is merely one of the labeling of states; in practice, with ρ_0 taken as linear, C_0 would be the more useful, relating directly to the particle interpretation of the free states (see below).

Apart from their unitarity, the fundamental mathematical property of the foregoing operators is their Lorentz invariance, i.e., independence of the Lorentz frame employed in the foregoing analysis. In analytical form, this means for the collision operators

$$\Gamma(L)^{-1}C\Gamma(L) = C \quad \text{or} \quad \Gamma_0(L)^{-1}C_0\Gamma_0(L) = C_0,$$

where $\Gamma(L)$ denotes the unitary transformation on $L_2(M)$ induced by the Lorentz transformation L, while $\Gamma_0(L)$ is the same for the "free" manifold. These relations are implied by similar transformation properties of the wave operators

$$W_\pm = \Gamma(L)W_\pm\Gamma_0(L),$$

for arbitrary Lorentz transformations L. This invariance is clear for the cited equations as regards space–time translation and spatial rotations, and requires verification only for the special transformations:

$$L_\theta : x' = x\cosh\theta + t\sinh\theta,$$
$$y' = y, \qquad z' = z,$$
$$t' = x\sinh\theta + t\cosh\theta.$$

Setting for notational convenience $\Gamma(\theta) = \Gamma(L_\theta)$ and $\Gamma_0(\theta) = \Gamma_0(L_\theta)$, it remains to show that

$$\lim_{t\to-\infty} \Gamma(\theta)W(t)\Gamma_0(\theta)^{-1} = \lim_{t\to-\infty} W(t),$$

or equivalently,

$$\lim_{t\to-\infty} \Gamma_0(\theta)W(t)^{-1} = \lim_{t\to-\infty} W(t)'\Gamma(\theta).$$

(The case $t \to +\infty$ follows by symmetry.) It suffices naturally to establish infinitesimal invariance.

Now the transformations $\Gamma(\theta)$, $W(t)$, $\Gamma_0(\theta)$ are all unitary transformations in function spaces induced from volume-preserving point transformations on the underlying manifolds (i.e., M and M_0 here). To assure the asymptotic character of the foregoing transformations on the function spaces, it is therefore sufficient to show that the corresponding point transformations are asymptotic, or

$$U_0(L_\theta)w(t)^{-1} \sim w(t)^{-1}U_0(L_\theta), \qquad t \to -\infty;$$

here $U_0(L_\theta)$ and $U(L_\theta)$ are the transformations on the manifolds M_0 and M (respectively) induced by the Lorentz transformation L_θ. This asymptotic condition means that for arbitrary ϕ_0 in M_0,

$$U_0(L_\theta)w(t)^{-1}\phi - w(t)^{-1}U(L_\theta)\phi \to 0, \qquad t \to -\infty,$$

in the sense that the difference of the Cauchy data at any fixed time goes to zero, for the elements $\phi_1^{(\theta)} = U_0(L_\theta)w(t)^{-1}\phi$ and $\phi_2^{(\theta)} = w(t)^{-1}U(L_\theta)\phi$ of M_0 involved in the foregoing. Infinitesimal invariance means that the derivative of the difference with respect to θ, evaluated at $\theta = 0$, tends to zero as $t \to -\infty$. Interchanging the independent operations $\partial/\partial\theta$ and $\lim_{t\to-\infty}$, the relation to be established becomes

$$\lim_{t\to-\infty} [(\partial/\partial\theta)(U(L_\theta)w(t)^{-1}\phi_0 - w(t)^{-1}U_0(L_\theta)\phi_0)]_{\theta=0} = 0.$$

It is convenient now to denote t by a different symbol $\tilde{t}$ so as to make explicit its complete independence from the time variable on which the elements of M or M_0 under discussion depend as functions on space–time, and which may now be denoted t. Setting $\phi_0 = w(\tilde{t})^{-1}\phi$ (so that ϕ_0 is a $\tilde{t}$-dependent element of M_0), $(\partial/\partial\theta)\phi_1^{(\theta)}\,|_{\theta=0}$ is a tangent function l_1 to M_0 at ϕ_0, represented in fact by the solution $X\phi_0$ of Eq. (5), where $X = x(\partial/\partial t) + t(\partial/\partial x)$ is the generator of the one-parameter group in question. On the other hand, $(\partial/\partial\theta)\phi_2^{(\theta)}\,|_{\theta=0}$ is another tangent function l_2 to M_0, which may be expressed as (grad $w(t)^{-1}$)$(X\phi)$, and so has Cauchy data at time $t = \tilde{t}$:

$$l_2(\mathbf{x}, \tilde{t}) = (X\phi)(\mathbf{x}, \tilde{t});$$

$$(\partial/\partial t)l_2(\mathbf{x}, \tilde{t}) = (\partial/\partial t)(X\phi)(\mathbf{x}, \tilde{t}).$$

The two tangent functions l_1 and l_2 agree for $t = \tilde{t}$, for the identity of ϕ and its first time derivative at time $t = \tilde{t}$ with the similar data for ϕ_0 imply that $X\phi$ and $X\phi_0$ agree when $t = \tilde{t}$. The time derivatives of $X\phi$ and $X\phi_0$ involve however the second-order time derivatives of ϕ and ϕ_0. These are given by the partial differential equations defining the manifolds M and M_0. In this way it may be computed that the time derivatives of l_1 and l_2, evaluated at $t = \tilde{t}$, differ by $x(\rho_0(\phi) - \rho(\phi))$ [here $\mathbf{x} = (x, y, z)$]. The assertion is then that the tangent function to M_0 determined by its vanishing at time $\tilde{t}$ and by its first derivative being given by the last expression, has its Cauchy data, for arbitrary fixed times, convergent to zero as $\tilde{t} \to -\infty$.

To conclude the argument, it is assumed that $\int x(\rho_0(\phi) - \rho(\phi))^2 \, d_3x \to 0$ for all ϕ in M, a condition that ρ_0 deviate not too strongly from ρ. For example, in the realistic case $\rho_0(\phi) = m^2\phi$ and $\rho(\phi) = m^2\phi + g^2\phi^3$, the condition becomes

$$\int x^2(\phi(\mathbf{x}, t))^6 \, d_3x \to 0 \quad \text{as} \quad t \to \pm\infty, \ .$$

which is not yet known to be the case, but for which there are material plausibility indications. [*Note added in proof.* This has now been proved for the case $m = 0$ by Walter Strauss, Compt. Rend. **256**, 5045 (1963).]

It may be noted that the asymptotic invariance would appear to be mathematically more generally valid than the strict existence of the collision operator, in the sense that the foregoing difference may well converge to zero, although the individual terms fail to converge. On the other hand, the asymptotic invariance is not a mere formal corollary to the relativistic character of the equations and the fundamental form, as may be seen from consideration of the case in which $\rho(\phi) = m^2\phi$ and $\rho_0(\phi) = m_0^2\phi$. There is no actual collision operator if $m \neq m_0$, and from the computations given above and properties of Klein–Gordon wavefunctions, it is readily verified that the norm of $l_1 - l_2$ does not go to zero as $\tilde{t} \to -\infty$.

The *reduced* wave and scattering operators, formally equivalent to those usually treated, are obtained by suppressing the multiplicity in the representation of states by (ray) vectors through the contraction of the gross operators to subspaces similar to those treated earlier for the case when a vacuum exists. More specifically, assume there exists a vacuum state representative v for the physical quantum field. This is an element of $L_2(M)$ which is left fixed by the induced action $L \to \Gamma(L)$ of the Lorentz group. Now it is no essential loss of generality to assume that v is an eigenstate of the collision operator C. For let $\mathfrak{N}$ be the closed linear manifold spanned by the $f(C)v$, where $f(C)$ is an arbitrary function of C. Then C leaves $\mathfrak{N}$ invariant, and its restriction to $\mathfrak{N}$, a unitary operator with a simple (or nondegenerate) spectrum. Now the representation operators $\Gamma(L)$ for the Lorentz group leave every element of $\mathfrak{N}$ fixed since they commute with C and leave v fixed. Any eigenstate of the restriction of C to $\mathfrak{N}$ is then a physical vacuum representative. The element $W_-^{-1}v = v_0$ (or alternatively, $W_+^{-1}v$, which differs from v_0 by a constant phase factor) is then a vacuum representative for the reference system determined by Eq. (7).

The *reduced reference field state space* is then definable as the smallest closed linear manifold containing v_0 and invariant under the field operators $\tilde{\phi}_0(x)$ for all x, as well as the collision operator C_0. (The last requirement may commonly be superfluous, as indicated by quasiphysical intuition, but as there is no explicit representation for the collision operator in terms of the field variables, there is no direct mathematical indication for this, and in fact, in more complex physical systems with degenerate vacuums may well be quite material.) The *reduced physical field state space* is then definable as $W_-\mathcal{K}$ or equivalently $W_+\mathcal{K}$. Conversely, the physical field space could be defined first and the reference field space obtained by application of the inverse of one of the wave operators. The spaces obtained are invariant under the respective (a) field operators $\tilde{\phi}(x)$ [or $\tilde{\phi}_0(x)$], (b) collision operator C (or C_0), and (c) representation operators $\Gamma(L)$ [or $\Gamma_0(L)$]. The restrictions of these operators to the reduced space therefore satisfy all algebraic relations between them satisfied prior to restriction, and define the reduced-field, collision, and momenta operators (the latter as generators of the action of one-parameter groups of finite Lorentz transformations).

Example. To show concretely how the collision matrix elements may be computed through the use of the foregoing formalism, consider a physical field determined by the equation

$$\Box\phi = m^2\phi + g'\phi^3 \qquad (g' > 0),$$

and reference field determined by the equation

$$\Box\phi = m^2\phi.$$

Consider the key problem of determining the C-matrix element between incoming and outgoing states each of which is constituted of finite sets of "particles". The effective applicability of a particle interpretation derives from the linearity of the reference field equation, which has previously played no role in the formalism.

At a sufficiently early time $t \sim -\infty$, the physical system is sufficiently closely represented by the reference particle system. Now particle states of the reference field may be represented in a well-known way in the Fock representation for the Bose–Einstein field in question. For present purposes, the states in question must be represented rather by functionals on the space of Cauchy data of the relevant linear relativistic equation. The existence of this alternative representation derives directly from the explicit representation for the variables obtained by specialization of the above Sec. 2 to the case of the scalar meson equation. Since the commutation relations obtained are identical with those of the Fock representation, there is a mutual correspondence between the observables of the two field representations. This implies a correspondence between the states of the two representations, and when the multiplicity of the functional representation is suppressed by working within the subspace $\mathcal{K}$ indicated earlier, there exists a unitary correspondence between the Fock representation space and the subspace $\mathcal{K}$ under which the field operators $\tilde{\phi}(x)$ and the Lorentz group actions $\Gamma(L)$ correspond. This correspondence can be carried through in closed analytic form (cf. Ref. 5, last section, for the basic theory). The Fock states correspond to the *holomorphic* elements of $L_2(M_0)$, when M_0 is regarded as a complex manifold with multiplication by real scalars as given but with multiplication by i defined as the Hilbert transform relative to time; an n-particle state corresponds simply to a polynomial of degree n, multiplied by the fundamental exponential of the negative of the Lorentz-invariant inner product representative of the vacuum.

The precise form of the functionals is not needed here, but for specificity, it may be mentioned that in terms of the Cauchy data at time t for the Klein–Gordon equation being considered here, the relevant functionals $f(\phi)$ have the form

$$f(\phi) = p(\langle\phi, \psi_1\rangle, \langle\phi, \psi_2\rangle, \cdots, \langle\phi, \psi_{n'}\rangle) \times \exp\left[-\tfrac{1}{2}\,||\phi||^2\right],$$

where $p(s_1, s_2, \cdots, s_{n'})$ is an ordinary complex polynomial in the complex variables $s_1, s_2, \cdots, s_{n'}$; the $\psi_1, \psi_2, \cdots, \psi_n$ are arbitrary Cauchy data; the Hermitian inner product $\langle\phi, \psi\rangle$ is given by the equation

$$\langle\phi, \psi\rangle = \int [(m^2 - \Delta)^{(1/2)}\phi(\mathbf{x}, t)]\cdot\psi(\mathbf{x}, t)\, d_3\mathbf{x} + \int [(m^2 - \Delta)^{-(1/2)}(\partial\phi(\mathbf{x}, t)/\partial t]\cdot(\partial\psi(\mathbf{x}, t)/\partial t)\, d_3\mathbf{x} + i\int \left(\phi(\mathbf{x}, t)\,\frac{\partial\psi}{\partial t}(\mathbf{x}, t) - \psi(\mathbf{x}, t)\,\frac{\partial\phi}{\partial t}(\mathbf{x}, t)\right) d_3\mathbf{x};$$

and $||\phi||^2 = \langle\phi, \phi\rangle$. (This results from direct substitution in the result applicable to arbitrary wave equations, making use of familiar properties of the scalar meson wavefunctions, and notably their expression in terms of Cauchy data.)

At a later time t', the physical field is the state represented by the transformation of the given functional $f(\phi)$ via the action induced from the fundamental nonlinear equation. In other terms, the state at time t' is represented by the functional g of the Cauchy data at that time given by the equation

$$g(\phi) = f(T_{t',t}^{-1}\phi),$$

where $T_{t',t}$ is the nonlinear transformation on the space of Cauchy data which carries the data at time t for a solution of Eq. (8) into its data at time t'. At a sufficiently late time $t' \sim \infty$, the physical system is sufficiently closely represented by the reference system, which is the state represented by the functional $g(\phi)$. The collision matrix element $\langle CF, F'\rangle$ between the states of the reference field represented by the functionals F and F' on M_0 is then the inner product $\langle g, f'\rangle$ between the functional g obtained above on the Cauchy data at time t', from the functional f representing F as a functional of the Cauchy data at time t, and the functional f', representing F' as a functional of the Cauchy data at time t'. This inner product is taken in the space of all square-integrable functionals over the space of Cauchy data.

The resulting collision matrix element is then completely well-defined, apart from the integration involved in forming the last inner product. Strictly speaking, the limit as $t \to -\infty$ and $t' \to \infty$, which should exist if the physical model is valid, of the expression given above, should be taken. As a computational procedure, however, the use of finite times t and t' approximating $-\infty$ and $+\infty$ is unexceptionable. All matters of convergence or divergence are consolidated in the existence of the integrals in function space which arise, and which,

for reasons indicated earlier, are appropriately treated elsewhere. It may be noted however that when an integral in function space exists, it can be approximated arbitrarily closely by an integral over an appropriate subspace of sufficiently high finite dimension. The C matrix can hence be computed, when it exists, within an arbitrarily small error, by the solution of Eq. (8) for arbitrary Cauchy data, which is known to be possible by methods affording estimates and bounds for the errors involved in numerical quadrature [cf. Ref. 9], together with the evaluation of integrals whose integrands arise from the integration of Eq. (8) over Euclidean spaces of sufficiently high dimensions.

It is noteworthy also that no hypothesis as to the existence of a physical vacuum is involved in the foregoing method for computing the C matrix. All that is required is a correspondence, or type of dictionary describing any given quasiempirical state of the incoming field in mathematical terms as a state of the reference field, without which the concept of a collision matrix would not appear to have definite meaning.

4. DISCUSSION

A. Relation to Ordinary Nonrelativistic Theory

To examine this relation, assume that the foregoing theory extends to the case of Fermi–Dirac fields with suitable modifications, and is thereby applicable to the Dirac equation. This equation, with a given external time-independent potential, determines a manifold of functions to which the foregoing analysis applies with the Lorentz group replaced by the group of displacements in time, augmented in most cases by the group of spatial rotations. The classical Dirac equation appears as the limiting case of the corresponding quantum field theory as the scale of the anticommutators tends to zero. The classical Dirac equation in turn yields as a limiting case (in a familiar way) the Schrödinger equation as $c \to \infty$.

The conventional Schrödinger equation thus appears only as an approximation within the present formalism. On the other hand, this formalism could be adapted somewhat to deal with nonrelativistic systems of finitely many degrees of freedom. This is formal application of quite a different nature from that made above, and, in view of the approximate character of any nonrelativistic model, can have no absolute validity. In general, the simplicity of the Schrödinger equation is such as to give it a great advantage over the present formalism for an approximate treatment, as is familiar also in conventional field theory. It is demonstrable however that the adaptation of the present formalism to a nonrelativistic system with a finite number of degrees of freedom leads to identical results with the usual theory when the Hamiltonian is at most quadratic in the canonical variables. When the Hamiltonian is more complicated, as for example in the case of the hydrogen atom, the present phase-space formalism is analytically considerably less simple than the usual treatment in physical space. It might however be of theoretical interest to explore the consequences of the phase-space formalism in such cases.

B. Manifestly Covariant Formalism

It was physically and mathematically convenient in the foregoing to work in a specific Lorentz frame. This formalism is probably more fundamental, from both points of view, than the earlier formalism[1] which was more manifestly relativistic. However, the Lorentz-invariant character of the theory is of great importance, and it is of interest to adapt the present formalism to a more manifestly relativistic one. This involves replacing the vector fields $X_{f,t}$ which generate simple displacements in the Cauchy data at time t, by vector fields X_F (where F is a smooth function on space–time vanishing at infinity) which generate one-parameter groups which can not be given in closed form. When F has the form $F(x') = f(\mathbf{x}')\delta'(t' - t)$, X_F agrees with $X_{f,t}$. In general, it assigns at any point ϕ of M the tangent function l to M at ϕ given by the equation

$$l(x) = \int D_\phi(x, y)F(y)\, d_4y,$$

where $D_\phi(x, y)$ is the Riemann, or commutator, function for the linear partial differential equation defining the tangent-function space at ϕ. The Cauchy data for $D_\phi(x, y)$ at equal times are such as to show directly that the $X_{f,t}$, which themselves do not form a linear set, are subsumable in the indicated way under the X_F, which do form a linear, relativistic set, transforming appropriately under the action of the Lorentz group. The commutators of two X_F's is computable in terms of the variational derivatives $(\partial/\partial\phi)D_\phi(x, y)$, from which $[\tilde{\phi}(x), \tilde{\phi}(x')]$ may be derived for arbitrary x and x' in a relatively explicit form. From the commutators it can also be explicitly checked that the following relations, which imply the closure of

[9] I. E. Segal, Ann. Math. 78, 339 (1963).

Ω, are valid (here $X = X_F$, $Y = X_G$, $Z = X_H$, for arbitrary F, G, and H):

$$\Omega(X, [Y, Z]) + \Omega(Z, [X, Y]) + \Omega(Y, [Z, X]) = 0,$$

$$X\Omega(Y, Z) + Z\Omega(X, Y) + Y\Omega(Z, X) = 0$$

(the covariant differential $3d\Omega$ is the sum of the foregoing terms). The closure is obvious in the less manifestly relativistic formulation used above, corresponding to the vanishing of the individual terms in the foregoing expressions when F, G, and G vanish except at a particular time (i.e., involve delta functions at the same time).

C. The Equation Satisfied by the Field

There is no difficulty in computing $\Box\tilde{\phi}(x)$ by the same method as that by which $\partial\tilde{\phi}(x)/\partial t$ was computed. The result is however not a simple function of the field variables $\tilde{\phi}(x)$, so far as can be seen. It can be stated as follows:

$$\Box(\tilde{\phi}(x) - \phi(x)) = \rho'(\phi(x))(\tilde{\phi}(x) - \phi(x));$$

here $\phi(x)$ denotes the operation of multiplication of a functional $f(\phi)$ of the classical field ϕ by the functional $\phi(x)$. This equation can be interpreted heuristically as implying that in the vicinity of a state which is closely approximated by a classical state, the deviation of the quantum from the classical field satisfies the first-order variational equation in the vicinity of the classical field. This is in the same direction as the assertion that the quantized field satisfies the classical equation, from an intuitive point of view, but differs from it in being mathematically clear-cut. (For its conceptual physical implications, cf. Ref. 1.)

D. Possible Role of Renormalization

The procedure given above determines the matrix element between a given state at an early time t and a given state at a late time t'. There can be no *a priori* assurance that the limit of this matrix element as $t \to -\infty$ and $t' \to +\infty$ will exist. Whether this is the case may depend on the equations in question, the states being considered, and the scale of the quantum field. It is also possible for the limit to exist somewhat more generally if the mass m in the reference equation, the constant g representing the scales of the reference and physical quantum fields, and the classical coupling constant g', in, say, the case of the equations discussed at the end of the preceding section, are simultaneously adjusted appropriately with the passage to the limit. While this would suggest that the basic partial differential equations have only approximate validity, it is, nevertheless, an entirely proper applied mathematical procedure to exploit an approximate equation in this way.

No less relevant is the approximation involved in the representation of the empirical states by mathematically well-defined states of the reference field. The rigorous theory gives the matrix elements only between dense classes of states of the reference field. There is no assurance that the particle states which are of primary interest are included in this dense class, and probable in fact that they are not, for theories of empirical relevance. It is then necessary to approximate to the particle states by the mathematically well-behaved states, which is readily done, but without any assurance that the corresponding matrix elements being discussed converge, as the mathematical states converge to the (mathematically well-defined, but not necessarily rigorously propagatable in the physical quantum field) particle states. In this circumstance, an applied mathematical procedure similar to that described above, i.e., some variety of renormalization, may be effective in computing physical results. The extent to which renormalization of this type may be required is then a measure of the extent to which two distinct features of the theoretical model for the empirical situation are valid: (i) the basic nonlinear partial differential equation for the physical field; (ii) the representation of the incoming physical particle states by the usual mathematical particle states in the Fock representation for the reference quantum field employed.

E. Role of Rigor

The integrals with respect to a Euclidean volume element in the space of Cauchy data employed above must properly be replaced by integration with respect to appropriate quasi-invariant weak probability distributions in the space of Cauchy data, which have definite mathematical existence (cf. Ref. 3). A number of analytical difficulties arise, of which the central one is the solution of a given nonlinear hyperbolic equation with a given quasi-invariant weak probability distribution in the space of Cauchy data, i.e., a stochastic Cauchy problem for such equations. The major problem here is that the nonlinear transformations defining the propagation by the given equation from one time to another do not in general represent bimeasurable transformations on the space of Cauchy data, with respect to a given distribution of the type indicated. (Compare, for the mathematical

background, Ref. 10 and 11, which give the best published results to date on the relevant transformation problem.) The physical interpretation of this is the complication referred to above, that not every normalizable state of an incoming reference field can be expected to be rigorously capable of propagation by the physical field equations, but only a dense subset. The theory provides the transition amplitude matrix between such "physical" states at arbitrary finite times.

All this is quite independent of the possible existence of a physical vacuum, for which there is no mathematical assurance. On the other hand, the transition matrix elements between elements of a dense class of states at different times appears to be all that is physically absolutely essential or indicated, from a conservative point of view. The particle states of primary empirical interest are approximable by members of this dense class. There is no general mathematical argument to show that the particle states in any sense fully span the state space for the physical field. It would be remarkable if in any case this were demonstrable, for there is no empirical assurance that complex states without any exact physical-particle interpretation can not occur as the result of any given class of (relativistic) interactions of elementary particles, and no determination in closed form of the states of the physical quantum field in terms of those of the reference field can be expected.

[10] L. Gross, Trans. Am. Math. Soc. 94, 404 (1960).
[11] L. Gross, Mem. Amer. Math. Soc. 46, (1963), p. 1.

Irreducible Tensor Expansion of Solid Spherical Harmonic-Type Operators in Quantum Mechanics*

YING-NAN CHIU
Laboratory of Molecular Structure and Spectra, Department of Physics, University of Chicago
(Received 14 August 1963)

A method is proposed to derive the general one- and two-center expansions of quantum mechanical operators that are in the form of regular or irregular solid spherical harmonics of any integral degree. The expansions are expressed as couplings of irreducible spherical tensors. The physical nature of such coupling and the parallelism to vector addition are illustrated. Possible uses of such expansions in the evaluation of molecular integral and their reduction to simple, known expansions obtained from other methods are briefly discussed.

QUANTUM mechanical operators may occur in the form of regular solid spherical harmonics, $r_{ij}^n P_n^k(\cos\Theta_{ij})$, or irregular solid spherical harmonics, $r_{ij}^{-n-1}P_n^k(\cos\Theta_{ij})$, or their variations.[1] The subscripts i and j may both refer to nuclei or both to electrons or one may refer to a nucleus the other to an electron. An example of the regular solid harmonic operator is the 2^n electric multipole moment of an atom or a molecule. An example of the irregular solid harmonic operator is to be found in the magnetic interaction between the magnetic moments of electrons and nuclei.[2]

For a two-electron operator, in general, it is desirable to expand it into the coordinates of each electron before integration. In particular, for integration of a multiple-center molecular integral it is advantageous to express the operator in a bipolar (two-center) expansion. With $n = k = 0$, we have the simplest kind of irregular solid harmonic operator, namely the Coulomb potential $1/r_{ij}$, the expansion of which, aside from the well-known Legendre expansion in surface harmonics of the first kind,[3] the Neumann expansion in surface harmonics of the first and second kind,[4] has also been investigated by Carlson and Rushbrooke,[5] Buehler and Hirschfelder,[6] Rose,[7] and Fontana.[8] For irregular solid harmonics, with $n = 2$, we have the magnetic-dipole interaction operator,[9] the one-center expansion of which has been given by Pitzer *et al.*[1] The present paper is to propose a general method for deriving the expansion for any n and k, which also permits easy extension to two-center expansions. We shall be more concerned with the formal aspects of the expansion than its eventual application.

We define the normalized nth-order surface spherical harmonic which transforms as an irreducible spherical tensor of the nth rank as follows[10]:

$$Y_n^k(\theta,\phi) = \left[\frac{2n+1}{4\pi}\frac{(n-k)!}{(n+k)!}\right]^{\frac{1}{2}} e^{ik\phi}(-)^k \times \frac{(\sin\theta)^k}{2^n n!}\left(\frac{d}{d\cos\theta}\right)^{n+k}(\cos^2\theta - 1)^n = \left[\frac{2n+1}{4\pi}\frac{(n-k)!}{(n+k)!}\right]^{\frac{1}{2}} e^{ik\phi}P_n^k(\cos\theta).$$

The expansion of the two-particle irregular solid spherical harmonic (of degree $-n-1$) operator

$$\Upsilon_n^k(\mathbf{r}_{12}) \equiv r_{12}^{-n-1}Y_n^k(\chi,\Phi)$$

may be achieved through the following steps:

Step 1. Starting at point e_1 (see Fig. 1) as the local origin of the coordinates, we can express the surface harmonic $Y_n^k(\chi,\Phi)$ of the vector $\mathbf{r}_{12}$ with respect to the local z' (as the new) polar axis in terms of the surface harmonics of the same vector $\mathbf{r}_{12}$ with respect to $\mathbf{r}_1$ as (the old) polar axis, through a rotational transformation.[11]

$$\Upsilon_n^k(\mathbf{r}_{12}) = \frac{Y_n^k(\chi,\Phi)}{r_{12}^{n+1}} = \sum_{m'=-n}^{n}\frac{D_{m'k}^n(0,-\theta_1,-\phi_1)Y_n^{m'}(\psi,\phi)}{r_{12}^{n+1}}, \quad (1)$$

* Work supported by a grant from the National Science Foundation.

[1] R. Pitzer, C. W. Kern, and W. N. Lipscomb, J. Chem. Phys. **37**, 267 (1962).
[2] N. F. Ramsey, *Molecular Beams* (Oxford University Press, London, 1956), p. 69.
[3] H. Eyring, J. Walter, and G. E. Kimball, *Quantum Chemistry* (John Wiley & Sons, Inc., New York, 1944), p. 371.
[4] K. Ruedenberg, J. Chem. Phys. **19**, 1459 (1951).
[5] B. C. Carlson and G. S. Rushbrooke, Proc. Camb. Phil. Soc., **46**, 626 (1950).
[6] R. J. Buehler and J. O. Hirschfelder, Phys. Rev. **83**, 628 (1951).
[7] M. E. Rose, J. Math. and Phys. **37**, 215 (1958).
[8] P. R. Fontana, J. Math. Phys. **2**, 825 (1961).
[9] Y. N. Chiu, J. Chem. Phys. **39**, 2736, 2749 (1963).
[10] Note the factor $(-)^k$ is part of our definition of $P_n^k(\cos\theta)$. This definition is the same as that of Carlson *et al.* (Ref. 5), Buehler *et al.* (Ref. 6), Rose (Refs. 7 and 11), Fontana (Ref. 8) and Hobson (Ref. 13) but different from that of Eyring *et al.* (Ref. 3) and Pitzer *et al.* (Ref. 1).
[11] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957), pp. 60, 61.

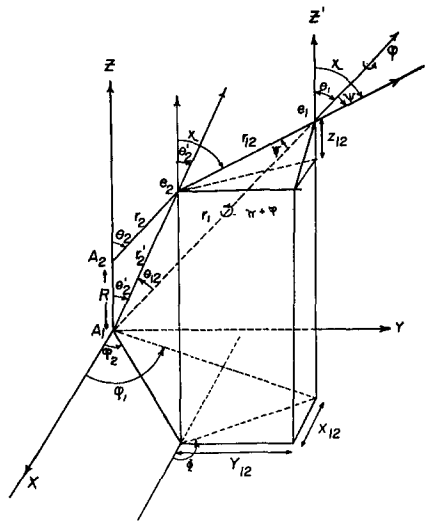


FIG. 1. Coordinate systems.

where $D^n_{m',k}$ is the rotation matrix. The angles, 0, $-\theta_1$, and $-\phi_1$ are the three Euler angles[12] that rotate the vector $\mathbf{r}_1$ to the position of the local polar (z') axis and correspond to the inverse of the rotations (ϕ_1, θ_1, γ) that carry the local polar (z') axis to the position of the vector $\mathbf{r}_1$. The third Euler angle of rotation γ is arbitrary and is set to zero here.

Step 2. We note that $Y_n^{m'}(\psi, \phi) = (-)^n Y_n^{m'}(\pi - \psi, \pi + \phi)$. After expressing $Y_n^{m'}$ in terms of $P_n^{m'}$, we expand the solid spherical harmonics at center e_1 in terms of that at center A_1 by use of a formula which is the reverse of that of Hobson[13] and Buehler *et al.*[6] for $r_2' > r_1$,

$$\frac{P_n^{m'}(\cos(\pi - \psi))}{r_{12}^{n+1}} = \sum_{l=n}^{\infty} (-)^{l-n} \times \frac{(l-m')!}{(l-n)!\,(n-m')!} \frac{(-r_1)^{l-n}}{(r_2')^{l+1}} P_l^{m'}(\cos\theta_{12}). \quad (2)$$

Then in turn we express

$$P_l^{m'}(\cos\theta_{12}) e^{im'(\pi+\phi)} \text{ back as } Y_l^{m'}(\theta_{12}, \pi + \phi),$$

and further note that the latter is the surface harmonic of the vector $\mathbf{r}_2'$ with respect to $\mathbf{r}_1$ as the (new) polar axis, and may be expressed[11] as before in terms of the surface harmonic of the same vector $\mathbf{r}_2'$ with respect to z as the (old) polar axis through a rotational transformation. We then get from (1)

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{m=-l}^{l} \sum_{m'=-n}^{n} (-)^n \frac{(l-m')!}{(l-n)!\,(n-m')!} \times \left[\frac{2n+1}{2l+1} \frac{(n-m')!\,(l+m')!}{(n+m)!\,(l-m)!}\right]^{\frac{1}{2}} \frac{r_1^{l-n}}{(r_2')^{l+1}} \times D^n_{m',k}(0, -\theta_1, -\phi_1) D^l_{mm'}(\phi_1, \theta_1, 0) Y_l^m(\theta_2'\phi_2'). \quad (3)$$

Step 3. We note[14] that $D^n_{m',k}(0, -\theta_1, -\phi_1) = (-)^{k-m'} D^n_{-k,-m'}(\phi_1, \theta_1, 0)$. From the standard Clebsch–Gordan series[15] for the coupling of the rotation matrix D's and, the explicit expression of the Wigner (Clebsch–Gordan) vector coupling coefficient[16] C's (see below), we get

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{m=-l}^{l} \sum_{j=|l-n|}^{l+n} \sum_{m'=-n}^{n} (-)^k \times \left[\frac{(2n+1)(2l)!}{(2l-2n+1)!\,2n!}\right]^{\frac{1}{2}} \times C(l, n, l-n; m', -m', 0) \times C(l, n, j; m, -k, m-k) \times C(l, n, j; m', -m', 0) \frac{r_1^{l-n}}{(r_2')^{l+1}} \times D^j_{m-k,0}(\phi_1, \theta_1, 0) Y_l^m(\theta_2', \phi_2'). \quad (4)$$

Step 4. From the orthogonality of the Wigner coefficients, we sum over m' to get $\delta_{j,l-n}$. From this Kronecker's delta, we eliminate the sum over j. Using its symmetry properties, we rearrange the variables of the remaining Wigner coefficient, and noting that[11]

$$D^{l-n}_{m-k,0}(\phi_1, \theta_1, 0) = (-)^{m-k}\left(\frac{4\pi}{2l-2n+1}\right)^{\frac{1}{2}} Y_{l-n}^{k-m}(\theta_1, \phi_1),$$

we get for $r_2' > r_1$, the one-center expansion

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{m=-l}^{l} (-)^l \left[\frac{4\pi(2l)!}{(2n)!\,(2l-2n+1)!}\right]^{\frac{1}{2}} \times C(l-n, l, n; k-m, m, k) \times \frac{r_1^{l-n}}{(r_2')^{l+1}} Y_{l-n}^{k-m}(\theta_1, \phi_1) Y_l^m(\theta_2', \phi_2'). \quad (5.1)$$

For $r_1 > r_2'$, starting[17] the expansions from point

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-)^l \times \left[\frac{4\pi(2l+2n)!}{(2l+1)!\,(2n)!}\right]^{\frac{1}{2}} C(l+n, l, n; k-m, m, k) \times \frac{(r_2')^l}{r_1^{n+l+1}} Y_{l+n}^{k-m}(\theta_1, \phi_1) Y_l^m(\theta_2', \phi_2'). \quad (6.1a)$$

[12] Ref. 11, pp. 50 and 51.
[13] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea Publishing Company, New York, 1955), pp. 140, 141.
[14] Ref. 11, p. 54.
[15] Ref. 11, p. 58.
[16] Ref. 11, pp. 38, 39.
[17] Starting from the point e_1, we get the essentially equivalent expansion.

e_2, we get a similar formula as above with r_1 and r_2' interchanged and with an extra parity factor of $(-)^n$ due to inverting the argument of the solid harmonic of order n, namely the one-center expansion,[18]

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{m=-l}^{l} (-)^{l+n} \left[\frac{4\pi(2l)!}{(2n)!\,(2l-2n+1)!}\right]^{\frac{1}{2}} \times C(l-n, l, n; k-m, m, k) \times \frac{(r_2')^{l-n}}{(r_1)^{l+1}} Y_{l-n}^{k-m}(\theta_2, \phi_2) Y_l^m(\theta_1, \phi_1). \quad (6.1)$$

Using the explicit expressions of the Wigner coefficients (see below), the definition of Y's, and the property of the associated Legendre polynomials,

$$P_l^{-m}(\cos\theta) = (-)^m \frac{(l-m)!}{(l+m)!} P_l^m(\cos\theta),$$

we have obtained from (5.1) for $n = k = 0$ the expansion of $1/r_{12}$ identical to that obtained by Eyring *et al.*[3] through the Fourier coefficient method; for $n = 2$, we have obtained the expansions of Pitzer *et al.*[1] In particular, starting from the Eyring *et al.* expansion for $1/r_{12}$ in the real form, letting

$$\frac{P_2^2(\cos\chi)\cos 2\Phi}{r_{12}^3} = \frac{3(x_{12}^2 - y_{12}^2)}{r_{12}^5} = \left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right) 1/r_{12} = 2\left(\frac{\partial^2}{\partial \xi_1^2} + \frac{\partial^2}{\partial \eta_1^2}\right) 1/r_{12}, \quad (7)$$

where $\xi_1 = x_1 + iy_1$, $\eta_1 = x_1 - iy_1$, and using the method of Hobson,[19] we have obtained for this operator five expressions which may be summarized into the form of Pitzer *et al.* and (5.1). In principle by recognizing that[19]

$$(i)^{\frac{1}{2}\mp\frac{1}{2}} \frac{\partial^{n-k}}{(\partial z_1)^{n-k}} \left\{\left(\frac{\partial}{\partial \xi_1}\right)^k \pm \left(\frac{\partial}{\partial \eta_1}\right)^k\right\} 1/r_{12} = \frac{(-)^{n-k}(n-k)!}{2^{k-1} r_{12}^{n+1}} P_n^k(\cos\chi) \begin{matrix}\cos k\Phi \\ \sin k\Phi,\end{matrix} \quad (8)$$

one could also derive our formula (5.1) for any n and k by Hobson's method. But this method, because of its having to distinguish cases for $k \leq m$, $k > m$, etc., will not yield a compact formula and is more liable to algebraic error.

Next we come to the two-electron (or two-particle) regular solid spherical harmonic operators. Starting from point e_1, we make similar expansions. At Step 2, in place of Eq. (2), we use another reversed formula of Hobson[13,6,1] as follows:

$$r_{12}^n P_n^{m'}(\cos(\pi - \psi)) = \sum_{l=m'}^{n} (-)^{n-l} \times \frac{(n+m')!}{(l+m')!\,(n-l)!} r_1^{n-l} (r_2')^l P_l^{m'}(\cos\theta_{12}). \quad (9)$$

At Step 3 we use the explicit expression for the Wigner coefficient $C(n, l, n-l; -m', m', 0)$. The resulting one-center expansion is

$$\mathcal{Y}_n^k(\mathbf{r}_{12}) = r_{12}^n Y_n^k(\chi, \Phi) = \sum_{l=m'}^{n} \sum_{m=-l}^{l} (-)^l \left[\frac{4\pi(2n+1)!}{(2l+1)!\,(2n-2l+1)!}\right]^{\frac{1}{2}} \times C(l, n-l, n; m, k-m, k) \times (r_2')^l Y_l^m(\theta_2', \phi_2') r_1^{n-l} Y_{n-l}^{k-m}(\theta_1, \phi_1), \quad (10.1)$$

where m' is set to zero. This is identical to the expansion (33) of Rose[7] derived from generalized electric multipole interaction of two-charge clusters. Although Rose started from two-center considerations, his expansion is essentially of one-center character because of his definition of $\mathbf{r}_{ab} = \mathbf{a} - \mathbf{b}$ as if both $\mathbf{a}$ and $\mathbf{b}$ had the same origin, instead of $\mathbf{a} - \mathbf{b} - \mathbf{R}$ with vectors $\mathbf{a}$ and $\mathbf{b}$ having origins displaced by a vector $\mathbf{R}$.

In the above derivations, we have taken care to make use of more familiar methods of analysis and to show the geometric relation so as to facilitate later extension to the two-center expansion. One speculative but simple method of derivation is to recognize that since the solid harmonics transform as irreducible tensors, they must couple according to similar rules,[20] except for a constant factor K and a summation sign over l which is related to the initial multipole-type expansion in $1/r_{12}$. We therefore write for $r_2' > r_1$

[18] For $\mathbf{r}_2' = \mathbf{r}_1$, i.e., $r_{12} = 0$, according to Pitzer *et al.* (Ref. 1), there are delta-function-type terms for $n - m' \geq 2$ in our Eq. (1). If the eventual integration of the molecular integral is to be performed with respect to center A_1, we may choose $\mathbf{r}_1$ as the polar axis and expand the product of electronic wavefunctions $f(\mathbf{r}_2')$ at center A_1, as in Pitzer *et al.* The delta-function-type terms which represent the operator at this contact point $\mathbf{r}_2' = \mathbf{r}_1'$ for integration purpose are:

$$\Upsilon_n^k(\mathbf{r}_{12}) = \left[\lim_{\epsilon \to 0} \int_{\epsilon}^{r_0} \sum_{l=0}^{\infty} \sum_{m'=0}^{n-2} \sum_{i=0}^{\frac{1}{2}(n-l-2)} r_{12}^{l+2i-n+1}\, dr_{12} \times \int_{\mu-}^{\mu+} \int_0^{2\pi} P_l^{m'}(\cos\psi) Y_n^{m'}(\psi, \phi) \times (a_{ilm'} \cos m'(\pi + \phi) + b_{ilm'} \sin m'(\pi + \phi)) \times D_{m'k}^n(0, -\theta_1, -\phi_1) d(\cos\psi) d\phi \right]_{f(\mathbf{r}_1) \to \delta(\mathbf{r}_1 - \mathbf{r}_1') = \delta(\mathbf{r}_{12})},$$

where we have followed their notations for the expansion and for the integration limits. As the a's and b's are functions of $f(\mathbf{r}_1)$, if, after the integration and the limiting process we replace every $f(\mathbf{r}_1)$ by $\delta(\mathbf{r}_{12})$, we get for the operator $\Upsilon_n^k(\mathbf{r}_{12})$ a function of $\delta(\mathbf{r}_{12})$ and its derivatives.

[19] Ref. 13, pp. 134–139.

[20] Ref. 11, p. 79.

$$\mathfrak{T}_n^k(\mathbf{r}_{12}) = \frac{Y_n^k(\chi, \Phi)}{r_{12}^{n+1}}$$
$$= K \sum_l \sum_m C(l-n, l, n; k-m, m, k) \times \mathfrak{Y}_{l-n}^{k-m}(\mathbf{r}_1)\mathfrak{T}_l^m(\mathbf{r}_2'), \quad (11)$$

and note that, if $\mathbf{r}_1$ is taken to be the polar axis, the above formula must reduce to a formula similar to that of Eq. (2) except for a parity factor $(-)^n$ due to inverting the direction of $\mathbf{r}_{12}$. Thus by setting $\cos\theta_1 = 0$, $k = m$, $\cos\theta_2' = \cos\theta_{12}$, we compare coefficients to get

$$K = (-)^l[4\pi(2l)!/(2l-2n+1)!\,(2n)!]^{\frac{1}{2}},$$

and obtain for (11) a formula identical to (5.1) derived from detailed step-by-step expansions. In fact if we define a regular solid harmonic irreducible tensor as

$$R_n^k(\mathbf{r}) = \left[\frac{4\pi}{(2n+1)!}\right]^{\frac{1}{2}}\mathfrak{Y}_n^k(\mathbf{r}) = \left[\frac{4\pi}{(2n+1)!}\right]^{\frac{1}{2}} r^n Y_n^k(\theta, \phi),$$

and an irregular solid harmonic irreducible tensor as

$$I_n^k(\mathbf{r}) = [4\pi(2n)!]^{\frac{1}{2}}\mathfrak{T}_n^k(\mathbf{r}) = [4\pi(2n)!]^{\frac{1}{2}}\frac{Y_n^k(\theta, \phi)}{r^{n+1}},$$

then our Eq. (5.1) for one-center expansion and $r_2' > r_1$ becomes

$$I_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty}\sum_{m=-l}^{l} C(l-n, l, n; k-m, m, k) \times R_{l-n}^{k-m}(\mathbf{r}_1)I_l^m(-\mathbf{r}_2'); \quad (5.1a)$$

our Eq. (10.1) for one-center expansion becomes

$$R_n^k(\mathbf{r}_{12}) = \sum_{l=0}^{n}\sum_{m=-l}^{l} C(l, n-l, n; m, k-m, k) \times R_l^m(-\mathbf{r}_2')R_{n-l}^{k-m}(\mathbf{r}_1). \quad (10.1a)$$

The parity factor $(-)^l$ has been absorbed as we reverse the direction of $\mathbf{r}_2'$ to show the vector-addition character of the expansion.[21] Our Eq. (5.1) is a type of addition (coupling or displacement) theorem for the irregular solid harmonics; it seems not to exist in literature in such general form.

In the spirit of deriving (11) from (2), we rewrite into convenient forms Eq. (2) and the other reversed Hobson's formula [Eq. (9)] as follows, both using $\mathbf{r}_1$ as the polar axis:

$$\mathfrak{T}_n^{m'}(-\mathbf{r}_{12}) = \frac{Y_n^{m'}(\pi-\psi, \pi+\phi)}{r_{12}^{n+1}} = \sum_{l=n}^{\infty}(-)^{l+n}\left[\frac{(2l)!}{(2n)!\,(2l-2n)!}\right]^{\frac{1}{2}} \times C(l-n, l, n; 0, m', m') \times \frac{r_1^{l-n}}{(r_2')^{l+1}} Y_l^{m'}(\theta_{12}, \pi+\phi) \quad (2a)$$

$$\mathfrak{Y}_n^{m'}(-\mathbf{r}_{12}) = r_{12}^n Y_n^{m'}(\pi-\psi, \pi+\phi) = \sum_{l=0}^{n}(-)^{l+n}\left[\frac{(2n+1)!}{(2n+1)!\,(2n-2l)!}\right]^{\frac{1}{2}} \times C(n-l, l, n; 0, m', m') \times r_1^{n-l}(r_2')^l Y_l^{m'}(\theta_{12}, \pi+\phi), \quad (9a)$$

where we have made use of $Y_{l-n}^0(0, \phi) = [(2l-2n+1)/4\pi]^{\frac{1}{2}}$ and a similar expression for $Y_{n-l}^0(0, \phi)$.[22]

To obtain a two-center expansion, we express the coordinates $\mathbf{r}_2'$ of the particle e_2 with respect to the center A_1 in terms of the coordinates $\mathbf{r}_2$ of the same particle with respect to the center A_2 (see Fig. 1). For irregular solid harmonics, we turn to one-center expansions (5.1) and (6.1). With (5.1) we further distinguish two cases: $R > r_2$ and $r_2 > R$. For the latter case, we use Hobson's expansion,[18] namely the reverse of our (2) and (2a); for the former case we use a new Hobson's formula[13] not referred to before, which is, with the polar axis along $\mathbf{R}$,

$$\frac{P_l^m(\cos\theta_2')}{r_2'^{l+1}} = \sum_{S=m}^{\infty}(-)^{S+m} \times \frac{(S+l)!}{(S+m)!\,(l-m)!}\frac{r_2^S}{R^{l+S+1}}P_S^m(\cos\theta_2), \quad (12)$$

and which we rewrite as

$$\mathfrak{T}_l^m(\mathbf{r}_2') = \sum_{S=m}^{\infty}\left[\frac{(2l+2S+1)!}{(2l)!\,(2S+1)!}\right]^{\frac{1}{2}} \times C(l+S, S, l; 0, m, m)\frac{r_2^S}{R^{l+S+1}}Y_S^m(\theta_2, \phi_2). \quad (12a)$$

The resulting two-center expansions for $\mathbf{R}$ along the z axis are the following:

[21] There is some vague resemblance to the standard vector model for the coupling of angular momenta (Ref. 11, p. 34). However, aside from the summation sign, there are reservations about drawing analogy between the R's and the wavefunction ψ's and identifying l to angular momentum j, and $n-l$ to j_2, etc., as the wavefunction in the composite space may not be spherical harmonics (Ref. 11). Besides, the "angular momenta" l and $n-l$ here appear to add algebraically only, namely, the resulting $j = j_1 \pm j_2$ without intermediate values.

[22] To reduce our formulas (2), (9) or (2a), (9a) to those with a geometry same as that of Hobson (Ref. 11, pp. 140, 141) we simply replace r_1 by $-r_1$, vector $-\mathbf{r}_{12}$ by $\mathbf{r}_2'$, and scalar r_2' by r_{12}; then we have the reversed formula of $\mathfrak{T}_n^{m'}(\mathbf{r}_2)$ in terms of $Y_l^{m'}(\pi-\psi, \pi+\phi)$ and r_{12}: To reduce them to formulas with a geometry relation same as that of Pitzer *et al.* (Ref. 1), we simply multiply the right-hand side by $(-)^{n+m'}$ and replace $\pi-\psi$ by ψ.

for $r_2' > r_1; R > r_2$,

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{S=m}^{\infty} \sum_{m=-l}^{l} (-)^l \times \left[\frac{4\pi(2l+2S+1)!}{(2l-2n+1)!\,(2S+1)!\,(2n)!}\right]^{\frac{1}{2}} \times C(l+S, S, l; 0, m, m) \times C(l-n, l, n; k-m, m, k) \times \frac{r_1^{l-n} r_2^S}{R^{l+S+1}} Y_{l-n}^{k-m}(\theta_1, \phi_1) Y_S^m(\theta_2, \phi_2); \tag{5.2a}$$

for $r_2' > r_1; r_2 > R$,

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{S=l}^{\infty} \sum_{m=-l}^{l} (-)^l \times \left[\frac{4\pi(2S)!}{(2l-2n+1)!\,(2S-2l)!\,(2n)!}\right]^{\frac{1}{2}} \times C(S-l, S, l; 0, m, m) \times C(l-n, l, n; k-m, m, k) \times \frac{r_1^{l-n} R^{S-l}}{r_2^{S+1}} Y_{l-n}^{k-m}(\theta_1, \phi_1) Y_S^m(\theta_2, \phi_2). \tag{5.2b}$$

With (6.1) we simply use Hobson's formula or the reverse of our (9) and (9a). The resulting two-center expansion for **R** along z axis is the following:

for $r_1 > r_2'$,

$$\Upsilon_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{S=k-m}^{l-n} \sum_{m=-l}^{l} (-)^{l+n} \times \left[\frac{4\pi(2l)!}{(2l-2n-2S)!\,(2n)!\,(2S+1)!}\right]^{\frac{1}{2}} \times C(l-n-S, S, l-n; 0, k-m, k-m) \times C(l-n, l, n; k-m, m, k) \times \frac{R^{l-n-S} r_2^S}{r_1^{l+1}} Y_S^{k-m}(\theta_2, \phi_2) Y_l^m(\theta_1, \phi_1). \tag{6.2}$$

For regular solid harmonics we turn to the one-center expansion (10.1). Simple application of Hobson's formula or the reverse of our (9a) yields the two-center expansion for **R** along z axis,

$$\mathcal{Y}_n^k(\mathbf{r}_{12}) = \sum_{l=s}^{n} \sum_{S=m}^{l} \sum_{m=-l}^{l} (-)^l \times \left[\frac{4\pi(2n+1)!}{(2n-2l+1)!\,(2l-2S)!\,(2S+1)!}\right]^{\frac{1}{2}} \times C(l-S, S, l; 0, m, m) \times C(l, n-l, n; m, k-m, k) \times R^{l-S} r_2^S r_1^{n-l} Y_S^m(\theta_2, \phi_2) Y_{n-l}^{k-m}(\theta_1, \phi_1). \tag{10.2}$$

It is easy to infer that, if the vector between two-centers, **R**, is not along the z axis, but the coordinate system of center A_2 is parallel to that of center A_1, instead of (12a) we should use

$$\Upsilon_l^m(\mathbf{r}_2') = \sum_{S=m'}^{\infty} \sum_{m'=-S}^{S} \left[\frac{4\pi(2l+2S)!}{(2l)!\,(2S+1)!}\right]^{\frac{1}{2}} \times C(l+S, S, l; m-m', m', m) \times \frac{r_2^S}{R^{l+S+1}} Y_{l+S}^{m-m'}(\mathbf{R}) Y_S^{m'}(\theta_2, \phi_2), \tag{12b}$$

a formula derived from (6.1a)[17] by adding the appropriate parity factor due to the relative orientation of $\mathbf{r}_2'$, **R**, and $\mathbf{r}_2$. We have used vector **R** to stand for the polar and azimuth angle of this vector in the coordinate system of center A_1.

Using similar general expansions derived from (5.1) and (10.1), we have the following two-center expansions written in compact form by use of the notations preceding (5.1a) and (10.1a). Irregular solid harmonics, two-center, direction of **R** arbitrary:

for $r_2' > r_1; R > r_2$,

$$I_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{S=m'}^{\infty} \sum_{m=-l}^{l} \sum_{m'=-S}^{S} (-)^l \times C(l+S, S, l; m-m', m', m) \times C(l-n, l, n; k-m, m, k) \times I_{l+S}^{m-m'}(\mathbf{R}) R_S^{m'}(\mathbf{r}_2) R_{l-n}^{k-m}(\mathbf{r}_1); \tag{5.2aa}$$

for $r_2' > r_1; r_2 > R$,

$$I_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{S} \sum_{S=l}^{\infty} \sum_{m=-l}^{l} \sum_{m'=-S}^{S} (-)^l \times C(S-l, S, l; m-m', m', m) \times C(l-n, l, n; k-m, m, k) \times R_{S-l}^{m-m'}(\mathbf{R}) I_S^{m'}(\mathbf{r}_2) R_{l-n}^{k-m}(\mathbf{r}_1); \tag{5.2bb}$$

for $r_1 > r_2'$,

$$I_n^k(\mathbf{r}_{12}) = \sum_{l=n}^{\infty} \sum_{S=m'}^{l-n} \sum_{m=-l}^{l} \sum_{m'=-S}^{S} (-)^{l+n} \times C(l-n-S, S, l-n; k-m-m', m', k-m) \times C(l-n, l, n; k-m, m, k) \times R_{l-n-S}^{k-m-m'}(\mathbf{R}) R_S^{m'}(\mathbf{r}_2) I_l^m(\mathbf{r}_1). \tag{6.2aa}$$

Regular solid harmonics, two-center, direction of **R** arbitrary:

$$R_n^k(\mathbf{r}_{12}) = \sum_{l=S}^{n} \sum_{S=m'}^{l} \sum_{m=-l}^{l} \sum_{m'=-S}^{S} (-)^l \times C(l-S, S, l; m-m', m', m) \times C(l, n-l, n; m, k-m, k) \times R_{l-S}^{m-m'}(\mathbf{R}) R_S^{m'}(\mathbf{r}_2) R_{n-l}^{k-m}(\mathbf{r}_1). \tag{10.2aa}$$

After the parity factors $(-)^l$ and $(-)^{l+n}$ are

eliminated by inverting the direction of the intermediate vector $\mathbf{r}_2'$, the compact formulas above illustrate the *formal vector-addition (coupling) property of solid spherical harmonics*, namely $\mathbf{R} + \mathbf{r}_2 = \mathbf{r}_2'$; $-\mathbf{r}_2' + \mathbf{r}_1 = \mathbf{r}_{12}$. If the coordinate system of the second center differs from that of the first through a rotation of the first by Euler angles α, β, γ, we may replace $Y_S^m(\mathbf{r}_2)$ by $\sum_i D^S_{im'}(-\gamma, -\beta, -\alpha) Y_S^i(\mathbf{r}_2')$ where $\mathbf{r}_2'$ is the vector coordinate of the particle e_2 in terms of the new system.

For nonoverlapping charges and $n = k = 0$, our formulas (5.2a) and (5.2aa) reduce to the ($\mathrm{II_b}$) and ($\mathrm{III_b}$), respectively, of Carlson *et al.*[5] For overlapping as well as nonoverlapping charges, we compare our expansion for $n = k = 0$ with the bipolar expansion of $1/r_{12}$ by Buehler *et al.*[6] obtained from the Fourier coefficient method. Our (5.2a) reduces to their expansion for their Region I ($R > r_1 + r_2$), (5.2b) reduces to their expansion for Region III ($r_2 > R + r_1$), and (6.2) reduces to their expansion for Region IV ($r_1 > R + r_2$). Note, however, that we do not impose an absolute value sign on the m's, which can take negative as well as positive values here. While (5.2a), (5.2b), and (6.2) all may cover their Region II ($r_1 + r_2 \geq R \geq |r_1 - r_2|$), no attempt was made to reduce our formula to the general formula for Region II given in their addenda.[23] The convergence property of our expansions in this region remains to be explored. A general two-center expansion for solid harmonics of $n \geq 1$ in terms of solid harmonics (spherical polar coordinates), as far as the author is aware, does not exist in literature.

We have made extensive use of the Wigner (Clebsch–Gordan) coefficients, partly for compactness and partly for the physical interpretation they give to the different stages of coupling. The more symmetric $3 - j$ symbols are not used because the extra factors may obscure this picture. The explicit expression for the necessary Wigner coefficients are summarized below for completeness.

$$C(l, n, l-n; m, -k, m-k) = (-)^{l-m}\left(\frac{2l-2n+1}{2n+1}\right)^{\frac{1}{2}} C(l-n, l, n; k-m, m, k)$$

$$= (-)^{n-k}\left(\frac{2l-2n+1}{2l+1}\right)^{\frac{1}{2}} C(n, l-n, l; k, m-k, m)$$

$$= (-)^{n-k}\left(\frac{2l-2n+1}{2l+1}\right)^{\frac{1}{2}} C(l-n, n, l; m-k, k, m)$$

$$= (-)^{n-k}\left[\frac{(2l-2n+1)!\,(2n)!\,(l-m)!\,(l+m)!}{(2l+1)!\,(n-k)!\,(n+k)!\,(l-n-m+k)!\,(l-n+m-k)!}\right]^{\frac{1}{2}},$$

$$C(l+n, l, n; k-m, m, k) = (-)^{l+m}\left(\frac{2n+1}{2l+2n+1}\right)^{\frac{1}{2}} C(n, l, n+l; -k, m, m-k)$$

$$= (-)^{l+m}\left[\frac{(2n+1)!\,(2l)!\,(l+n+k-m)!\,(l+n-k+m)!}{(2l+2n+1)!\,(l-m)!\,(l+m)!\,(n-k)!\,(n+k)!}\right]^{\frac{1}{2}}.$$

All of the coefficients used in this paper may be obtained by symmetry properties[16] or by changing the dummy indices l, n, k, m, or by setting $k = m$.

Note added after completion of the manuscript. The author recently had the pleasure of a talk with Robert A. Sack, from whom he understands that Dr. Sack, by a differential equation method (to be published), has been doing the one-center expansion of $r_{ij}^n P_{n'}^k(\cos \Theta_{ij})$ and the two-center expansion of r_{ij}^n, where n may be equal or unequal to n' and may be a negative as well as a fractional number. It appears that our derivation here puts more emphasis on the method and the physical nature of the problem.

ACKNOWLEDGMENT

The author takes pleasure in thanking Professor Robert S. Mulliken for making his stay at the University of Chicago possible.

[23] R. J. Buehler and J. O. Hirschfelder, Phys. Rev. **85**, 149 (1952).

Some Properties of Triangular Representations of $SU(3)$*

S. P. ROSEN
Department of Physics, Purdue University, Lafayette, Indiana
(Received 26 September 1963)

The occurrence of an equal-spacing mass rule in the unitary decuplet can be explained either by observing that the isotopic spin and hypercharge of each particle are related by $T = 1 + \frac{1}{2}Y$, or by making use of a theorem due to Diu and Ginibre. This theorem states that, in all triangular representations of $SU(3)$, the matrix elements of an arbitrary tensor operator depend upon one reduced matrix element instead of two. Here we present a new proof of the Diu–Ginibre theorem and use it show that relations of the form $T = \lambda \pm \frac{1}{2}Y$ exist for all triangular representations. We also show that the L and K spins are related to their corresponding hypercharges by $L = \lambda \pm \frac{1}{2}Y_L$ and $K = \lambda \pm \frac{1}{2}Y_K$. One consequence is that the masses and magnetic moments of particles in a triangular multiplet are equally spaced. Other consequences are also discussed.

1. INTRODUCTION

IT has been pointed out by Gell-Mann[1] and Glashow and Sakurai[2] that the mass spectrum of a ten-dimensional unitary multiplet (or decuplet[3]) is governed by an "equal-spacing" rule. The simplest way to explain this rule is to notice that, as a result of the relation

$$T = 1 + \tfrac{1}{2}Y \tag{1}$$

between the isotopic spin T and hypercharge Y of each particle in the decuplet, Okubo's first-order mass formula[4] reduces to a linear function of hypercharge.[2] A more sophisticated explanation follows from a theorem proved by Diu[5] and by Ginibre[6] in connection with the Wigner–Eckart theorem for $SU(3)$. Here we wish to demonstrate the equivalence of these two explanations by showing how the relation (1) can be derived from the Diu–Ginibre theorem. In fact we show that there exist relations of the form

$$T = \lambda \pm \tfrac{1}{2}Y, \tag{2}$$

where λ is independent of T and Y, for a whole class of unitary multiplets; consequently, the corresponding mass spectra are all equally spaced.

The multiplets of interest are those that form bases of the triangular representations[7] of $SU(3)$. A triangular representation is one whose characteristic numbers[8] (f_1, f_2, f_3) satisfy either

$$f_1 = f_2, \tag{3a}$$

or

$$f_2 = f_3, \tag{3b}$$

and is so called because its dimension[8] is a triangular number,[9] and its weight diagram[10] a triangle. The decuplet is characterized by the numbers $(2, -1, -1)$ and satisfies (3b). [It should be noted that many authors[5–7,10] use two numbers to characterize a representation of $SU(3)$; they are

$$\mu_1 = f_1 - f_2, \qquad \mu_2 = f_2 - f_3, \tag{4}$$

and the triangular conditions (3a), (3b) become

$$\mu_1 = 0, \tag{5a}$$

$$\mu_2 = 0, \tag{5b}$$

respectively].

According to the Diu–Ginibre theorem, the matrix elements of any traceless tensor R^μ_ν in a triangular representation depend upon one reduced matrix element instead of two. More precisely, in an arbitrary representation, each matrix element of R^μ_ν is a linear combination of the corresponding matrix elements of two tensors, B^μ_ν and C^μ_ν, which are constructed from the infinitesimal generators of $SU(3)$ (see Ref. 4 and the second section below); the matrix elements of B^μ_ν and C^μ_ν are linearly independent in all representations except the

* Work supported in part by the United States Air Force.
[1] M. Gell-Mann, *Proceedings of the 1962 International Conference on High Energy Physics* (CERN, Geneva, Switzerland, 1962), p. 805.
[2] S. L. Glashow and J. J. Sakurai, Nuovo Cimento **26**, 622 (1962).
[3] S. L. Glashow and A. H. Rosenfeld, Phys. Rev. Letters **10**, 192 (1963).
[4] S. Okubo, Progr. Theoret. Phys. (Kyoto) **27**, 949 (1961).
[5] B. Diu, Nuovo Cimento **28**, 466 (1963).
[6] J. Ginibre, J. Math. Phys. **4**, 720 (1963).
[7] S. Gasiorowicz, Argonne National Laboratory Report, ANL-6729 (1963).
[8] H. Weyl, *The Classical Theory of Groups* (Princeton University Press, Princeton, New Jersey, 1939); see also A. R. Edmonds [Proc. Roy. Soc. (London) **A268**, 567 (1962)], who discusses the conditions under which the representations of $U(3)$ and $SU(3)$ are equivalent.
[9] G. James and R. C. James, *Mathematics Dictionary* (D. Van Nostrand Company, Inc., New York, 1959), 2nd ed., p. 271.
[10] R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).

triangular ones. As a result, mass differences in a triangular multiplet are proportional to differences in the expectation values of B_3^3,[4,5] i.e., the hypercharge.

In order to derive (1) and (2) from this theorem, we present another proof which is similar in spirit to that of Diu,[5] but which differs considerably from it in its details. By means of various identities among the products of infinitesimal generators, we show that a relation

$$C_\nu^\mu = gB_\nu^\mu \tag{6}$$

holds only for matrix elements in a triangular representation. The coefficient g is determined in terms of the appropriate f_i. Equations (1) and (2) then follow from the identification of certain generators with isotopic spin and hypercharge.

Another outcome of our work concerns the various ways of classifying particles within a given multiplet. Besides the usual classification based on isotopic spin and hypercharge, there are two others,[11] each of which involves an isotopic spin type quantity and a corresponding hypercharge; the appropriate pairs of quantum numbers are denoted by[12] (K, Y_K) and (L, Y_L). We show that whenever (2) is satisfied, the corresponding relations

$$K = \lambda \pm \tfrac{1}{2}Y_K, \tag{7}$$

$$L = \lambda \pm \tfrac{1}{2}Y_L \tag{8}$$

are also satisfied.

Equation (7) is not without physical significance, for the K spin plays a very important role in electromagnetic interactions,[13] It has, indeed, been shown[12] that the lowest-order formula for magnetic moments has exactly the same form as Okubo's first-order mass formula,[4] with (K, Y_K) substituted for (T, Y). Consequently, whenever the mass spectrum of a multiplet is equally spaced, the spectrum of magnetic moments is also equally spaced.

Our proof of the Diu–Ginibre theorem is given in the next section, and the relation between isotopic spin and hypercharge is derived in the third. The paper concludes with a detailed summary of our results, and a discussion of further consequences for triangular multiplets.

[11] C. A. Levinson, H. J. Lipkin, and S. Meshkov, Nuovo Cimento **23**, 236 (1961); Phys. Letters **1**, 44 (1962); Phys Rev. Letters **10**, 361 (1963). These authors use the symbol U instead of K.

[12] S. P. Rosen, Phys. Rev. Letters **11**, 100 (1963). Notice that the present definitions of L and the hypercharges are slightly different from those of this reference.

[13] C. A. Levinson, H. J. Lipkin, and S. Meshkov, Phys. Letters, **7**, 81 (1963).

2. THE DIU-GINIBRE THEOREM

From the infinitesimal generators A_ν^μ of $U(3)$,[8] it is possible to construct only two independent tensor operators with zero trace. They are

$$\begin{aligned} B_\nu^\mu &= A_\nu^\mu - \tfrac{1}{3}\delta_\nu^\mu M_1, \\ C_\nu^\mu &= A_\alpha^\mu A_\nu^\alpha - \tfrac{1}{3}\delta_\nu^\mu M_2, \end{aligned} \tag{9}$$

where M_1 and M_2 are two of the three Casimir operators[4]

$$M_1 = A_\alpha^\alpha, \quad M_2 = A_\beta^\alpha A_\alpha^\beta, \quad M_3 = A_\beta^\alpha A_\gamma^\beta A_\alpha^\gamma, \tag{10}$$

associated with $U(3)$. Okubo's analog of the Wigner–Eckart theorem states that the matrix elements of any tensor operator R_ν^μ can be expressed as[4]

$$\begin{aligned} \langle j\{f\} \,|R_\nu^\mu|\, k\{f\}\rangle &= b\langle j\{f\} \,|B_\nu^\mu|\, k\{f\}\rangle \\ &\quad + c\langle j\{f\} \,|C_\nu^\mu|\, k\{f\}\rangle, \end{aligned} \tag{11}$$

where $|j\{f\}\rangle$ and $|k\{f\}\rangle$ are basis vectors of a given representation $\{f\} \equiv U(f_1, f_2, f_3)$.[4] The coefficients b and c are independent of the suffices μ, ν ,and of the quantum numbers represented by j, k, (e.g., isotopic spin and hypercharge), but may depend on f_1, f_2, f_3. To prove the Diu–Ginibre theorem, we must show that the relation

$$\langle j\{f\} \,|C_\nu^\mu|\, k\{f\}\rangle = g\langle j\{f\} \,|B_\nu^\mu|\, k\{f\}\rangle \tag{12}$$

holds only when $\{f\}$ is a triangular representation of $U(3)$.

Our proof is based upon an analysis of the operator

$$D_\nu^\mu = A_\alpha^\mu A_\beta^\alpha A_\nu^\beta. \tag{13}$$

We assume that (12) holds in some representation $\{f\}$ and then express $\langle j\{f\}|\, D_\nu^\mu \,|k\{f\}\rangle$ as a linear combination of $\langle j\{f\}|\, A_\nu^\mu \,|k\{f\}\rangle$ and $\langle j\{f\}|\, \delta_\nu^\mu \,|k\{f\}\rangle$ in two different ways. Since the matrix elements of A_ν^μ and δ_ν^μ are linearly independent, we can equate their coefficients in the two expressions for $\langle j\{f\}|\, D_\nu^\mu \,|k\{f\}\rangle$ and obtain two equations for the coefficient g. The condition for the validity of our original assumption about (12) is found by eliminating g from these equations. It is satisfied in only two cases: (i) $f_1 = f_2$, and (ii) $f_2 = f_3$.

We need the following properties of the Casimir operators (10) and of B_ν^μ, C_ν^μ. Since the commutation relations

$$[A_\beta^\alpha, A_\nu^\mu] = \delta_\beta^\mu A_\nu^\alpha - \delta_\nu^\alpha A_\beta^\mu \tag{14}$$

imply that M_1, M_2, and M_3 commute with every A_ν^μ and with each other, the Casimir operators are constant in each representation of $U(3)$; their eigenvalues in the representation $U(f_1, f_2, f_3)$ are[4]

$$M_1 = -(f_1 + f_2 + f_3),$$
$$M_2 = f_1^2 + f_2^2 + f_3^2 + 2(f_1 - f_3),$$
$$M_3 = -(f_1^3 + f_2^3 + f_3^3) + (-\tfrac{3}{2}f_1^3 + \tfrac{3}{2}f_2^2 + \tfrac{3}{2}f_3^2) - \tfrac{1}{2}(f_1 + f_2 + f_3)^2 + 2(f_1 + f_2 - 2f_3). \quad (15)$$

Both operators B^μ_ν and C^μ_ν commute with the Casimir operators, and hence,

$$\langle j\{f\} \; |B^\mu_\nu| \; k\{f'\}\rangle = \langle j\{f\} \; |C^\mu_\nu| \; k\{f'\}\rangle = 0, \quad \text{for} \quad \{f\} \neq \{f'\}. \quad (16)$$

The first expression for the matrix element of D^μ_ν is obtained by rewriting (13) in the form

$$D^\mu_\nu = (C^\mu_\beta + \tfrac{1}{3}\delta^\mu_\beta M_2)(C^\beta_\nu + \tfrac{1}{3}\delta^\nu_\beta M_2). \quad (17)$$

We now take the matrix element of (17), insert a complete set of states between the factors of the right-hand side, and make use of (16) to obtain

$$\langle j\{f\} \; |D^\mu_\nu| \; k\{f\}\rangle = \sum_l \langle j\{f\} \; |C^\mu_\beta + \tfrac{1}{3}\delta^\mu_\beta M_2| \; l\{f\}\rangle \times \langle l\{f\} \; |C^\beta_\nu + \tfrac{1}{3}\delta^\beta_\nu M_2| \; k\{f\}\rangle.$$

Next we assume that (12) is true in $\{f\}$, so that

$$\langle j\{f\} \; |D^\mu_\nu| \; k\{f\}\rangle = \sum_l \langle j\{f\} \; |gA^\mu_\beta + \tfrac{1}{3}\delta^\mu_\beta(M_2 - gM_1)| \; l\{f\}\rangle \times \langle l\{f\} \; |gA^\beta_\nu + \tfrac{1}{3}\delta^\beta_\nu(M_2 - gM_1)| \; k\{f\}\rangle.$$

Using (16) again, we can replace the sum over l by a sum over a complete set of states; then from closure, the definition of C^μ_ν (9), and Eq. (12) we obtain

$$9\langle j\{f\} \; |D^\mu_\nu| \; k\{f\}\rangle = [9g^3 + 6g(M_2 - gM_1)] \times \langle j\{f\} \; |A^\mu_\nu| \; k\{f\}\rangle + (M_2 - gM_1) \times [3g^2 + M_2 - gM_1]\langle j\{f\} \; |\delta^\mu_\nu| \; k\{f\}\rangle. \quad (18)$$

The second expression is found by writing D^μ_ν in the form

$$D^\mu_\nu = (A^\mu_\alpha A^\alpha_\beta A^\beta_\gamma)A^\gamma_\nu,$$

and then using the identity[14]

$$6(A^\lambda_\alpha A^\alpha_\beta A^\beta_\rho) = 6(M_1 + 3)C^\lambda_\rho - 3[(M_1 + 2)^2 - M_2]B^\lambda_\rho + 2M_3\delta^\lambda_\rho$$

to express D^μ_ν as linear combination of δ^μ_ν, B^μ_ν and C^μ_ν. With the assumption that (12) holds in $\{f\}$, it follows that

$$9\langle j\{f\} \; |D^\mu_\nu| \; k\{f\}\rangle = [\tfrac{3}{2}(3g - M_1)(M_2 + 2g(M_1 + 3) - (M_1 + 2)^2) + 3M_3]\langle j\{f\} \; |A^\mu_\nu| \; k\{f\}\rangle + \tfrac{3}{2}(M_2 - gM_1)(M_2 + 2g(M_1 + 3) - (M_1 + 2)^2)\langle j\{f\} \; |\delta^\mu_\nu| \; k\{f\}\rangle. \quad (19)$$

[14] See Ref. 4, Theorem III.

A comparison of the coefficients of $\langle j\{f\}| \; \delta^\mu_\nu \; |k\{f\}\rangle$ in (18) and (19) yields

$$3(M_2 + 2g(M_1 + 3) - (M_1 + 2)^2) = 6g^2 + 2(M_2 - gM_1), \quad (20)$$

or equivalently

$$6g^2 - 2g(4M_1 + 9) + 3(M_1 + 2)^2 - M_2 = 0. \quad (21)$$

With the aid of (20), the coefficients of $\langle j\{f\}| \; A^\mu_\nu \; |k\{f\}\rangle$ lead to a second equation in g, namely,

$$g(3M_2 - M_1^2) = 3M_3 - M_1M_2. \quad (22)$$

The condition for (21) and (22) to have a common root is

$$6(3M_3 - M_1M_2)^2 - 2(4M_1 + 9)(3M_3 - M_1M_2)(3M_2 - M_1^2) + [3(M_1 + 2)^2 - M_2](3M_2 - M_1^2)^2 = 0. \quad (23)$$

In order to solve (23), we reexpress it in terms of the f_i [see (15)]; the resulting equation is[15]

$$(f_1 - f_2)(f_2 - f_3)(f_1 - f_3 + 1) \times \{(f_1 - f_2)(f_2 - f_3)(f_1 - f_3 + 3) + 2(f_1 - f_3)(f_1 - f_3 + 5) + 12\} = 0. \quad (24)$$

Because[8]

$$f_1 \geq f_2 \geq f_3, \quad (25)$$

the factor in the curly bracket of (24) and $(f_1 - f_3 + 1)$ are always greater than zero; the only factors that can vanish are the first two. Therefore the only cases in which (24) is satisfied are

$$f_1 = f_2, \quad (3a)$$

and

$$f_2 = f_3. \quad (3b)$$

The corresponding values of g are [see (15) and (22)]

$$g = \begin{cases} 2 - f_2 - f_3 & (f_1 = f_2), \\ 1 - f_1 - f_2 & (f_2 = f_3). \end{cases} \quad (26)$$

3. RELATION BETWEEN ISOTOPIC SPIN AND HYPERCHARGE

As a consequence of the commutation rules (14) and the unitary restriction[4]

$$(A^\mu_\nu)^\dagger = A^\nu_\mu, \quad (27)$$

[15] The easiest way to obtain (24) is to observe that (23) is invariant under $f_i \to f_i + \epsilon (i = 1, 2, 3)$. A subsidiary condition, $M_1 = 0$, can therefore be imposed on the f_i.

the infinitesimal generators of $U(3)$ can be divided into three sets[12]:

$$T_+ = -A_1^2, \quad T_- = -A_2^1, \quad T_3 = \tfrac{1}{2}(A_2^2 - A_1^1),$$
$$Y_T = A_3^3 - \tfrac{1}{3}M_1 \equiv B_3^3, \qquad (28)$$

$$L_+ = -A_3^1, \quad L_- = -A_1^3,$$
$$L_3 = \tfrac{1}{2}(A_1^1 - A_3^3), \quad Y_L = B_2^2, \qquad (29)$$

$$K_+ = -A_2^3, \quad K_- = -A_3^2,$$
$$K_3 = \tfrac{1}{2}(A_3^3 - A_2^2), \quad Y_K = B_1^1, \qquad (30)$$

each of which contains an angular-momentum-type operator and a corresponding hypercharge. It is customary to identify the first set (28) with isotopic spin and the usual hypercharge,

$$Y_T = B + S, \qquad (31)$$

where B denotes baryon number and S strangeness. The operators of the second and third sets can be expressed in terms of Y_T and the electric charge Q[12]:

$$L_3 = -\tfrac{1}{2}(Q + Y_T), \quad Y_L = (Q - Y_T),$$
$$K_3 = Y_T - \tfrac{1}{2}Q, \quad Y_K = -Q. \qquad (32)$$

Every particle in a unitary multiplet is a simultaneous eigenstate of T_3, L_3, K_3, Y_T, Y_L, Y_K, but it is not always an eigenstate of all three operators $\mathbf{T}^2$, $\mathbf{L}^2$, $\mathbf{K}^2$.

The operators in (28), (29), and (30), satisfy several identities which are useful in the following discussion. They are

$$T_3 + L_3 + K_3 \equiv 0, \qquad (33)$$

$$Y_T + Y_L + Y_K \equiv 0, \qquad (34)$$

and[4,16]

$$C_1^1 \equiv \alpha + \beta Y_K + \tfrac{1}{4}Y_K^2 - \mathbf{K}^2,$$
$$C_2^2 \equiv \alpha + \beta Y_L + \tfrac{1}{4}Y_L^2 - \mathbf{L}^2, \qquad (35)$$
$$C_3^3 \equiv \alpha + \beta Y_T + \tfrac{1}{4}Y_T^2 - \mathbf{T}^2,$$

with $\alpha = \frac{1}{6}(M_2 - \frac{1}{3}M_1^2)$, $\beta = (\frac{3}{2} + \frac{2}{3}M_1)$. Since C_ν^μ is a traceless tensor, Eqs. (34) and (35) yield another identity,

$$\mathbf{T}^2 + \mathbf{L}^2 + \mathbf{K}^2 \equiv \tfrac{1}{2}(M_2 - \tfrac{1}{3}M_1^2) + \tfrac{1}{4}(Y_T^2 + Y_L^2 + Y_K^2). \qquad (36)$$

In order to derive the relation between T and Y_T, we shall require the following lemma:

Lemma. Every particle in a triangular multiplet is a simultaneous eigenstate of $\mathbf{T}^2$, $\mathbf{L}^2$, $\mathbf{K}^2$.

Proof: From (35) and the commutation rules for an arbitrary tensor operator,[17]

$$[A_\beta^\alpha, C_\nu^\mu] = \delta_\beta^\mu C_\nu^\alpha - \delta_\nu^\alpha C_\beta^\mu,$$

it follows that $\mathbf{T}^2$, $\mathbf{L}^2$, $\mathbf{K}^2$ commute with T_3, L_3, K_3, Y_T, Y_L, Y_K.

From the Diu–Ginibre theorem (12) and an argument similar to the one leading from (17) to (18), it can be shown that

$$\langle j\{f\} \,|[C_\beta^\alpha, C_\nu^\mu]|\, k\{f\}\rangle = g^2\langle j\{f\} \,|[A_\beta^\alpha, A_\nu^\mu]|\, k\{f\}\rangle, \qquad (37)$$

whenever $\{f\}$ is a triangular representation. Taken together with (14) and (35), Eq. (37) implies that in a triangular representation of $U(3)$ the matrices corresponding to $\mathbf{T}^2$, $\mathbf{L}^2$, and $\mathbf{K}^2$, commute with one another. Whence the lemma.

It now follows that each particle in a triangular multiplet can be assigned six quantum numbers $(T', Y_T'; L', Y_L'; K', Y_K')$. (Note that we use primes to distinguish eigenvalues from their corresponding operators).

Consider the expectation value of C_3^3 for a particle m which belongs to a multiplet with $f_2 = f_3$: from (12) and (26),

$$\langle m \,|C_3^3|\, m\rangle = (1 - f_1 - f_2)\langle m \,|B_3^3|\, m\rangle. \qquad (38)$$

By means of (35), (28), and the eigenvalues of M_1, M_2 [see (15)], we reduce (38) to

$$T'(T' + 1) = (\tfrac{1}{2}Y_T' + \tfrac{1}{3}(f_1 - f_2)) \times (\tfrac{1}{2}Y_T' + \tfrac{1}{3}(f_1 - f_2) + 1). \qquad (39)$$

Similarly, from the expectation values of C_2^2 and C_1^1 we obtain

$$L'(L' + 1) = (\tfrac{1}{2}Y_L' + \tfrac{1}{3}(f_1 - f_2)) \times (\tfrac{1}{2}Y_L' + \tfrac{1}{3}(f_1 - f_2) + 1) \qquad (40)$$

and

$$K'(K' + 1) = (\tfrac{1}{2}Y_K' + \tfrac{1}{3}(f_1 - f_2)) \times (\tfrac{1}{2}Y_K' + \tfrac{1}{3}(f_1 - f_2) + 1). \qquad (41)$$

Each of Eqs. (39), (40), (41) has two possible solutions,

$$X' = \tfrac{1}{2}Y_X' + \tfrac{1}{3}(f_1 - f_2), \qquad (42a)$$

and

$$X' = -\tfrac{1}{2}Y_X' - \tfrac{1}{3}(f_1 - f_2) - 1, \qquad (42b)$$

and it follows that if we consider the set of three solutions, one for T', one for L', and one for K',

[16] S. P. Rosen, Phys. Rev. **132**, 1234 (1963).

[17] See Ref. 4, Eq. (A.4).

there are eight different possibilities. We must therefore find an argument for rejecting all but one of them.

Since T, K, L, have been introduced into the analysis in a highly symmetrical fashion [see for example (33), (34) and (36)], the relations between T', K', L', and their corresponding hypercharges should all be of the same form; we may therefore reject all those sets in which two relations are given by (42a), and one by (42b), or vice versa. Suppose now that each relation is of the form (42b): then

$$T' + L' + K' = -\tfrac{1}{2}(Y'_T + Y'_L + Y'_K) - (f_1 - f_2) - 3. \quad (43)$$

From (33),

$$Y'_T + Y'_L + Y'_K = \langle m \,|Y_T + Y_L + Y_K|\, m\rangle = 0,$$

and so (43) becomes

$$T' + L' + K' = -(f_1 - f_2) - 3. \quad (44)$$

But T', L', K', and $(f_1 - f_2)$ are all positive numbers, and hence (44) can never be satisfied. On the other hand, if the relations are all given by (42a), we obtain an equation

$$T' + L' + K' = (f_1 - f_2), \quad (45)$$

which can be satisfied. We therefore conclude that the quantum numbers of a particle belonging to a triangular multiplet with $f_2 = f_3$ obey

$$\begin{aligned} T' &= \tfrac{1}{2}Y'_T + \tfrac{1}{3}(f_1 - f_2), \\ L' &= \tfrac{1}{2}Y'_L + \tfrac{1}{3}(f_1 - f_2), \\ K' &= \tfrac{1}{2}Y'_K + \tfrac{1}{3}(f_1 - f_2). \end{aligned} \quad (46)$$

In a similar way, we can show that for a particle in a multiplet with $f_1 = f_2$,

$$\begin{aligned} T' &= \tfrac{1}{3}(f_2 - f_3) - \tfrac{1}{2}Y'_T, \\ L' &= \tfrac{1}{3}(f_2 - f_3) - \tfrac{1}{2}Y'_L, \\ K' &= \tfrac{1}{3}(f_2 - f_3) - \tfrac{1}{2}Y'_K. \end{aligned} \quad (47)$$

To illustrate the use of these formulas, we consider the decuplet with $f_1 = 2$, $f_2 = f_3 = -1$. From (46) the quantum numbers of each particle belonging to it satisfy

$$\begin{aligned} T' &= 1 + \tfrac{1}{2}Y'_T, \\ L' &= 1 + \tfrac{1}{2}Y'_L, \\ K' &= 1 + \tfrac{1}{2}Y'_K. \end{aligned} \quad (48)$$

The conjugate decuplet is characterized by $f_1 = f_2 = 1$, $f_3 = -2$, and from (47),

$$\begin{aligned} T' &= 1 - \tfrac{1}{2}Y'_T, \\ L' &= 1 - \tfrac{1}{2}Y'_L, \\ K' &= 1 - \tfrac{1}{2}Y'_L. \end{aligned} \quad (49)$$

4. CONCLUSIONS

We have proved the Diu–Ginibre–theorem[5,6] by showing that in any triangular representation of $SU(3)$ [see (3a), (3b)], the matrix elements of C^μ_ν are proportional to the corresponding matrix elements of B^μ_ν [see (9), (12), (26)]. Using this form of the theorem, we have shown that any particle belonging to a triangular multiplet is a simultaneous eigenstate of the operators $\mathbf{T}^2$, $\mathbf{L}^2$, $\mathbf{K}^2$ [see (28), (29), (30)], and that each of its quantum numbers T', L', K', is linearly related to the corresponding hypercharge. Details of these relations are given in (46) and (47).

Some of the physical consequences of (46) and (47), namely the equal-spacing rules, have already been discussed in the introduction. One other consequence of physical interest is that no triangular multiplet contains two or more particles with the same hypercharge, but different isotopic spins. This behavior is to be contrasted with that of non-triangular multiplets, all of which contain at least one pair of particles which are degenerate with respect to hypercharge.

In conclusion, it is amusing to notice that (46) and (47) can be used to demonstrate that the weight diagram[10] of a triangular representation is indeed a triangle. The number of particles with a given hypercharge is $(2T' + 1)$ which is equal to $1 + \frac{2}{3}(f_1 - f_2) + Y'_T$ [from (46)], or $1 + \frac{2}{3}(f_2 - f_3) - Y'_T$ [from (47)]; therefore as Y'_T increases by steps of one unit, the number of points in successive rows of the weight diagram increases (46), or decreases (47), by one. Moreover, all unitary multiplets have one particle with $T' = 0$[7], and so the initial row, or the final row, contains only one point.

ACKNOWLEDGMENT

Part of this work was carried out at the 1963 Summer Institute for Theoretical Physics, University of Wisconsin. The author would like to thank Professor R. G. Sachs for the hospitality of the Institute.

The Representations of the Inhomogeneous Lorentz Group in Terms of an Angular Momentum Basis

J. S. LOMONT*
Polytechnic Institute of Brooklyn, Brooklyn, New York
AND
H. E. MOSES
Lincoln Laboratory,† Massachusetts Institute of Technology, Lexington, Massachusetts
(Received 26 September 1963)

The irreducible ray representations of the proper, orthochronous, inhomogeneous Lorentz group were originally given by Wigner in terms of a basis in which the energy and linear momenta are diagonal. In the present paper we show how the infinitesimal generators of the irreducible representations act on a basis in which the energy, the square of the angular momentum, the component of the angular momentum along the z axis, and the helicity (or circular polarization) are diagonal.

We consider representations corresponding to particles of nonzero mass, and any spin and of zero mass and finite spin. The continuous-spin case is to be treated in a later paper.

I. INTRODUCTION

THE irreducible unitary ray representations of the proper, orthochronous, inhomogeneous Lorentz group were first given in global form by Wigner.[1] The basis in Hilbert space which is used to express the operators is one in which the linear momenta are diagonal. Subsequent workers who have worked in the field have, for the most part, also used bases in which the linear momentum operators are diagonal.[2-6] An interesting exception is where the authors express the irreducible representations of the inhomogeneous Lorentz group in terms of the irreducible representations of the homogeneous Lorentz group,[7,8] the effect being that the irreducible representations of the inhomogeneous group are given in terms of a basis in which the Casimir operators or invariants of the homogeneous group are diagonal.

However, for physical applications it is very useful to express the irreducible representations in terms of a basis in which the energy, the square of the angular momentum, the z component of the angular momentum, and the helicity are diagonal. While it is difficult to give the global form of the representations in this basis, it is possible to show explicitly how the infinitesimal generators operate.

The object of the present paper, then, is to give explicit formulas which show how the infinitesimal generators operate in the angular momentum basis which we have described above. We restrict ourselves to the physically interesting cases of nonzero mass and arbitrary spin, and zero-mass, discrete-spin cases. The continuous-spin case is to be given in a later paper.

It should be mentioned that Pauli[9] gives a very terse derivation for the massless, discrete-spin case. However, the final results for the generators are not given explicitly.

We give the derivation of all our results in a subsequent paper. However, it is possible to verify the results directly by showing that the operators are Hermitian, satisfy the necessary commutation rules, and have the correct values for the invariants.

We denote the ten infinitesimal generators of the irreducible representations by P^μ and $J^{\mu\nu} = -J^{\nu\mu}$ ($\mu, \nu = 0, 1, 2, 3$; $g^{\mu\nu} = g_{\mu\nu} = 0$ if $\mu \neq \nu$, $g^{00} = g_{00} = -1$, $g_{11} = g_{22} = g_{33} = 1$). The commutation rules which these Hermitian operators are required to satisfy are

$$[P^\mu, P^\nu] = 0, \tag{1.1}$$

$$[J_{\mu\nu}, P_\alpha] = i[g_{\mu\alpha}P_\nu - g_{\nu\alpha}P_\mu], \tag{1.2}$$

$$[J_{\mu\nu}, J_{\alpha\beta}] = i[g_{\mu\alpha}J_{\nu\beta} - g_{\nu\alpha}J_{\mu\beta} + g_{\mu\beta}J_{\alpha\nu} - g_{\nu\beta}J_{\alpha\mu}]. \tag{1.3}$$

* This work was performed while working at Lincoln Laboratory.
† Operated with support from the United States Advanced Research Projects Agency.

[1] E. P. Wigner, Ann. Math. 40, 149 (1939).
[2] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U. S., 34, 211 (1948).
[3] L. L. Foldy, Phys. Rev. 102, 568 (1956).
[4] Iu. M. Shirokov, Zh. Eksperim. i Teor. Fiz. 33, 1196 (1957) [English transl.: Soviet Phys.—JETP 6, 919 (1958)].
[5] V. I. Ritus, Zh. Eksperim. i Teor. Fiz. 40, 352 (1961) [English transl.: Soviet Phys.—JETP 13, 240 (1961)].
[6] J. S. Lomont and H. E. Moses, J. Math. Phys. 3, 405 (1962).
[7] I. S. Shapiro, Dokl. Akad. Nauk SSSR 4, 647 (1955) [English transl.: Soviet Phys.—Doklady 1, 91 (1956)].
[8] Chou Kuang-Chao and L. G. Zastavenko, Zh. Experim. i Teor. Fiz. 35, 1417 (1958) [English transl.: Soviet Phys.—JETP 8, 990 (1959)].
[9] W. Pauli, CERN Rept. 56-31, Geneva (1956).

On introducing the conventional notation $J_1=J_{23}$, $J_2 = J_{31}$, $J_3 = J_{12}$, $\mathcal{J}_i = J_{0i}$ $(i = 1, 2, 3)$, and $H = P^0$, the commutation relations may be written

$$\begin{aligned}[H, P_i] &= 0,\\ [P_i, P_j] &= 0;\end{aligned} \tag{1.4}$$

$$\begin{aligned}[J_i, P_i] &= 0,\\ [J_i, H] &= 0,\\ [J_i, \mathcal{J}_i] &= 0;\end{aligned} \tag{1.5}$$

$$\begin{aligned}[\mathcal{J}_i, H] &= iP_i,\\ [\mathcal{J}_i, P_j] &= i\delta_{ij}H;\end{aligned} \tag{1.6}$$

$$\begin{aligned}[J_1, J_2] &= iJ_3,\\ [J_2, J_3] &= iJ_1,\\ [J_3, J_1] &= iJ_2;\end{aligned} \tag{1.7}$$

$$\begin{aligned}[\mathcal{J}_1, \mathcal{J}_2] &= -iJ_3,\\ [\mathcal{J}_2, \mathcal{J}_3] &= -iJ_1,\\ [\mathcal{J}_3, \mathcal{J}_1] &= -iJ_2;\end{aligned} \tag{1.8}$$

$$\begin{aligned}[J_1, P_2] = [P_1, J_2] &= iP_3,\\ [J_2, P_3] = [P_2, J_3] &= iP_1,\\ [J_3, P_1] = [P_3, J_1] &= iP_2;\end{aligned} \tag{1.9}$$

$$\begin{aligned}[J_1, \mathcal{J}_2] = [\mathcal{J}_1, J_2] &= i\mathcal{J}_3,\\ [J_2, \mathcal{J}_3] = [\mathcal{J}_2, J_3] &= i\mathcal{J}_1,\\ [J_3, \mathcal{J}_1] = [\mathcal{J}_3, J_1] &= i\mathcal{J}_2.\end{aligned} \tag{1.10}$$

The operator H is the Hamiltonian, the operators J_i $(i = 1, 2, 3)$ are components of the angular momentum, the operator

$$\mathbf{P}\cdot\mathbf{J} = \sum_{i=1}^{3} P_iJ_i$$

is the helicity operator, and

$$\mathbf{J}^2 = \sum_{i=1}^{3} J_i^2$$

is the square of the angular momentum.

2. REPRESENTATION WITH NONVANISHING MASS

Let us denote the mass of the particle by μ and the spin of the particle by s, where s is a nonnegative integer or half-odd integer. We now introduce a space of complex functions $\varphi(E, j, m, \alpha)$ in which E is a continuous variable $\mu < E < \infty$ for positive-energy representations and $-\infty < E < -\mu$ for negative-energy representations. The variable α takes on the values $\alpha = -s, -s + 1, -s + 2, \cdots, s - 2, s - 1, s$. For a fixed value of α, the variable j takes on the values, $j = |\alpha|, |\alpha| + 1, |\alpha| + 2, \cdots$. For a fixed value of j, the variable m takes on the values $m = -j, -j + 1, -j + 2, \cdots, j - 2, j - 1, j$. The inner product between two functions in the space $\psi(E, j, m, \alpha)$ and $\varphi(E, j, m, \alpha)$ is given by

$$(\psi, \varphi) = \int \frac{dE}{p} \sum_{j,m,\alpha} \psi^*(E, j, m, \alpha)\varphi(E, j, m, \alpha),$$

where p is the function of E defined by

$$p = (E^2 - \mu^2)^{\frac{1}{2}}.$$

The infinitesimal generators are the following:

$$J_3\varphi(E, j, m, \alpha) = m\varphi(E, j, m, \alpha); \tag{2.1}$$

$$\begin{aligned}&(J_2 + iJ_1)\varphi(E, j, m, \alpha)\\ &= [(j - m)(j + m + 1)]^{\frac{1}{2}}\varphi(E, j, m + 1, \alpha);\end{aligned} \tag{2.2}$$

$$\begin{aligned}&(J_2 - iJ_1)\varphi(E, j, m, \alpha)\\ &= [(j + m)(j - m + 1)]^{\frac{1}{2}}\varphi(E, j, m - 1, \alpha);\end{aligned} \tag{2.3}$$

$$H\varphi(E, j, m, \alpha) = E\varphi(E, j, m, \alpha); \tag{2.4}$$

$$\begin{aligned}P_3\varphi(E, j, m, \alpha) = p\Bigg[&\frac{m\alpha}{j(j+1)}\,\varphi(E, j, m, \alpha) + \frac{1}{(j+1)}\\ &\times\left[\frac{(j-m+1)(j+m+1)(j-\alpha+1)(j+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\\ &\times\varphi(E, j+1, m, \alpha)\\ &+\frac{1}{j}\left[\frac{(j-m)(j+m)(j-\alpha)(j+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\\ &\times\varphi(E, j-1, m, \alpha)\Bigg];\end{aligned} \tag{2.5}$$

$$\begin{aligned}&(P_2 + iP_1)\varphi(E, j, m, \alpha)\\ &= p\Bigg[\frac{\alpha}{j(j+1)}\,[(j-m)(j+m+1)]^{\frac{1}{2}}\\ &\times\varphi(E, j, m+1, \alpha) - \frac{1}{(j+1)}\\ &\times\left[\frac{(j+m+1)(j+m+2)(j-\alpha+1)(j+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\\ &\times\varphi(E, j+1, m+1, \alpha)\\ &+\frac{1}{j}\left[\frac{(j-m-1)(j-m)(j-\alpha)(j+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\\ &\times\varphi(E, j-1, m+1, \alpha)\Bigg];\end{aligned} \tag{2.6}$$

$$(P_2 - iP_1)\varphi(E, j, m, \alpha)$$

$$= p\Big\{\frac{\alpha}{j(j+1)}\,[(j+m)(j-m+1)]^{\frac{1}{2}} \times \varphi(E, j, m-1, \alpha) + \frac{1}{j+1}\left[\frac{(j-m+1)(j-m+2)(j-\alpha+1)(j+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}} \times \varphi(E, j+1, m-1, \alpha) - \frac{1}{j}\left[\frac{(j+m-1)(j+m)(j-\alpha)(j+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}} \times \varphi(E, j-1, m-1, \alpha)\Big\}; \tag{2.7}$$

$$\mathcal{J}_3\varphi(E, j, m, \alpha) = \frac{im}{j(j+1)}\Big\{\alpha p\,\frac{\partial}{\partial E}\,\varphi(E, j, m, \alpha) + \frac{\mu}{2p}\,[(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}} \times \varphi(E, j, m, \alpha-1) - \frac{\mu}{2p}\,[(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m, \alpha+1)\Big\}$$

$$+ \frac{i}{(j+1)}\left[\frac{(j-m+1)(j+m+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha+1)(j+\alpha+1)]^{\frac{1}{2}}\Big((j+1)\frac{E}{p} + p\,\frac{\partial}{\partial E}\Big)\varphi(E, j+1, m, \alpha)$$

$$+ \frac{\mu}{2p}\,[(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j+1, m, \alpha-1)$$

$$+ \frac{\mu}{2p}\,[(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j+1, m, \alpha+1)\Big\}$$

$$- \frac{i}{j}\left[\frac{(j-m)(j+m)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha)(j+\alpha)]^{\frac{1}{2}}\Big(j\frac{E}{p} - p\,\frac{\partial}{\partial E}\Big)\varphi(E, j-1, m, \alpha)$$

$$+ \frac{\mu}{2p}\,[(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j-1, m, \alpha-1)$$

$$+ \frac{\mu}{2p}\,[(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j-1, m, \alpha+1)\Big\}; \tag{2.8}$$

$$(\mathcal{J}_2 + i\mathcal{J}_1)\varphi(E, j, m, \alpha) = \frac{i}{j(j+1)}\,[(j-m)(j+m+1)]^{\frac{1}{2}}\Big\{\alpha p\,\frac{\partial}{\partial E}\,\varphi(E, j, m+1, \alpha)$$

$$+ \frac{\mu}{2p}\,[(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j, m+1, \alpha-1)$$

$$- \frac{\mu}{2p}\,[(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m+1, \alpha+1)\Big\}$$

$$- \frac{i}{(j+1)}\left[\frac{(j+m+1)(j+m+2)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha+1)(j+\alpha+1)]^{\frac{1}{2}}$$

$$\times \Big((j+1)\frac{E}{p} + p\,\frac{\partial}{\partial E}\Big)\varphi(E, j+1, m+1, \alpha)$$

$$+ \frac{\mu}{2p}\,[(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j+1, m+1, \alpha-1)$$

$$+ \frac{\mu}{2p}\,[(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j+1, m+1, \alpha+1)\Big\}$$

$$- \frac{i}{j}\left[\frac{(j-m-1)(j-m)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha)(j+\alpha)]^{\frac{1}{2}}\Big(j\frac{E}{p} - p\,\frac{\partial}{\partial E}\Big)\varphi(E, j-1, m+1, \alpha)$$

$$+ \frac{\mu}{2p}\,[(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j-1, m+1, \alpha-1)$$

$$+ \frac{\mu}{2p}\,[(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j-1, m+1, \alpha+1)\Big\}. \tag{2.9}$$

$$(\mathcal{J}_2 - i\mathcal{J}_1)\varphi(E, j, m, \alpha) = \frac{i}{j(j+1)}[(j+m)(j-m+1)]^{\frac{1}{2}}\Big\{\alpha p \frac{\partial}{\partial E}\varphi(E, j, m-1, \alpha)$$

$$+ \frac{\mu}{2p}[(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j, m-1, \alpha-1)$$

$$- \frac{\mu}{2p}[(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m-1, \alpha+1)\Big\}$$

$$+ \frac{i}{(j+1)}\left[\frac{(j-m+1)(j-m+2)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha+1)(j+\alpha+1)]^{\frac{1}{2}}\Big((j+1)\frac{E}{p} + p\frac{\partial}{\partial E}\Big)$$

$$\times \varphi(E, j+1, m-1, \alpha) + \frac{\mu}{2p}[(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j+1, m-1, \alpha-1)$$

$$+ \frac{\mu}{2p}[(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j+1, m-1, \alpha+1)\Big\}$$

$$+ \frac{i}{j}\left[\frac{(j+m-1)(j+m)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\Big\{[(j-\alpha)(j+\alpha)]^{\frac{1}{2}}\Big(j\frac{E}{p} - p\frac{\partial}{\partial E}\Big)\varphi(E, j-1, m-1, \alpha)$$

$$+ \frac{\mu}{2p}[(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j-1, m-1, \alpha-1)$$

$$+ \frac{\mu}{2p}[(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j-1, m-1, \alpha+1)\Big\}. \tag{2.10}$$

In the above and subsequent equations, terms which have a factor j in the denominator are to be replaced by zero when $j = 0$.

Using the above forms for the infinitesimal generators, one sees that

$$\mathbf{J}^2\varphi(E, j, m, \alpha) = j(j+1)\varphi(E, j, m, \alpha), \tag{2.11}$$

$$\mathbf{P}\cdot\mathbf{J}\varphi(E, j, m, \alpha) = p\alpha\varphi(E, j, m, \alpha). \tag{2.12}$$

We can also give the operators which correspond to the spin operators of the particle. These operators are denoted by S_i $(i = 1, 2, 3)$, and are the same operators which appear in Eq. (4.2) of Ref. 6. The operators $J_i - S_i$ are then the orbital angular momentum operators. It is convenient to introduce the operator P defined by

$$P\varphi(E, j, m, \alpha) = (E^2 - \mu^2)^{\frac{1}{2}}\varphi(E, j, m, \alpha).$$

Then the operators S_i are given by

$$S_i = T_i + P^{-2}(\mathbf{P}\cdot\mathbf{J})P_i, \tag{2.13}$$

where the operators T_i act in the given basis in the following way:

$$T_3\varphi(E, j, m, \alpha) = -\frac{m}{2j(j+1)}[(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j, m, \alpha-1)$$

$$- \frac{1}{2(j+1)}\left[\frac{(j-m+1)(j+m+1)(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m, \alpha-1)$$

$$+ \frac{1}{2j}\left[\frac{(j-m)(j+m)(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m, \alpha-1)$$

$$- \frac{m}{2j(j+1)}[(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m, \alpha+1)$$

$$+ \frac{1}{2(j+1)}\left[\frac{(j-m+1)(j+m+1)(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m, \alpha+1)$$

$$- \frac{1}{2j}\left[\frac{(j-m)(j+m)(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m, \alpha+1); \tag{2.14}$$

$$
\begin{aligned}
&(T_2 + iT_1)\varphi(E, j, m, \alpha) \\
&= -\frac{1}{2j(j+1)}[(j-m)(j+m+1)(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j, m+1, \alpha-1) \\
&+\frac{1}{2(j+1)}\left[\frac{(j+m+1)(j+m+2)(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m+1, \alpha-1) \\
&+\frac{1}{2j}\left[\frac{(j-m-1)(j-m)(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m+1, \alpha-1) \\
&-\frac{1}{2j(j+1)}[(j-m)(j+m+1)(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m+1, \alpha+1) \\
&-\frac{1}{2(j+1)}\left[\frac{(j+m+1)(j+m+2)(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m+1, \alpha+1) \\
&-\frac{1}{2j}\left[\frac{(j-m-1)(j-m)(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m+1, \alpha+1); \qquad (2.15)
\end{aligned}
$$

$$
\begin{aligned}
&(T_2 - iT_1)\varphi(E, j, m, \alpha) \\
&= -\frac{1}{2j(j+1)}[(j-m+1)(j+m)(j-\alpha+1)(j+\alpha)(s-\alpha+1)(s+\alpha)]^{\frac{1}{2}}\varphi(E, j, m-1, \alpha-1) \\
&-\frac{1}{2(j+1)}\left[\frac{(j-m+1)(j-m+2)(j-\alpha+1)(j-\alpha+2)(s-\alpha+1)(s+\alpha)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m-1, \alpha-1) \\
&-\frac{1}{2j}\left[\frac{(j+m-1)(j+m)(j+\alpha-1)(j+\alpha)(s-\alpha+1)(s+\alpha)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m-1, \alpha-1) \\
&-\frac{1}{2j(j+1)}[(j-m+1)(j+m)(j-\alpha)(j+\alpha+1)(s-\alpha)(s+\alpha+1)]^{\frac{1}{2}}\varphi(E, j, m-1, \alpha+1) \\
&+\frac{1}{2(j+1)}\left[\frac{(j-m+1)(j-m+2)(j+\alpha+1)(j+\alpha+2)(s-\alpha)(s+\alpha+1)}{(2j+1)(2j+3)}\right]^{\frac{1}{2}}\varphi(E, j+1, m-1, \alpha+1) \\
&+\frac{1}{2j}\left[\frac{(j+m-1)(j+m)(j-\alpha-1)(j-\alpha)(s-\alpha)(s+\alpha+1)}{(2j-1)(2j+1)}\right]^{\frac{1}{2}}\varphi(E, j-1, m-1, \alpha+1). \qquad (2.16)
\end{aligned}
$$

3. REPRESENTATIONS WITH VANISHING MASS AND DISCRETE SPIN

The representations with vanishing mass and discrete spin are closely related to the previous representations. However, they are considerably simpler in terms of the angular momentum basis than the previous representations. In the present case, the variable α can take on only *one* value: either $\alpha = s$ or $\alpha = -s$, corresponding to one of two values of circular polarization. The space then consists of a space of functions $\varphi(E, j, m)$ such that $j = |\alpha|, |\alpha| + 1, |\alpha| + 2, \cdots$ and $m = -j, -j+1, \cdots, j$. The inner product of two functions is the same as before but with $\mu = 0$ and no summation over α. The infinitesimal generators are given by Eqs. (2.1)–(2.10), where $\varphi(E, j, m, \alpha)$ is replaced by $\varphi(E, j, m)$. The functions $\varphi(E, j, m, \alpha \pm 1)$ do not appear. Equations (2.11) and (2.12) continue to hold with the indicated substitutions for φ.