JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Canonical Transformations and Spectra of Quantum Operators

BORIS LEAF

Department of Physics, State University of New York at Cortland, Cortland, New York,
and
Department of Physics, State University of New York at Binghamton, Binghamton, New York

(Received 13 March 1969)

The claim that unitary transformations in quantum mechanics correspond to the canonical transformations of classical mechanics is not correct. The spectra of operators produced by unitary transformation of Cartesian coordinate and momentum operators $(\mathbf{q}, \mathbf{k})$ are necessarily continuous over the entire real domain of their eigenvalues. Operators with spectra which are not everywhere continuous are generated from $(\mathbf{q}, \mathbf{k})$ by one-sided unitary transformations U for which $U^\dagger U = 1$ but for which $UU^\dagger$ commutes with either $\mathbf{q}$ or $\mathbf{k}$ (but not both). If $UU^\dagger$ commutes with $\mathbf{k}$, the new coordinates and momenta $(\mathbf{r}, \mathbf{s})$ satisfy commutation relations $[s_m, r_n] = 2\pi i \mathbf{1}\delta_{m,n}$, $[s_m, s_n] = 0$, but $[r_m, r_n] \neq 0$; $(\mathbf{r}, \mathbf{s})$ are canonical only for one-dimensional systems. The properties of one-sided unitary transformations are described; they are characterized by $\phi(\mathbf{K})$, the eigenvalue of $UU^\dagger$. The one-dimensional case for which the one-sided unitary transformation is canonical is discussed in detail. A prescription is given for obtaining the operator canonically conjugate to any one-dimensional observable. Generalization to higher dimensions is also discussed.

1. INTRODUCTION

It is generally accepted that the quantum-mechanical transformations corresponding to the canonical transformations of classical mechanics are unitary transformations.[1] For each canonical transformation from classical Cartesian coordinate and momentum variables $(\mathbf{Q}, \mathbf{K})$ to generalized coordinates and momenta $(\mathbf{R}, \mathbf{S})$, there should correspond a unitary transformation from the corresponding quantum operators $(\mathbf{q}, \mathbf{k})$ to $(\mathbf{r}, \mathbf{s})$. But it can be shown that the only operators $(\mathbf{r}, \mathbf{s})$ generated from $(\mathbf{q}, \mathbf{k})$ by unitary transformation are those with continuous spectra whose eigenvectors are normalizable to a delta function over the entire real domain of the eigenvalues. Since the quantum operators corresponding to classical, canonical variables often have discrete spectra (e.g., in the case of classical angle and action variables), an anomaly exists. The spectra of the operators $(\mathbf{r}, \mathbf{s})$ generated by unitary transformation are considered in Sec. 2.

The transformations of $(\mathbf{q}, \mathbf{k})$, which produce operators $(\mathbf{r}, \mathbf{s})$ for which $\mathbf{s}$ (or $\mathbf{r}$) has a spectrum that is not continuous over the entire real domain of its eigenvalues, are generated by operators U which are not unitary, but only one-sided unitary. If $U^\dagger U = 1$, then $UU^\dagger \neq 1$; but if $UU^\dagger$ commutes with $\mathbf{k}$, the spectrum of $\mathbf{s}$ is not everywhere continuous. The components of $\mathbf{r}$ and $\mathbf{s}$ satisfy the commutation rules for canonical operators, except that the components of $\mathbf{r}$ do not commute with each other. Accordingly, $\mathbf{r}$ and $\mathbf{s}$ are canonically conjugate only for one-dimensional systems.

The properties of one-sided unitary transformations are described in Sec. 3. If $UU^\dagger$ commutes with $\mathbf{k}$, then these properties are determined by a factor $\phi(\mathbf{K})$, the eigenvalue of $UU^\dagger$.

In Sec. 4, the important one-dimensional case is described for which the operators (r, s) produced by one-sided unitary transformation are canonical. The case of a discrete spectrum is discussed first. A method is given for constructing from a discrete basis $\{|\mu\rangle\}$ a basis $\{|S\rangle\}$ which is defined for every real value of S. The kets $|\mu\rangle$ and $|S\rangle$ are both eigenvectors of s. The method takes into account the presence of squared delta functions which arise in the case of discrete spectra. The factor $\phi(K)$ is singular in this case and it includes the arbitrary constant introduced by the presence of the squares of delta functions.[2,3] It is shown that the eigenvalues S which differ from the points μ of the discrete spectrum of s are not measurable. The general one-dimensional case in which the μ basis has continuous ranges as well as discrete points is also discussed in Sec. 4. A prescription is given for obtaining canonically conjugate operators for one-dimensional systems starting from any given basis of eigenvectors.

The generalization to higher dimensions is discussed in Sec. 5. As in the one-dimensional case, operators whose spectra are not everywhere continuous are generated only by one-sided unitary transformations.

2. UNITARY TRANSFORMATIONS

Consider the Hilbert space in which the eigenvectors of the coordinate operator $\mathbf{q}$ are complete.

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1947), 3rd ed., p. 106.

[2] W. Güttinger, Progr. Theoret. Phys. (Kyoto) **13**, 613 (1955).

[3] J. R. MacDonald and M. K. Bruchman, Rev. Mod. Phys. **28**, 393 (1956).

The vector $\mathbf{q}$ denotes the complete commuting set[4] of one-dimensional position operators for each degree of freedom of the system $q_1, q_2, \cdots, q_f$. We write the eigenvalue equation and completeness condition

$$\mathbf{q}\,|\mathbf{Q}\rangle = \mathbf{Q}\,|\mathbf{Q}\rangle, \quad \int d\mathbf{Q}\,|\mathbf{Q}\rangle\langle\mathbf{Q}| = \mathbf{1}, \tag{2.1}$$

where $\mathbf{Q}$ denotes the set of eigenvalues $\{Q_1, Q_2, \cdots, Q_f\}$ and $|\mathbf{Q}\rangle$ is the simultaneous eigenvector of the set of operators $\mathbf{q}$ belonging to the eigenvalue $\mathbf{Q}$. The Hilbert space is the tensor product of the f one-dimensional spaces spanned by the eigenvectors of the components of $\mathbf{q}$.[5] The vector $\mathbf{Q}$ locates a point in an f-dimensional configuration space. Each component Q_i ranges over the entire real domain. Similarly, the momentum operator $\mathbf{p} = 2\pi\hbar\mathbf{k}$ is a complete commuting set of component operators $p_1, p_2, \cdots, p_f$, for which

$$\mathbf{k}\,|\mathbf{K}\rangle = \mathbf{K}\,|\mathbf{K}\rangle, \quad \int d\mathbf{K}\,|\mathbf{K}\rangle\langle\mathbf{K}| = \mathbf{1}. \tag{2.2}$$

The components of the eigenvalue $\mathbf{K} = \{K_1, K_2, \cdots, K_f\}$ range over the entire real domain. $\{|\mathbf{Q}\rangle\}$ and $\{|\mathbf{K}\rangle\}$ are orthonormal sets with delta-function normalization:

$$\langle\mathbf{Q}'\,|\,\mathbf{Q}\rangle = \delta(\mathbf{Q}' - \mathbf{Q}), \quad \langle\mathbf{K}'\,|\,\mathbf{K}\rangle = \delta(\mathbf{K}' - \mathbf{K}). \tag{2.3}$$

The commutation relations for canonical operators hold:

$$\begin{aligned}[q_m, q_n] &= [k_m, k_n] = 0,\\ [k_m, q_n] &= (2\pi i)^{-1}\delta_{m,n}\mathbf{1}.\end{aligned} \tag{2.4}$$

Accordingly,

$$\langle\mathbf{Q}\,|\,\mathbf{K}\rangle = \exp(2\pi i\mathbf{Q}\cdot\mathbf{K}), \tag{2.5}$$

so that $|\mathbf{K}\rangle$ is the Fourier transform of $|\mathbf{Q}\rangle$:

$$|\mathbf{K}\rangle = \int d\mathbf{Q}\exp(2\pi i\mathbf{Q}\cdot\mathbf{K})\,|\mathbf{Q}\rangle. \tag{2.6}$$

Consider the transformation from $\mathbf{q}$, $\mathbf{k}$ to new operators $\mathbf{r}$, $\mathbf{s}$ generated by the operator U:

$$\mathbf{r} = U^\dagger\mathbf{q}U, \quad \mathbf{s} = U^\dagger\mathbf{k}U. \tag{2.7}$$

If U is unitary,

$$U^\dagger U = UU^\dagger = \mathbf{1}, \tag{2.8}$$

then $\mathbf{r}$ and $\mathbf{s}$ are Hermitian, and they are canonical operators since

$$\begin{aligned}[r_m, r_n] &= [s_m, s_n] = 0,\\ [s_m, r_n] &= (2\pi i)^{-1}\delta_{m,n}\mathbf{1}.\end{aligned} \tag{2.9}$$

It is generally accepted that unitary transformations are the quantum-mechanical transformations corresponding to the canonical transformations of classical mechanics.[1] But it is easy to show that the operators $\mathbf{r}$ and $\mathbf{s}$ produced by unitary transformation of $\mathbf{q}$, $\mathbf{k}$ necessarily have the same spectra as $\mathbf{q}$, $\mathbf{k}$; their eigenvectors are orthonormal sets with delta-function normalization for all values of their eigenvalues in the real domain. According to (2.1) and (2.7),

$$\mathbf{r}U^\dagger\,|\mathbf{Q}\rangle = U^\dagger\mathbf{q}\,|\mathbf{Q}\rangle = \mathbf{Q}U^\dagger\,|\mathbf{Q}\rangle, \tag{2.10}$$

so that $U^\dagger\,|\mathbf{Q}\rangle$ is an eigenket of $\mathbf{r}$ belonging to $\mathbf{Q}$. Define the ket

$$|\mathbf{R}\rangle = U^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}. \tag{2.11}$$

$|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}$ is the eigenket of $\mathbf{q}$ belonging to the eigenvalue $\mathbf{Q} = \mathbf{R}$. Then

$$\mathbf{r}\,|\mathbf{R}\rangle = U^\dagger\mathbf{q}\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}} = \mathbf{R}U^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}} = \mathbf{R}\,|\mathbf{R}\rangle, \tag{2.12}$$

so that $|\mathbf{R}\rangle$ is the eigenket of $\mathbf{r}$ belonging to eigenvalue $\mathbf{R}$. Similarly,

$$|\mathbf{S}\rangle = U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}}, \quad \mathbf{s}\,|\mathbf{S}\rangle = \mathbf{S}\,|\mathbf{S}\rangle. \tag{2.13}$$

From (2.11) and (2.13),

$$U^\dagger = \int d\mathbf{R}\,|\mathbf{R}\rangle\langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}} = \int d\mathbf{S}\,|\mathbf{S}\rangle\langle\mathbf{K}|_{\mathbf{K}=\mathbf{S}}. \tag{2.14}$$

The sets $\{|\mathbf{R}\rangle\}$ and $\{|\mathbf{S}\rangle\}$ are complete, since

$$\begin{aligned}\int d\mathbf{R}\,|\mathbf{R}\rangle\langle\mathbf{R}| &= U^\dagger\int d\mathbf{R}\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}\langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}}\,U\\ &= U^\dagger\int d\mathbf{Q}\,|\mathbf{Q}\rangle\langle\mathbf{Q}|\,U = \mathbf{1}.\end{aligned} \tag{2.15}$$

They are also orthonormal sets with delta-function normalization, i.e.,

$$\begin{aligned}\langle\mathbf{R}'\,|\,\mathbf{R}\rangle &= \langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}'}\,UU^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}\\ &= \langle\mathbf{Q}\,|_{\mathbf{Q}=\mathbf{R}'}\,\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}} = \delta(\mathbf{R}' - \mathbf{R}).\end{aligned} \tag{2.16}$$

Similarly,

$$\int d\mathbf{S}\,|\mathbf{S}\rangle\langle\mathbf{S}| = \mathbf{1}, \quad \langle\mathbf{S}'\,|\,\mathbf{S}\rangle = \delta(\mathbf{S}' - \mathbf{S}). \tag{2.17}$$

From (2.12), (2.13), and the completeness of $\{|\mathbf{R}\rangle\}$ and $\{|\mathbf{S}\rangle\}$, we have

$$\mathbf{r} = \int d\mathbf{R}\,|\mathbf{R}\rangle\,\mathbf{R}\,\langle\mathbf{R}|, \quad \mathbf{s} = \int d\mathbf{S}\,|\mathbf{S}\rangle\,\mathbf{S}\,\langle\mathbf{S}|. \tag{2.18}$$

The Fourier transform relation (2.5) is invariant to the unitary transformation,

$$\langle\mathbf{R}\,|\,\mathbf{S}\rangle = \langle\mathbf{Q}\,|_{\mathbf{Q}=\mathbf{R}}\,\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = \exp(2\pi i\mathbf{R}\cdot\mathbf{S}), \tag{2.19}$$

so that

$$|\mathbf{S}\rangle = \int d\mathbf{R}\exp(2\pi i\mathbf{R}\cdot\mathbf{S})\,|\mathbf{R}\rangle. \tag{2.20}$$

[4] See Ref. 1, p. 57.

[5] A. Messiah, *Quantum Mechanics* (John Wiley & Sons, Inc., New York, 1961), Vol. 1, p. 307.

Since the operators $\mathbf{r}$ and $\mathbf{s}$, resulting from the unitary transformation (2.7), have continuous spectra over the entire real domain with delta-function normalization of their eigenvectors, unitary transformations cannot generate canonical operators with discrete or noncontinuous spectra. It will be shown that operators with noncontinuous spectra are generated by one-sided unitary transformations.

As an example of a unitary transformation to illustrate the preceding formalism, consider the case

$$U^\dagger = \int d\mathbf{Q} \int d\mathbf{Q}' \exp(2\pi i \mathbf{Q}' \cdot \mathbf{Q}) \, |\mathbf{Q}'\rangle\langle\mathbf{Q}|, \quad (2.21)$$

where (2.8) is satisfied. Define

$$|\mathbf{K}\rangle = U^\dagger \, |\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{K}} = \int d\mathbf{Q} \exp(2\pi i \mathbf{Q} \cdot \mathbf{K}) \, |\mathbf{Q}\rangle.$$

Then

$$\begin{aligned} U^\dagger \mathbf{q} U &= U^\dagger \int d\mathbf{Q} \, |\mathbf{Q}\rangle \, \mathbf{Q} \, \langle\mathbf{Q}| \, U \\ &= \int d\mathbf{K} U^\dagger \, |\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{K}} \, \mathbf{K} \, \langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{K}} \, U \\ &= \int d\mathbf{K} \, |\mathbf{K}\rangle \, \mathbf{K} \, \langle\mathbf{K}| = \mathbf{k}. \end{aligned}$$

Since

$$U^\dagger U^\dagger = \int d\mathbf{Q} \, |\mathbf{Q}\rangle\langle -\mathbf{Q}|, \quad UU = \int d\mathbf{Q} \, |-\mathbf{Q}\rangle\langle\mathbf{Q}|,$$

we have, therefore,

$$\begin{aligned} U^\dagger \mathbf{k} U = U^\dagger U^\dagger \mathbf{q} UU &= U^\dagger U^\dagger \int d\mathbf{Q} \, |\mathbf{Q}\rangle \, \mathbf{Q} \, \langle\mathbf{Q}| \, UU \\ &= \int d\mathbf{Q} \, |-\mathbf{Q}\rangle \, \mathbf{Q} \, \langle -\mathbf{Q}| = -\mathbf{q}. \end{aligned}$$

This unitary transformation corresponds to the classical canonical transformation which interchanges coordinate and momentum variables.[6] The correspondence between infinitesimal unitary transformations for which $U = \mathbf{1} + i\epsilon F$, where ϵ is an infinitesimal number and F is a Hermitian operator, and infinitesimal classical canonical transformations is well known.[1]

3. ONE-SIDED UNITARY TRANSFORMATION

Consider the operator U for which

$$U^\dagger U = \mathbf{1}, \quad \text{but} \quad UU^\dagger \neq \mathbf{1}. \quad (3.1)$$

Let us see under what conditions the canonical commutation relations (2.9) can be obtained for the Hermitian operators $\mathbf{r}$, $\mathbf{s}$ defined by (2.7). From (2.4) and (3.1),

$$\begin{aligned} U^\dagger [k_m, q_n] U &= U^\dagger U U^\dagger k_m q_n U - U^\dagger q_n k_m U U^\dagger U \\ &= (2\pi i)^{-1} \delta_{m,n} \mathbf{1}. \end{aligned}$$

If we require that $UU^\dagger$ commute with $\mathbf{k}$ (but not with $\mathbf{q}$),

$$[UU^\dagger, \mathbf{k}] = 0, \quad (3.2)$$

then $[s_m, r_n] = (2\pi i)^{-1} \delta_{m,n} \mathbf{1}$. It also follows from (3.2) that $[s_m, s_n] = 0$. But in general $[r_m, r_n] \neq 0$, since

$$r_m r_n = U^\dagger q_m U U^\dagger q_n U \quad (3.3)$$

and $UU^\dagger$ does not commute with $\mathbf{q}$. For a one-dimensional system, the one-sided unitary transformation satisfying (3.2) is canonical; for higher dimensions, the components of the new coordinate operator r do not commute with each other. If $UU^\dagger$ commutes with $\mathbf{q}$ (but not with $\mathbf{k}$), it is the components of the new momentum operator $\mathbf{s}$ which do not commute; *mutatis mutandis*, discussion of this case would present nothing new. Note that if $UU^\dagger$ commutes with both $\mathbf{q}$ and $\mathbf{k}$, then $UU^\dagger = \mathbf{1}$; this is the unitary case already considered. Since it will be shown that the transformations of $(\mathbf{q}, \mathbf{k})$ which generate operators $(\mathbf{r}, \mathbf{s})$, where $\mathbf{s}$ (or $\mathbf{r}$) has a noncontinuous spectrum, are one-sided unitary transformations in which $UU^\dagger$ commutes with $\mathbf{k}$ (or $\mathbf{q}$), their properties are now considered.

Define the kets $|\mathbf{R}\rangle$ and $|\mathbf{S}\rangle$ as in (2.11) and (2.13). Then, just as in (2.15) and (2.17), the sets are complete, according to (3.1). Also, as in (2.18),

$$\begin{aligned} \mathbf{r} &= U^\dagger \int d\mathbf{Q} \, |\mathbf{Q}\rangle \, \mathbf{Q} \, \langle\mathbf{Q}| \, U \\ &= U^\dagger \int d\mathbf{R} \, |\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}} \, \mathbf{R} \, \langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}} \, U \\ &= \int d\mathbf{R} \, |\mathbf{R}\rangle \, \mathbf{R} \, \langle\mathbf{R}|, \\ \mathbf{s} &= \int d\mathbf{S} \, |\mathbf{S}\rangle \, \mathbf{S} \, \langle\mathbf{S}|. \end{aligned} \quad (3.4)$$

Since $UU^\dagger$ commutes with $\mathbf{k}$, it can be written in diagonal form as

$$UU^\dagger = \int d\mathbf{K} \, |\mathbf{K}\rangle \, \phi(\mathbf{K}) \, \langle\mathbf{K}|, \quad \phi(\mathbf{K}) = \mathrm{Tr}\,[UU^\dagger \, |\mathbf{K}\rangle\langle\mathbf{K}|], \quad (3.5)$$

where $\phi(\mathbf{K})$ is the eigenvalue of $UU^\dagger$ to which the ket $|\mathbf{K}\rangle$ belongs:

$$UU^\dagger \, |\mathbf{K}\rangle = \phi(\mathbf{K}) \, |\mathbf{K}\rangle. \quad (3.6)$$

[6] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1950), p. 245.

Since $UU^\dagger$ is Hermitian, $\phi(\mathbf{K})$ is real. If we define

$$\phi(\mathbf{S}) = [\phi(\mathbf{K})]_{\mathbf{K}=\mathbf{S}}, \tag{3.7}$$

then

$$\begin{aligned}\phi(\mathbf{S})\,|\mathbf{S}\rangle &= \phi(\mathbf{K})_{\mathbf{K}=\mathbf{S}}U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = U^\dagger[\phi(\mathbf{K})|\mathbf{K}]_{\mathbf{K}=\mathbf{S}}\\ &= U^\dagger UU^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = |\mathbf{S}\rangle.\end{aligned} \tag{3.8}$$

Also $|\mathbf{S}\rangle$ is an eigenket of $\mathbf{s}$ since, according to (3.2),

$$\begin{aligned}\mathbf{s}\,|\mathbf{S}\rangle &= U^\dagger\mathbf{k}U\cdot U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = U^\dagger\mathbf{k}\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}}\\ &= \mathbf{S}U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = \mathbf{S}\,|\mathbf{S}\rangle.\end{aligned} \tag{3.9}$$

The eigenvectors of the Hermitian operator $\mathbf{s}$ are orthogonal but without delta-function normalization, since

$$\langle\mathbf{S}'\mid\mathbf{S}\rangle = \langle\mathbf{K}|_{\mathbf{K}=\mathbf{S}'}\,UU^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = \phi(\mathbf{S})\delta(\mathbf{S}'-\mathbf{S}). \tag{3.10}$$

Note that if $\phi(\mathbf{S}) = 1$ for all $\mathbf{S}$, then $UU^\dagger = \mathbf{1}$. This is the unitary case with delta-function normalization excluded in (3.1). On the other hand, the kets $|\mathbf{R}\rangle$ are not eigenkets of the Hermitian operator $\mathbf{r}$, since

$$\mathbf{r}\,|\mathbf{R}\rangle = U^\dagger\mathbf{q}UU^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}, \tag{3.11}$$

and $UU^\dagger$ does not commute with $\mathbf{q}$. The kets $|\mathbf{R}\rangle$ are not orthogonal; instead, according to (2.5),

$$\begin{aligned}\langle\mathbf{R}'\mid\mathbf{R}\rangle &= \langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}'}\,UU^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}\\ &= \int d\mathbf{K}\phi(\mathbf{K})\,\langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}'}\,|\mathbf{K}\rangle\langle\mathbf{K}\mid\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}\\ &= \int d\mathbf{K}\phi(\mathbf{K})\exp[2\pi i\mathbf{K}\cdot(\mathbf{R}'-\mathbf{R})]\\ &= \int d\mathbf{S}\phi(\mathbf{S})\exp[2\pi i\mathbf{S}\cdot(\mathbf{R}'-\mathbf{R})].\end{aligned} \tag{3.12}$$

Orthogonality is destroyed unless $\phi(\mathbf{S}) = 1$ for all $\mathbf{S}$. From (3.4) and (3.12), we have

$$\begin{aligned}\mathbf{r}\,|\mathbf{R}\rangle &= \int d\mathbf{R}'\,|\mathbf{R}'\rangle\,\mathbf{R}'\langle\mathbf{R}'\mid\mathbf{R}\rangle\\ &= \int d\mathbf{R}'\mathbf{R}'\,|\mathbf{R}'\rangle\int d\mathbf{S}\phi(\mathbf{S})\exp[2\pi i\mathbf{S}\cdot(\mathbf{R}'-\mathbf{R})],\end{aligned} \tag{3.13}$$

$$\begin{aligned}|\mathbf{R}\rangle &= \int d\mathbf{R}'\,|\mathbf{R}'\rangle\langle\mathbf{R}'\mid\mathbf{R}\rangle\\ &= \int d\mathbf{R}'\,|\mathbf{R}'\rangle\int d\mathbf{S}\phi(\mathbf{S})\exp[2\pi i\mathbf{S}\cdot(\mathbf{R}'-\mathbf{R})].\end{aligned} \tag{3.14}$$

Since (3.12) can be written as

$$\langle\mathbf{R}'\mid\mathbf{R}\rangle = \phi\left(\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{R}}\right)\delta(\mathbf{R}'-\mathbf{R}), \tag{3.15}$$

it follows that

$$\mathbf{r}\,|\mathbf{R}\rangle = \phi\left(\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{R}}\right)\mathbf{R}\,|\mathbf{R}\rangle, \quad |\mathbf{R}\rangle = \phi\left(\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{R}}\right)|\mathbf{R}\rangle. \tag{3.16}$$

The Fourier transform relations are preserved. Thus,

$$\begin{aligned}|\mathbf{S}\rangle &= U^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = U^\dagger\int d\mathbf{Q}\,|\mathbf{Q}\rangle\langle\mathbf{Q}\mid\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}}\\ &= U^\dagger\int d\mathbf{Q}\,|\mathbf{Q}\rangle\exp(2\pi i\mathbf{Q}\cdot\mathbf{S})\\ &= \int d\mathbf{R}\exp(2\pi i\mathbf{R}\cdot\mathbf{S})U^\dagger\,|\mathbf{Q}\rangle_{\mathbf{Q}=\mathbf{R}}\\ &= \int d\mathbf{R}\exp(2\pi i\mathbf{R}\cdot\mathbf{S})\,|\mathbf{R}\rangle.\end{aligned} \tag{3.17}$$

Similarly,

$$|\mathbf{R}\rangle = \int d\mathbf{S}\exp(-2\pi i\mathbf{R}\cdot\mathbf{S})\,|\mathbf{S}\rangle. \tag{3.18}$$

But

$$\langle\mathbf{R}\mid\mathbf{S}\rangle = \langle\mathbf{Q}|_{\mathbf{Q}=\mathbf{R}}UU^\dagger\,|\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}} = \phi(\mathbf{S})\exp(2\pi i\mathbf{R}\cdot\mathbf{S}). \tag{3.19}$$

According to (3.9), any function $f(\mathbf{s})$ can be written in diagonal form

$$f(\mathbf{s}) = \int d\mathbf{S}\,|\mathbf{S}\rangle f(\mathbf{S})\,\langle\mathbf{S}|. \tag{3.20}$$

Since

$$\langle\mathbf{R}|\,f(\mathbf{s})\,|\mathbf{S}\rangle = f(\mathbf{S})\langle\mathbf{R}\mid\mathbf{S}\rangle = \phi(\mathbf{S})f(\mathbf{S})\exp(2\pi i\mathbf{R}\cdot\mathbf{S}),$$

according to (3.8) and (3.18), we have

$$\begin{aligned}f(\mathbf{s}) &= \int d\mathbf{R}\int d\mathbf{S}\,|\mathbf{R}\rangle\,\phi(\mathbf{S})f(\mathbf{S})\exp(2\pi i\mathbf{R}\cdot\mathbf{S})\,\langle\mathbf{S}|\\ &= \int d\mathbf{R}\,|\mathbf{R}\rangle f\left((2\pi i)^{-1}\frac{\partial}{\partial\mathbf{R}}\right)\int d\mathbf{S}\,\langle\mathbf{S}|\exp(2\pi i\mathbf{R}\cdot\mathbf{S})\\ &= \int d\mathbf{R}\,|\mathbf{R}\rangle f\left((2\pi i)^{-1}\frac{\partial}{\partial\mathbf{R}}\right)\langle\mathbf{R}|.\end{aligned} \tag{3.21}$$

Consider the unitary operator $\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{s})$ where $\boldsymbol{\beta}$ is a real constant:

$$\begin{aligned}\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{s}) &= \int d\mathbf{S}\,|\mathbf{S}\rangle\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{S})\,\langle\mathbf{S}|\\ &= \int d\mathbf{R}\,|\mathbf{R}\rangle\exp\left(\boldsymbol{\beta}\cdot\frac{\partial}{\partial\mathbf{R}}\right)\langle\mathbf{R}|\\ &= \int d\mathbf{R}\,|\mathbf{R}\rangle\langle\mathbf{R}+\boldsymbol{\beta}|.\end{aligned} \tag{3.22}$$

Accordingly, $\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{s})$ is a displacement operator for the set $\{|\mathbf{R}\rangle\}$:

$$\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{s})\,|\mathbf{R}\rangle = \int d\mathbf{S}\,|\mathbf{S}\rangle \exp[-2\pi i\mathbf{S}\cdot(\mathbf{R}-\boldsymbol{\beta})] = |\mathbf{R}-\boldsymbol{\beta}\rangle. \tag{3.23}$$

From (3.13), (3.17), and (3.19),

$$\mathbf{r}\,|\mathbf{R}\rangle = \int d\mathbf{S}\phi(\mathbf{S})\exp(-2\pi i\mathbf{R}\cdot\mathbf{S})(2\pi i)^{-1}\frac{\partial}{\partial\mathbf{S}} \times\int d\mathbf{R}'\exp(2\pi i\mathbf{R}'\cdot\mathbf{S})\,|\mathbf{R}'\rangle$$
$$= \int d\mathbf{S}\langle\mathbf{S}\mid\mathbf{R}\rangle(2\pi i)^{-1}\frac{\partial}{\partial\mathbf{S}}|\mathbf{S}\rangle,$$
$$\mathbf{r} = \int d\mathbf{S}\left((2\pi i)^{-1}\frac{\partial}{\partial\mathbf{S}}|\mathbf{S}\rangle\right)\langle\mathbf{S}| = \int d\mathbf{S}\,|\mathbf{S}\rangle\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{S}}\langle\mathbf{S}|. \tag{3.24}$$

As shown in (3.3), the components of $\mathbf{r}$ do not commute; from (3.24),

$$r_m r_n = \int d\mathbf{S}\,|\mathbf{S}\rangle\left(\frac{i}{2\pi}\frac{\partial}{\partial S_m}\right)\phi(\mathbf{S})\left(\frac{i}{2\pi}\frac{\partial}{\partial S_n}\right)\langle\mathbf{S}|, \tag{3.25}$$

so that commutivity occurs for the unitary case, $\phi(\mathbf{S}) = 1$ for all $\mathbf{S}$. For integral m, we can write

$$(2\pi i\boldsymbol{\beta}\cdot\mathbf{r})^m = \int d\mathbf{S}\,|\mathbf{S}\rangle\left(-\phi(\mathbf{S})\boldsymbol{\beta}\cdot\frac{\partial}{\partial\mathbf{S}}\right)\langle\mathbf{S}|,$$

so that the unitary operator $\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{r})$ becomes

$$\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{r}) = \int d\mathbf{S}\,|\mathbf{S}\rangle\exp\left(-\phi(\mathbf{S})\boldsymbol{\beta}\cdot\frac{\partial}{\partial\mathbf{S}}\right)\langle\mathbf{S}|, \tag{3.26}$$

which is not a displacement operator for the set $\{|\mathbf{S}\rangle\}$. But consider the class of operators g defined by

$$g = \int d\mathbf{R}\,|\mathbf{R}\rangle\, g(\mathbf{R})\,\langle\mathbf{R}|. \tag{3.27}$$

If $g(\mathbf{R})$ is real, then g is Hermitian. The operator $\mathbf{r}$ itself is a g operator according to (3.4), but the operator g should not be confused with $g(\mathbf{r})$. Now

$$g\,|\mathbf{S}\rangle = \int d\mathbf{R}\,|\mathbf{R}\rangle\, g(\mathbf{R})\phi(\mathbf{S})\exp(2\pi i\mathbf{R}\cdot\mathbf{S}) = \phi(\mathbf{S})g\left((2\pi i)^{-1}\frac{\partial}{\partial\mathbf{S}}\right)|\mathbf{S}\rangle,$$
$$g = \int d\mathbf{S}\left\{g\left((2\pi i)^{-1}\frac{\partial}{\partial\mathbf{S}}\right)|\mathbf{S}\rangle\right\}\langle\mathbf{S}| = \int d\mathbf{S}\,|\mathbf{S}\rangle\, g\left(\frac{i}{2\pi}\frac{\partial}{\partial\mathbf{S}}\right)\langle\mathbf{S}|. \tag{3.28}$$

A displacement operator can be obtained as the g operator:

$$v^\dagger(\boldsymbol{\beta}) = \int d\mathbf{R}\,|\mathbf{R}\rangle\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{R})\,\langle\mathbf{R}| = \int d\mathbf{S}\,|\mathbf{S}\rangle\exp\left(-\boldsymbol{\beta}\cdot\frac{\partial}{\partial\mathbf{S}}\right)\langle\mathbf{S}| = \int d\mathbf{S}\,|\mathbf{S}\rangle\langle\mathbf{S}-\boldsymbol{\beta}|, \tag{3.29}$$

$$v^\dagger(\boldsymbol{\beta})\,|\mathbf{S}\rangle = \int d\mathbf{S}'\,|\mathbf{S}'\rangle\langle\mathbf{S}'-\boldsymbol{\beta}\mid\mathbf{S}\rangle = \int d\mathbf{S}'\,|\mathbf{S}'\rangle\,\delta(\mathbf{S}'-\boldsymbol{\beta}-\mathbf{S})\phi(\mathbf{S}) = \phi(\mathbf{S})\,|\mathbf{S}+\boldsymbol{\beta}\rangle. \tag{3.30}$$

In general, $v(\boldsymbol{\beta})$ is not unitary unless $\phi(\mathbf{S}) = 1$ for all $\mathbf{S}$, since

$$v^\dagger(\boldsymbol{\beta})v(\boldsymbol{\beta}) = \int d\mathbf{S}\,|\mathbf{S}\rangle\,\phi(\mathbf{S}-\boldsymbol{\beta})\,\langle\mathbf{S}|,$$
$$v(\boldsymbol{\beta})v^\dagger(\boldsymbol{\beta}) = \int d\mathbf{S}\,|\mathbf{S}\rangle\,\phi(\mathbf{S}+\boldsymbol{\beta})\,\langle\mathbf{S}|. \tag{3.31}$$

Note from (3.23) and (3.30) that

$$\exp(2\pi i\boldsymbol{\beta}\cdot\mathbf{s})\mathbf{r}\exp(-2\pi i\boldsymbol{\beta}\cdot\mathbf{s}) = \int d\mathbf{R}\,|\mathbf{R}-\boldsymbol{\beta}\rangle\,\mathbf{R}\,\langle\mathbf{R}-\boldsymbol{\beta}| = \mathbf{r}+\boldsymbol{\beta}\mathbf{1}, \tag{3.32}$$

$$v(\boldsymbol{\beta})\mathbf{s}v^\dagger(\boldsymbol{\beta}) = \int d\mathbf{S}\,|\mathbf{S}-\boldsymbol{\beta}\rangle\,\phi(\mathbf{S})\mathbf{S}\,\langle\mathbf{S}-\boldsymbol{\beta}| = \int d\mathbf{S}\,|\mathbf{S}\rangle\,\phi(\mathbf{S}+\boldsymbol{\beta})(\mathbf{S}+\boldsymbol{\beta})\,\langle\mathbf{S}|. \tag{3.33}$$

The properties of the one-sided unitary transformations satisfying (3.2) are controlled by the ubiquitous factor $\phi(\mathbf{S})$. If $\phi(\mathbf{S})$ depends on Planck's constant in such a way that as $\hbar\to 0$, $\phi(\mathbf{S})\to 1$ for all $\mathbf{S}$, then the same classical limit would be obtained whether $\mathbf{r}$, $\mathbf{s}$ are generated from $\mathbf{q}$, $\mathbf{k}$ by unitary transformation or by one-sided unitary transformations satisfying (3.2).

4. THE ONE-DIMENSIONAL CASE

In Sec. 2 it was shown that unitary transformation of the coordinate and momentum operators cannot generate operators whose spectra are discrete or not continuous on the entire real domain. For a one-dimensional system, the one-sided unitary transformation satisfying (3.2) is canonical, as shown in Sec. 3. In this section we show that for one-dimensional systems, operators with discrete spectra or spectra which are not continuous over the entire real domain can be generated by such canonical, one-sided, unitary transformations.

In the one-dimensional space spanned by the eigenvectors of position operator q and momentum operator $p = 2\pi\hbar k$, consider the transformation from the basis of eigenkets $|K\rangle$ of k to a basis of kets $|\mu\rangle$ specified by a parameter μ which takes on discrete values on the real axis (discrete spectrum). The set $\{|\mu\rangle\}$ is assumed complete, and orthonormal with Kronecker-delta normalization, so that

$$\Sigma_\mu |\mu\rangle\langle\mu| = \mathbf{1}, \quad \langle\mu' | \mu\rangle = \delta_{\mu',\mu}. \tag{4.1}$$

Note that the μ basis is nondegenerate, since, by definition, for a one-dimensional system the set $\{|\mu\rangle\}$ must be a complete commuting set.[4] If, for example, an eigenvalue were doubly degenerate, two indices μ_1 and μ_2 would be required to specify a complete commuting set; the system would be two-dimensional.

The transformation operator U is specified by

$$U^\dagger |K\rangle = \epsilon^{\frac{1}{2}}\Sigma_\mu \delta(K - \mu)\,|\mu\rangle \tag{4.2}$$

or

$$U^\dagger = \epsilon^{\frac{1}{2}}\int dK \Sigma_\mu \delta(K - \mu)\,|\mu\rangle\langle K|. \tag{4.3}$$

The real constant ϵ will be chosen so that the condition

$$U^\dagger U = \mathbf{1} \tag{4.4}$$

is satisfied. Thus,

$$\begin{aligned} U^\dagger U &= \epsilon\int dK \Sigma_\mu \Sigma_{\mu'} \delta(K - \mu)\delta(K - \mu')\,|\mu\rangle\langle\mu'| \\ &= \epsilon\int dK \Sigma_\mu \Sigma_{\mu'} \delta_{\mu,\mu'}[\delta(K - \mu)]^2\,|\mu\rangle\langle\delta'|, \end{aligned} \tag{4.5}$$

since $\delta(K - \mu)\delta(K - \mu') = 0$ if $\mu' \neq \mu$. It has been shown by use of distribution theory[2,3] that

$$[\delta(K - \mu)]^2 = \epsilon'\delta(K - \mu), \tag{4.6}$$

where ϵ' is a *finite* constant. Choose $\epsilon' = \epsilon^{-1}$. Then, from (4.1),

$$U^\dagger U = \epsilon\epsilon'\int dK \Sigma_\mu \delta(K - \mu)\,|\mu\rangle\langle\mu| = \Sigma_\mu |\mu\rangle\langle\mu| = \mathbf{1}. \tag{4.7}$$

On the other hand,

$$\begin{aligned} UU^\dagger &= \epsilon\int dK \int dK' \Sigma_\mu \delta(K - \mu)\delta(K' - \mu)\,|K'\rangle\langle K| \\ &= \epsilon\int dK\,|K\rangle \Sigma_\mu \delta(K - \mu)\langle K| = \epsilon\Sigma_\mu \delta(k - \mu). \end{aligned} \tag{4.8}$$

Accordingly, the transformation from the K basis, where K is a continuous spectrum on the entire real axis, to the μ basis, where μ is a discrete spectrum of points on the real axis, is a canonical one-sided unitary transformation for which $UU^\dagger$ commutes with k but not with q. It is characterized by the factor $\phi(K)$ which, according to (3.5) and (4.8), is

$$\phi(K) = \epsilon\Sigma_\mu \delta(K - \mu). \tag{4.9}$$

Equations (4.5) and (4.7) are equivalent to

$$\begin{aligned} &\epsilon\int dK \delta(K - \mu)\delta(K - \mu') \\ &\qquad = \delta_{\mu,\mu'}\epsilon\int dK[\delta(K - \mu)]^2 \\ &\qquad = \delta_{\mu,\mu'}\int dK \delta(K - \mu) = \delta_{\mu,\mu'}, \end{aligned} \tag{4.10}$$

when μ and μ' assume discrete values. The validity of this equation permits interpretation of the undetermined constant ϵ as arising from the indefinite limit of integration in the following equation:

$$\begin{aligned} &\epsilon\delta(K - \mu)\delta(K - \mu') \\ &\qquad = \int dK \delta(K - \mu)\int_{\mu' - \frac{1}{2}\epsilon}^{\mu' + \frac{1}{2}\epsilon} dx \delta(K - x), \end{aligned} \tag{4.11}$$

where ϵ lies in the range $0 < \frac{1}{2}\epsilon < |\mu' - \mu|$, for $\mu' \neq \mu$. From (4.11),

$$\begin{aligned} &\epsilon\int dK \delta(K - \mu)\delta(K - \mu') \\ &\qquad = \int dK \delta(K - \mu)\int_{\mu' - \frac{1}{2}\epsilon}^{\mu' + \frac{1}{2}\epsilon} dx \delta(K - x) \\ &\qquad = \int_{\mu' - \frac{1}{2}\epsilon}^{\mu' + \frac{1}{2}\epsilon} dx \delta(x - \mu) = \delta_{\mu',\mu}, \end{aligned}$$

in agreement with (4.10), where ϵ is restricted by the smallest separation between points of the discrete spectrum. In the continuum, $\epsilon \to 0$ so that

$$\begin{aligned} &\epsilon\int dK \delta(K - \mu)\delta(K - \mu') \\ &\qquad = \int_{\mu' - \frac{1}{2}\epsilon}^{\mu' + \frac{1}{2}\epsilon} dx \delta(x - \mu) \to \delta(\mu' - \mu). \end{aligned}$$

In this limit, for all K,

$$\phi(K) = \epsilon\Sigma_\mu \delta(K - \mu) \to 1, \quad \text{so that} \quad UU^\dagger \to \mathbf{1}; \tag{4.12}$$

the transformation becomes unitary.

Using the one-sided unitary operator U defined by (4.3), define the kets $|S\rangle$ and $|R\rangle$, as in (2.11) and (2.13):

$$\begin{aligned} |S\rangle &= U^\dagger |K\rangle_{K=S} = \epsilon^{\frac{1}{2}}\Sigma_\mu \delta(S - \mu)\,|\mu\rangle, \\ |R\rangle &= U^\dagger |Q\rangle_{Q=R} = \epsilon^{\frac{1}{2}}\Sigma_\mu \exp(-2\pi i\mu R)\,|\mu\rangle. \end{aligned} \tag{4.13}$$

The sets $\{|S\rangle\}$ and $\{|R\rangle\}$ are complete in accordance with the discussion in Sec. 3. From (3.10) and (3.12),

$$\langle S' \mid S\rangle = \phi(S)\delta(S' - S) = \epsilon\Sigma_\mu\delta(S - \mu)\delta(S - S'),$$
$$\langle R' \mid R\rangle = \int dS\phi(S)\exp[2\pi iS(R' - R)]$$
$$= \epsilon\Sigma_\mu \exp[2\pi i\mu(R' - R)]. \quad (4.14)$$

It may be noted that, in agreement with (4.10),

$$\langle\mu' \mid \mu\rangle = \int dR\langle\mu' \mid R\rangle\langle R \mid \mu\rangle$$
$$= \epsilon\int dR \exp[2\pi iR(\mu - \mu')] = \delta_{\mu,\mu'}. \quad (4.15)$$

The Hermitian operators r and s become

$$r = \int dR\,|R\rangle\, R\,\langle R|$$
$$= \epsilon\Sigma_\mu\Sigma_{\mu'}\int dRR\exp[2\pi iR(\mu' - \mu)]\,|\mu\rangle\langle\mu'|, \quad (4.16)$$

$$s = \int dS\,|S\rangle\, S\,\langle S|$$
$$= \epsilon\Sigma_\mu\Sigma_{\mu'}\int dSS\delta(S - \mu)\delta(S - \mu')\,|\mu\rangle\langle\mu'|$$
$$= \Sigma_\mu\,|\mu\rangle\,\mu\,\langle\mu|. \quad (4.17)$$

The operator s has the discrete spectrum μ, but also has the spectrum S which is defined on the entire real axis:

$$s\,|\mu\rangle = \Sigma_{\mu'}\,|\mu'\rangle\,\mu'\langle\mu' \mid \mu\rangle = \mu\,|\mu\rangle, \quad (4.18)$$
$$s\,|S\rangle = \epsilon^{\frac{1}{2}}\Sigma_\mu\delta(S - \mu)s\,|\mu\rangle$$
$$= \epsilon^{\frac{1}{2}}\Sigma_\mu\delta(S - \mu)\mu\,|\mu\rangle$$
$$= \epsilon^{\frac{1}{2}}\Sigma_\mu\delta(S - \mu)S\,|\mu\rangle = S\,|S\rangle. \quad (4.19)$$

For every Hermitian operator s which has an orthonormal set of eigenkets $|\mu\rangle$ belonging to a discrete spectrum of eigenvalues μ, a complete set of eigenkets $|S\rangle$ can be constructed which is defined for every eigenvalue S on the real axis. However, the kets $|S\rangle$ are not normalized to a delta function but are normalized according to (4.14); the norm of $|S\rangle$ vanishes unless S is one of the discrete points of $\{\mu\}$; only when $\phi(S) = 1$ for all S is the normalization to a delta function for all S.

The operators r and s in (4.16) and (4.17) are canonically conjugate with $[s, r] = (2\pi i)^{-1}\mathbf{1}$. The kets $|R\rangle$ and $|S\rangle$ are Fourier transforms in accordance with (3.17) for the one-dimensional case. From (3.22) and (3.26) the unitary operator $\exp(2\pi i\beta s)$ is a displacement operator, but $\exp(2\pi i\beta r)$ is not. The displacement operator for the set $\{|S\rangle\}$ is $v^\dagger(\beta)$, which is defined in (3.29); (3.30) becomes

$$v^\dagger(\beta)\,|S\rangle = \phi(S)\,|S + \beta\rangle = \epsilon\Sigma_\mu\delta(S - \mu)\,|S + \beta\rangle$$
$$= \epsilon^{\frac{3}{2}}\Sigma_\mu\Sigma_{\mu'}\delta(S - \mu)\delta(S + \beta - \mu')\,|\mu'\rangle$$
$$= \epsilon^{\frac{1}{2}}\Sigma_\mu\Sigma_{\mu'}\delta_{\mu',\mu+\beta}\delta(S + \beta - \mu')\,|\mu'\rangle. \quad (4.20)$$

The displacement operator $v^\dagger(\beta)$ can also be applied to a ket $|\mu\rangle$ of the discrete spectrum:

$$v^\dagger(\beta)\,|\mu\rangle = \int dSv^\dagger(\beta)\,|S\rangle\langle S \mid \mu\rangle = \Sigma_{\mu'}\delta_{\mu',\mu+\beta}\,|\mu'\rangle, \quad (4.21)$$

which equals $|\mu + \beta\rangle$ if $\mu + \beta$ belongs to the set $\{\mu\}$ and vanishes otherwise.

The question arises whether the eigenvalues of s in the spectrum S are measurable. The answer is negative unless S is one of the points in the discrete set μ. Let $\rho(s)$ be a density operator which commutes with s so that it represents a mixture of eigenstates of s. Since $\rho(s)$ is diagonal both in the S representation and in the μ representation,

$$\rho(s) = \int dS\,|S\rangle\,\rho(S)\,\langle S| = \Sigma_\mu\,|\mu\rangle\,\rho(\mu)\,\langle\mu|, \quad (4.22)$$

where $\rho(S)$ and $\rho(\mu)$ must be nonnegative. Now,

$$\mathrm{Tr}\,\rho(s) = \int dS\rho(S)\,\mathrm{Tr}\,|S\rangle\langle S| = \int dS\rho(S)\phi(S)$$
$$= \epsilon\int dS\rho(S)\Sigma_\mu\delta(S - \mu), \quad (4.23)$$

$$\mathrm{Tr}\,\rho(s) = \Sigma_\mu\rho(\mu). \quad (4.24)$$

In order to satisfy the normalization condition

$$\mathrm{Tr}\,\rho(s) = 1 \quad (4.25)$$

for both (4.23) and (4.24), it is necessary that

$$\rho(S) = \Sigma_\mu\delta(S - \mu)\rho(\mu). \quad (4.26)$$

The expectation of s in the state specified by $\rho(s)$ is, accordingly,

$$\mathrm{Tr}\,[s\rho(S)] = \int dS\rho(S)\,\mathrm{Tr}\,[s\,|S\rangle\langle S|]$$
$$= \int dSS\rho(S)\phi(S)$$
$$= \Sigma_\mu\mu\rho(\mu). \quad (4.27)$$

The probability density $\rho(S)$ vanishes for $S \neq \mu$; only the eigenvalues μ of the discrete spectrum are measurable in the state $\rho(s)$ with probability weights $\rho(\mu)$.

If the one-dimensional basis of (4.1), in addition to the discrete set of kets $\{|\mu\rangle\}$, has a continuous range of kets $\{|\nu\rangle\}$ in the interval $\alpha < \eta < \beta$, then, in place of

(4.1), the completeness condition becomes

$$\Sigma_\mu |\mu\rangle\langle\mu| + \int_\alpha^\beta d\nu\, |\nu\rangle\langle\nu| = \mathbf{1}, \tag{4.28}$$

with orthogonality conditions

$$\langle\mu' \mid \mu\rangle = \delta_{\mu',\mu}, \quad \langle\nu \mid \mu\rangle = 0, \quad \langle\nu' \mid \nu\rangle = \delta(\nu' - \nu). \tag{4.29}$$

(Several continuous ranges could occur, but we illustrate the method by allowing only one.) The transformation operator U is defined, as in (4.2), by

$$U^\dagger |K\rangle = \epsilon^{\frac{1}{2}}\Sigma_\mu \delta(K - \mu)\, |\mu\rangle + \int d\nu\theta(\nu; \alpha, \beta)\delta(K - \nu)\, |\nu\rangle, \tag{4.30}$$

where $\theta(\nu; \alpha, \beta)$ is a step function equal to unity for $\alpha < \nu < \beta$, vanishing for $\nu < \alpha$ or $\nu > \beta$. Again, ϵ is chosen so that $U^\dagger U = \mathbf{1}$, i.e.,

$$\begin{aligned} U^\dagger U = &\int dK \Big\{\epsilon^{\frac{1}{2}}\Sigma_\mu \delta(K - \mu)\, |\mu\rangle \\ &+ \int d\nu\theta(\nu; \alpha, \beta)\delta(K - \nu)\, |\nu\rangle\Big\} \\ &\times \Big\{\epsilon^{\frac{1}{2}}\Sigma_{\mu'}\delta(K - \mu')\, \langle\mu'| \\ &+ \int d\nu'\theta(\nu'; \alpha, \beta)\delta(K - \nu')\, \langle\nu'|\Big\} \\ = &\Sigma_\mu |\mu\rangle\langle\mu| + \int d\nu\theta(\nu; \alpha, \beta)\, |\nu\rangle\langle\nu| = \mathbf{1}, \end{aligned} \tag{4.31}$$

when ϵ is specified as in (4.7) so that (4.10) is valid. But

$$\begin{aligned} UU^\dagger = &\int dK\, |K\rangle \Big\{\epsilon\Sigma_\mu \delta(K - \mu) \\ &+ \int d\nu\theta(\nu; \alpha, \beta)\delta(K - \nu)\Big\}\, \langle K| \\ = &\epsilon\Sigma_\mu \delta(k - \mu) + \int d\nu\theta(\nu; \alpha, \beta)\delta(k - \nu). \end{aligned} \tag{4.32}$$

Again, U is a one-sided unitary operator, and $UU^\dagger$ commutes with k.

$$\phi(K) = \epsilon\Sigma_\mu \delta(K - \mu) + \theta(K; \alpha, \beta). \tag{4.33}$$

Using the operator U defined by (4.30), define the kets $|S\rangle$ and $|R\rangle$ as

$$\begin{aligned} |S\rangle = U^\dagger |K\rangle_{K=S} &= \epsilon^{\frac{1}{2}}\Sigma_\mu \delta(S - \mu)\, |\mu\rangle \\ &+ \int d\nu\theta(\nu; \alpha, \beta)\delta(S - \nu)\, |\nu\rangle, \\ |R\rangle = U^\dagger |Q\rangle_{Q=R} &= \epsilon^{\frac{1}{2}}\Sigma_\mu \exp(-2\pi i\mu R)\, |\mu\rangle \\ &+ \int d\nu\theta(\nu; \alpha, \beta) \exp(-2\pi i\nu R)\, |\nu\rangle, \end{aligned} \tag{4.34}$$

for every real value of S and R. $\{|S\rangle\}$ and $\{|R\rangle\}$ are complete sets. The operators r, s defined according to (2.7) are canonical for the one-dimensional case. The operator s has the set of eigenkets $\{|\mu\rangle, |\nu\rangle\}$ of (4.28), i.e.,

$$s\, |\mu\rangle = \mu\, |\mu\rangle, \quad s\, |\nu\rangle = \nu\, |\nu\rangle, \tag{4.35}$$

but also the set $\{|S\rangle\}$,

$$s\, |S\rangle = S\, |S\rangle, \tag{4.36}$$

just as in (4.18) and (4.19). If $\phi(S) = 1$ for all S, then $UU^\dagger = 1$ so that U is unitary. Equation (4.33) shows that $\phi(S) = 1$ when S is in the range $\alpha < S < \beta$ of the continuous spectrum. In the general case of several continuous ranges, additional terms like $\theta(K; \alpha, \beta)$ would appear in (4.33), one for each continuous range. U is unitary when the continuous range extends over the entire real domain.

From the discussion of this section a general prescription can be given for obtaining canonically conjugate operators for one-dimensional systems. If, for any quantum observable, the solution of its eigenvalue problem yields the set of orthonormal eigenvectors $\{|\mu\rangle, |\nu\rangle\}$, or if a basis of such vectors is given as in (4.28) and (4.29), construct the set $\{|S\rangle\}$ according to (4.34). The set $\{|R\rangle\}$ can be obtained by Fourier inversion according to (3.18). The canonical operators r and s are then defined by

$$r = \int dR\, |R\rangle\, R\, \langle R|, \quad s = \int dS\, |S\rangle\, S\, \langle S|.$$

The properties of these operators are determined by $\phi(S)$ as in (4.33).

5. GENERALIZATION TO HIGHER DIMENSIONS

The number of dimensions of a system is the number of operators required to constitute a complete commuting set. The concept of a complete commuting set of observables has been discussed by Dirac.[1] To each set of eigenvalues $\{Q_1, Q_2, \cdots, Q_f\} = \mathbf{Q}$ of the complete commuting set $\{q_1, q_2, \cdots, q_f\} = \mathbf{q}$ corresponds one and only one eigenvector $|Q_1, Q_2, \cdots, Q_f\rangle = |\mathbf{Q}\rangle$. The dimensionality of the system is f. In (2.1) and (2.2) the vectors $\mathbf{Q}$ and $\mathbf{K}$ locate points in an f-dimensional configuration space and an f-dimensional momentum space, respectively; the pair $(\mathbf{Q}, \mathbf{K})$ specifies a point in the $2f$-dimensional phase space. The transformations (2.7) when U is unitary do not affect the dimensionality of the space. The operators $\mathbf{r}$ and $\mathbf{s}$ have f components, and for each value of $\mathbf{R}$ and $\mathbf{S}$ the eigenvectors $|\mathbf{R}\rangle = |R_1, R_2, \cdots, R_f\rangle$ and $|\mathbf{S}\rangle = |S_1, S_2, \cdots, S_f\rangle$ are simultaneous eigenvectors for the f components of $\mathbf{r}$ and $\mathbf{s}$, respectively. But the

spectra of $\mathbf{r}$ and $\mathbf{s}$ when U is unitary are continuous with delta-function normalization on the entire real domain of $\mathbf{R}$ and $\mathbf{S}$. Noncontinuous (discrete) spectra of $\mathbf{s}$ (or $\mathbf{r}$) require U to be one-sided unitary, with $U^\dagger U = \mathbf{1}$ and $UU^\dagger$ commuting with $\mathbf{k}$ (or $\mathbf{q}$). In this case the f components of $\mathbf{s}$ are a complete commuting set of observables, with simultaneous eigenvectors $|\mathbf{S}\rangle$ (not, however, normalized to a delta function for all $\mathbf{S}$), but the f components of $\mathbf{r} = U^\dagger \mathbf{q} U$ do not commute with each other and $|\mathbf{R}\rangle = U^\dagger |\mathbf{Q}\rangle$ is not their eigenvector.

The generalization to higher dimensions of the discussion in Sec. 4 of the one-dimensional case with noncontinuous spectrum is straightforward, though notationally involved. Consider an f-dimensional basis vector $|M_1, M_2, \cdots, M_f\rangle$. Each parameter M_i may assume discrete or continuous values. For simplicity allow only one continuous range for each M_i. Then, with the notation $M_i = \mu_i$ when μ_i is discrete and $M_i = \nu_i$ when ν_i is in the continuous range $\alpha_i < \nu_i < \beta_i$, there are f^2 kinds of kets $|M_1, \cdots, M_f\rangle$, since each M_i can be a μ_i or a ν_i. Take the case $f = 2$; the four kets are $|\mu_1, \mu_2\rangle$, $|\mu_1, \nu_2\rangle$, $|\nu_1, \mu_2\rangle$, $|\nu_1, \nu_2\rangle$. These kets are complete and orthonormal:

$$\begin{aligned}\Sigma_{\mu_1}\Sigma_{\mu_2} |\mu_1, \mu_2\rangle\langle\mu_1, \mu_2| &+ \Sigma_{\mu_1}\int d\nu_2 |\mu_1, \nu_2\rangle\langle\mu_1, \nu_2| \\ &+ \Sigma_{\mu_2}\int d\nu_1 |\nu_1, \mu_2\rangle\langle\nu_1, \mu_2| \\ &+ \int d\nu_1 \int d\nu_2 |\nu_1, \nu_2\rangle\langle\nu_1, \nu_2| = \mathbf{1}, \quad (5.1)\end{aligned}$$

$$\begin{aligned}\langle\mu_1', \mu_2' | \mu_1, \mu_2\rangle &= \delta_{\mu_1', \mu_1}\delta_{\mu_2', \mu_2}, \\ \langle\nu_1', \nu_2' | \nu_1, \nu_2\rangle &= \delta(\nu_1' - \nu_1)\delta(\nu_2' - \nu_2), \\ \langle\mu_1', \nu_2' | \mu_1, \nu_2\rangle &= \delta_{\mu_1', \mu_1}\delta(\nu_2' - \nu_2), \\ \langle\nu_1', \mu_2' | \nu_1, \mu_2\rangle &= \delta(\nu_1' - \nu_1)\delta_{\mu_2', \mu_2}. \end{aligned} \quad (5.2)$$

Scalar products of all other combinations vanish.

For the case $f = 2$, construct the ket $|S_1, S_2\rangle$ defined for all real values of S_1, S_2 as

$$\begin{aligned}|S_1, S_2\rangle = \epsilon\Sigma_{\mu_1}\Sigma_{\mu_2}\delta(S_1 - \mu_1)\delta(S_2 - \mu_2) |\mu_1, \mu_2\rangle \\ + \epsilon^{\frac{1}{2}}\Sigma_{\mu_1}\int d\nu_2 \theta(\nu_2; \alpha_2, \beta_2) \\ \times \delta(S_1 - \mu_1)\delta(S_2 - \nu_2) |\mu_1, \nu_2\rangle \\ + \epsilon^{\frac{1}{2}}\Sigma_{\mu_2}\int d\nu_1 \theta(\nu_1; \alpha_1, \beta_1) \\ \times \delta(S_1 - \nu_1)\delta(S_2 - \nu_2) |\nu_1, \mu_2\rangle \\ + \int d\nu_1 \int d\nu_2 \theta(\nu_1; \alpha_1, \beta_1)\theta(\nu_2; \alpha_2, \beta_2) \\ \times \delta(S_1 - \nu_1)\delta(S_2 - \nu_2) |\nu_1, \nu_2\rangle. \quad (5.3)\end{aligned}$$

The analogous ket $|S_1, \cdots, S_f\rangle$ would have f^2 terms on the right-hand side. The transformation operator U is given by

$$|S_1, \cdots, S_f\rangle = |\mathbf{S}\rangle = U^\dagger |\mathbf{K}\rangle_{\mathbf{K}=\mathbf{S}}. \quad (5.4)$$

This is the generalization of (4.34). Again ϵ is chosen according to (4.7) and (4.10) so that

$$U^\dagger U = \int d\mathbf{S} |\mathbf{S}\rangle\langle\mathbf{S}| = \mathbf{1}. \quad (5.5)$$

But $UU^\dagger$ is an operator which commutes with $\mathbf{k}$. The important point is that U is a one-sided unitary operator of the type considered in Sec. 3. Only if the basis $|M_1, M_2, \cdots, M_f\rangle$ is continuous with delta-function normalization over the whole real domain of $\{M_1, M_2, \cdots, M_f\}$ is U unitary. The Hermitian operators $\mathbf{r}$, $\mathbf{s}$ defined by (2.7) satisfy the commutation relations on their components:

$$[s_m, r_n] = (2\pi i)^{-1}\delta_{m,n}\mathbf{1}, \quad [s_m, s_n] = 0;$$

but, in general, $[r_m, r_n] \neq 0$ for $m \neq n$. The kets $|\mathbf{S}\rangle$ in (5.4) are simultaneous eigenkets for the f components of $\mathbf{s}$, as are the kets $|M_1, M_2, \cdots, M_f\rangle$. When the M basis is not continuous, $\{|M_1, \cdots, M_f\rangle\}$ is a set of eigenvectors for an observable s, but that observable cannot be generated by unitary transformation of coordinate and momentum operators.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Canonical Operators for the Simple Harmonic Oscillator

BORIS LEAF
Department of Physics, State University of New York at Cortland, Cortland, New York,
and
Department of Physics, State University of New York at Binghamton, Binghamton, New York

(Received 30 April 1969)

Angle and action operators (w, j) for the simple harmonic oscillator are treated as resulting from a canonical transformation of coordinate and momentum operators (q, k) generated by a one-sided unitary operator U such that $U^\dagger U = \mathbf{1}$ and $UU^\dagger$ commutes with k but not with q. From the discrete spectrum of the number operator n, eigenvectors $|\eta\rangle$ are constructed for every real value of η; the set $\{|\eta\rangle\}$ is complete and orthogonal. Another complete set $\{|W\rangle\}$ is obtained, consisting of the Fourier transforms of the kets in the set $\{|\eta\rangle\}$. The angle operator is $w = U^\dagger q U = \int dW\,|W\rangle\, W\,\langle W|$. The set $\{|W\rangle\}$ is not orthogonal; $|W\rangle$ is not an eigenvector of w. If v is defined as $\int dW\,|W\rangle \exp(-2\pi i W)\langle W|$, then the creation and destruction operators are given by $a = vn^{\frac{1}{2}}$, $a^\dagger = n^{\frac{1}{2}}v^\dagger$. v is a one-sided unitary operator such that $vv^\dagger = \mathbf{1}$, but $v^\dagger v = \mathbf{1} - |0\rangle\langle 0|$, where $|0\rangle$ is the ground state of the oscillator; v and $v^\dagger$ are similar to the operators E_- and E_+ of Carruthers and Nieto. The Weyl transforms of w and $j = 2\pi\hbar(n + \frac{1}{2}\mathbf{1})$ are the classical angle and action variables of the oscillator. The Weyl transform is formulated in terms of the coherent states of the oscillator. A time operator canonical to the Hamiltonian is defined as $t = 2\pi w/\omega$ ($\omega/2\pi$ = frequency). The observables for the oscillator are also given in the Heisenberg picture and their classical limits are considered.

1. INTRODUCTION

In the preceding paper,[1] it is shown that quantum observables, with spectra which are not continuous over the entire real domain, are generated by one-sided unitary transformation of the Cartesian coordinate or momentum operators. For one-dimensional systems such transformations are canonical. In this paper, the formalism is applied to the case of the simple harmonic oscillator. In addition to providing an example of the formalism, the problem of the oscillator is of interest in itself. Susskind and Glogower[2] have shown that the expression for the destruction operator a, as given by Dirac,[3] in terms of the Hermitian angle operator w and number operator n, in the form

$$a = vn^{\frac{1}{2}} \tag{1.1}$$

with

$$v = \exp(-2\pi i w), \tag{1.2}$$

cannot be correct. It can be shown that v in (1.1) annihilates the ground state of the oscillator so that

$$\langle 0|\, v^\dagger v\,|0\rangle = 0, \tag{1.3}$$

where $|0\rangle$ represents the ground state, in contradiction to the unitarity of v in (1.2). A review of angle variables in quantum mechanics by Carruthers and Nieto has recently appeared[4] in which problems associated with the oscillator are considered. Susskind and Glogower deny the existence of an angle operator for the oscillator. Carruthers and Nieto develop properties of operators C and S which are Hermitian analogs of the classical functions $\cos\phi$ and $\sin\phi$ of the angle variable.

Basically, the difficulties of the oscillator problem stem from the discreteness of the spectrum of the Hamiltonian and number operators. Because of this discreteness, the operator U which generates the canonical transformation to the Hermitian angle and number operators (w, n) from the coordinate and momentum operators (q, k) according to

$$w = U^\dagger q U, \quad n = U^\dagger k U \tag{1.4}$$

is a one-sided unitary operator such that $UU^\dagger$ commutes with k but not with q.[1] The operator w is Hermitian and can be written in diagonal form in the complete set of kets $\{|W\rangle\}$:

$$w = \int dW\,|W\rangle\, W\,\langle W|. \tag{1.5}$$

However, $|W\rangle$ is not an eigenket of w. The set $\{|W\rangle\}$ is not orthogonal. Properties of n and w and their spectra are considered in Sec. 2.

In Sec. 3, the displacement operator for the spectra of w and n are considered. In particular, the lowering operator for the spectrum of n is

$$v = \int dW\,|W\rangle \exp(-2\pi i W)\langle W|. \tag{1.6}$$

If $|N\rangle$ is an eigenket of the discrete spectrum of n,

[1] B. Leaf, J. Math. Phys. **10**, 1971 (1969).
[2] L. Susskind and J. Glogower, Physics **1**, 49 (1964).
[3] P. A. M. Dirac, Proc. Roy. Soc. (London), Ser. A, **114**, 243 (1927).
[4] P. Carruthers and M. M. Nieto, Rev. Mod. Phys. **40**, 411 (1968).

then it is shown that

$$v^\dagger |N\rangle = |N+1\rangle; \quad v|N\rangle = |N-1\rangle \quad \text{for} \quad N>0, \qquad v|0\rangle = 0. \quad (1.7)$$

But v is not unitary. Since

$$vv^\dagger = 1, \quad \text{but} \quad v^\dagger v = \sum_{N=1}^{\infty} |N\rangle\langle N|, \qquad (1.8)$$

v and $v^\dagger$ are similar to the operators E_- and E_+, respectively, given in Sec. 5 of the Carruthers and Nieto paper.[4] In fact, from (3.16) shown below, it follows that the choices

$$C = \tfrac{1}{2}(v + v^\dagger), \quad S = \frac{1}{2i}(v - v^\dagger), \qquad (1.9)$$

satisfy the commutation relation

$$[C, n] = iS, \quad [S, n] = -iC, \qquad (1.10)$$

from which the properties of the operators C and S are developed.

The expression for v in (1.6) rather than that in (1.2) is correct. It is shown in Sec. 4 that with v given by (1.6), the creation and destruction operators $(a^\dagger, a)$ are

$$a = vn^{\frac{1}{2}}, \quad a^\dagger = n^{\frac{1}{2}}v^\dagger. \qquad (1.11)$$

The relationship of the operators of the oscillator to the classical dynamical variables is examined in Sec. 5. The Weyl transforms[5] of the operators $j = 2\pi\hbar(n + \frac{1}{2})$ and w are the classical action and angle variables, respectively. A formulation of the Weyl transform in terms of the coherent states of the oscillator[6] is also given in Sec. 5.

In Sec. 6, a time operator for the oscillator is defined as $t = 2\pi w/\omega$, where $\omega/2\pi$ is the frequency. It is the operator canonically conjugate to the Hamiltonian. Expressions in the Heisenberg picture are given for w and t. The Weyl transform of the Heisenberg time operator is the same as the physical time measured by the increase in phase of the oscillator. It is an internal property of the oscillator.[7]

2. NUMBER AND PHASE (ACTION AND ANGLE) OPERATORS

The Hamiltonian for a simple harmonic oscillator, with coordinate q and momentum $p = 2\pi\hbar k$, is

$$H = p^2/2m + \tfrac{1}{2}m\omega^2 q^2, \qquad (2.1)$$

m = mass, $\omega/2\pi$ = frequency. It can also be written

$$H = \hbar\omega(n + \tfrac{1}{2}\mathbf{1}) = j\omega/2\pi, \qquad (2.2)$$

where n is the number operator and j is the action operator. The angle or phase operator w canonically conjugate to j and $2\pi\hbar n$ must satisfy the commutation relation

$$[w, j] = i\hbar\mathbf{1} \quad \text{or} \quad [w, n] = i\mathbf{1}/2\pi. \qquad (2.3)$$

The eigenvalue problem for n,

$$n|N\rangle = N|N\rangle, \qquad (2.4)$$

has a solution with a discrete spectrum,

$$|N\rangle = (b/\pi^{\frac{1}{2}}2^N N!)^{\frac{1}{2}} \int dQ h_N(bQ) \exp(-\tfrac{1}{2}b^2Q^2)|Q\rangle, \qquad (2.5)$$

for integral $N = 0, 1, 2, \cdots$. The kets $|Q\rangle$ are eigenvectors of the position operator q; $b = (m\omega/\hbar)^{\frac{1}{2}}$ is a constant with dimensions (length)$^{-1}$, and h_N is the Hermite polynomial of order N. The N-basis is complete and orthonormal with Kronecker-delta normalization:

$$\Sigma_N |N\rangle\langle N| = 1, \quad \langle N' \mid N\rangle = \delta_{N',N}. \qquad (2.6)$$

Construct the ket

$$|\eta\rangle = \epsilon^{\frac{1}{2}}\Sigma_N \delta(\eta - N)|N\rangle \qquad (2.7)$$

for every real number η. Then

$$n|\eta\rangle = \epsilon^{\frac{1}{2}}\Sigma_N \delta(\eta - N)N|N\rangle = \eta|\eta\rangle, \qquad (2.8)$$

so that $|\eta\rangle$ is an eigenket of n belonging to the eigenvalue η. The constant ϵ is determined by the requirement that the set $\{|\eta\rangle\}$ be complete:

$$\int_{-\infty}^{\infty} d\eta |\eta\rangle\langle\eta| = \int_{-\infty}^{\infty} d\eta \epsilon \Sigma_N \Sigma_{N'} \delta(\eta - N)\delta(\eta - N')|N\rangle\langle N'|. \qquad (2.9)$$

But ϵ can be chosen so that[1]

$$\epsilon\delta(\eta - N)\delta(\eta - N') = \epsilon\delta_{N',N}|\delta(\eta - N)|^2 = \delta_{N',N}\delta(\eta - N). \qquad (2.10)$$

Accordingly,

$$\int_{-\infty}^{\infty} d\eta |\eta\rangle\langle\eta| = \mathbf{1}. \qquad (2.11)$$

On the other hand, the set $\{|\eta\rangle\}$ is not normalized to a delta function, since

$$\langle\eta' \mid \eta\rangle = \epsilon\Sigma_N \delta(\eta' - N)\delta(\eta - N) = \delta(\eta' - \eta)\epsilon\Sigma_N \delta(\eta - N). \qquad (2.12)$$

[5] B. Leaf, J. Math. Phys. 9, 65 and 769 (1968).
[6] R. S. Glauber, "Photon Statistics," in *Fundamental Problems in Statistical Mechanics*, E. G. D. Cohen, Ed. (John Wiley & Sons, Inc., New York, 1968), p. 155.
[7] Y. Aharanov and D. Bohm, Phys. Rev. 122, 1649 (1961).

The norm of $|\eta\rangle$ vanishes unless η is a point in the discrete set $\{0, 1, 2, \cdots\}$. The ket $|\eta\rangle$ is obtained by one-sided unitary transformation of $|K\rangle$, the eigenket of momentum operator $k = p/2\pi\hbar$. The transformation operator U is defined by

$$|\eta\rangle = U^\dagger |K\rangle_{K=\eta}, \tag{2.13}$$

$$U^\dagger |K\rangle = \epsilon^{\frac{1}{2}}\Sigma_N \delta(K - N)|N\rangle,$$

$$U^\dagger = \epsilon^{\frac{1}{2}}\int_{-\infty}^{\infty} dK \Sigma_N \delta(K - N)|N\rangle\langle K|, \tag{2.14}$$

so that

$$U^\dagger U = 1, \quad UU^\dagger = \int_{-\infty}^{\infty} dK |K\rangle \epsilon\Sigma_N \delta(K - N)\langle K|. \tag{2.15}$$

$UU^\dagger$ commutes with k, so that

$$UU^\dagger |K\rangle = \phi(K)|K\rangle. \tag{2.16}$$

Accordingly, $|K\rangle$ is an eigenket of the Hermitian operator $UU^\dagger$ belonging to the eigenvalue

$$\phi(K) = \epsilon\Sigma_N \delta(K - N). \tag{2.17}$$

From (2.12),

$$|\eta\rangle = \int_{-\infty}^{\infty} d\eta' \delta(\eta' - \eta)\epsilon\Sigma_N \delta(\eta - N)|\eta'\rangle$$

$$= \epsilon\Sigma_N \delta(\eta - N)|\eta\rangle = \phi(\eta)|\eta\rangle. \tag{2.18}$$

The ket $|W\rangle$ is defined, for all values of the real number W, as

$$|W\rangle = U^\dagger |Q\rangle_{Q=W}. \tag{2.19}$$

According to (2.13), (2.16), and (2.17),

$$\langle W \mid \eta\rangle = \langle Q|_{Q=W} UU^\dagger |K\rangle_{K=\eta} = \phi(\eta)\langle Q \mid_{Q=W} K\rangle_{K=\eta}$$

$$= \epsilon\Sigma_N \delta(\eta - N)\exp(2\pi i W\eta), \tag{2.20}$$

since $\langle Q \mid K\rangle = \exp(2\pi i QK)$. From (2.18), $|W\rangle$ is the Fourier transform of $|\eta\rangle$:

$$|W\rangle = \int d\eta |\eta\rangle\langle\eta \mid W\rangle = \int d\eta |\eta\rangle \phi(\eta)\exp(-2\pi i W\eta)$$

$$= \int d\eta |\eta\rangle \exp(-2\pi i W\eta). \tag{2.21}$$

In the discrete N-representation, the ket $|W\rangle$ becomes

$$|W\rangle = \epsilon^{\frac{1}{2}}\Sigma_N |N\rangle \exp(-2\pi i NW). \tag{2.22}$$

It is periodic with period one in W. The set $\{|W\rangle\}$ is complete; from (2.10),

$$\int_{-\infty}^{\infty} dW |W\rangle\langle W|$$

$$= \epsilon \int dW \Sigma_N \Sigma_{N'} \exp[2\pi i W(N' - N)]|N\rangle\langle N'|$$

$$= \Sigma_N |N\rangle\langle N| = \mathbf{1}. \tag{2.23}$$

On the other hand, $\{|W\rangle\}$ is not orthogonal, and

$$\langle W' \mid W\rangle = \epsilon\Sigma_N \exp[2\pi i N(W' - W)]. \tag{2.24}$$

The angle operator w is diagonal in $\{|W\rangle\}$. It is defined as

$$w = \int_{-\infty}^{\infty} dW |W\rangle W \langle W|. \tag{1.5'}$$

From (2.8) and (2.13), we have

$$n = \int_{-\infty}^{\infty} d\eta |\eta\rangle \eta \langle\eta|$$

$$= U^\dagger \int_{-\infty}^{\infty} d\eta |K\rangle_{K=\eta} \eta \langle K|_{K=\eta} U = U^\dagger k U, \tag{2.25}$$

and similarly from (1.5') and (2.19),

$$w = U^\dagger \int_{-\infty}^{\infty} dW |Q\rangle_{Q=W} W \langle Q|_{Q=W} U = U^\dagger q U. \tag{2.26}$$

Since $UU^\dagger$ commutes with k, according to (2.15) we have

$$[w, n] = U^\dagger q UU^\dagger k U - U^\dagger k UU^\dagger q U$$

$$= U^\dagger [q, k] U = i\mathbf{1}/2\pi. \tag{2.27}$$

Accordingly, (2.3) is satisfied; w and n are canonically conjugate as defined. They are also Hermitian operators. From (2.8), $|\eta\rangle$ is an eigenket of n; $\{|\eta\rangle\}$ is an orthogonal set in (2.12). But from (2.24), $\{|W\rangle\}$ is not an orthogonal set; $|W\rangle$ is not an eigenket of w. From (2.24),

$$|W\rangle = \int_{-\infty}^{\infty} dW' |W'\rangle \epsilon\Sigma_N \exp[2\pi i N(W' - W)], \tag{2.28}$$

$$w|W\rangle = \int_{-\infty}^{\infty} dW' W' |W'\rangle \epsilon\Sigma_N \exp[2\pi i N(W' - W)]. \tag{2.29}$$

3. DISPLACEMENT OPERATORS

Any function $f(n)$ can be written in diagonal form. From (2.8),

$$f(n)|\eta\rangle = f(\eta)|\eta\rangle, \tag{3.1}$$

so that

$$f(n) = \int d\eta |\eta\rangle f(\eta)\langle\eta|. \tag{3.2}$$

Since

$$\langle W| f(n)|\eta\rangle = f(\eta)\epsilon\Sigma_N \delta(\eta - N)\exp(2\pi i W\eta)$$

$$= f\left[(2\pi i)^{-1}\frac{\partial}{\partial W}\right]\langle W \mid \eta\rangle,$$

we find

$$f(n) = \int dW |W\rangle f\left[(2\pi i)^{-1}\frac{\partial}{\partial W}\right]\langle W|. \tag{3.3}$$

In particular, if β is a real constant, then the unitary operator $\exp(2\pi i\beta n)$ is a displacement operator for $\{|W\rangle\}$. Thus,

$$\begin{aligned}\exp(2\pi i\beta n) &= \int d\eta\, |\eta\rangle \exp(2\pi i\beta\eta)\langle\eta| \\ &= \int dW\, |W\rangle \exp\left(\beta\frac{\partial}{\partial W}\right)\langle W| \\ &= \int dW\, |W\rangle\langle W+\beta|. \end{aligned} \tag{3.4}$$

From (2.18) and (2.20), we have

$$\begin{aligned}&\exp(2\pi i\beta n)\,|W\rangle \\ &= \int d\eta \exp(2\pi i\beta\eta)\,|\eta\rangle\langle\eta \mid W\rangle \\ &= \int d\eta\,|\eta\rangle \exp[-2\pi i\eta(W-\beta)] = |W-\beta\rangle. \end{aligned} \tag{3.5}$$

Therefore,

$$\begin{aligned}&\exp(2\pi i\beta n) w \exp(-2\pi i\beta n) \\ &\quad = \int dW(W-\beta)\, W\, \langle W-\beta| \\ &\quad = \int dW\, |W\rangle\, (W+\beta)\,\langle W| = w+\beta\mathbf{1}. \end{aligned} \tag{3.6}$$

On the other hand, a function of w is not, in general, diagonal in the set $\{|W\rangle\}$ even though, according to (1.5′), w is itself diagonal. This is a consequence of the nonorthogonality of $\{|W\rangle\}$. But a class of diagonal operators can be defined by $\int dW\,|W\rangle\, g(W)\,\langle W|$, where $g(W)$ is any function of W. Since

$$\begin{aligned}&\int dW\,|W\rangle\, g(W)\langle W\mid\eta\rangle \\ &\quad = \int dW\,|W\rangle\, g(W)\epsilon\Sigma_{N'}\delta(\eta-N)\exp(2\pi iW\eta) \\ &\quad = \epsilon\Sigma_N\delta(\eta-N) g\left[(2\pi i)^{-1}\frac{\partial}{\partial\eta}\right]|\eta\rangle, \end{aligned}$$

we find, from (2.18),

$$\begin{aligned}\int dW\,|W\rangle\, g(W)\,\langle W| &= \int d\eta\left\{g\left[(2\pi i)^{-1}\frac{\partial}{\partial\eta}\right]|\eta\rangle\right\}\langle\eta| \\ &= \int d\eta\,|\eta\rangle\, g\left[\frac{i}{2\pi}\frac{\partial}{\partial\eta}\right]\langle\eta|. \end{aligned} \tag{3.7}$$

In particular, a displacement operator for $\{|\eta\rangle\}$ can be defined by

$$\begin{aligned}v^\dagger(\beta) &= \int dW\,|W\rangle \exp(2\pi i\beta W)\,\langle W| \\ &= \int d\eta\,|\eta\rangle \exp\left(-\beta\frac{\partial}{\partial\eta}\right)\langle\eta| \\ &= \int d\eta\,|\eta\rangle\langle\eta-\beta|, \end{aligned} \tag{3.8}$$

$$\begin{aligned}v^\dagger(\beta)\,|\eta\rangle &= \int d\eta'\,|\eta'\rangle\,\delta(\eta'-\beta-\eta)\epsilon\Sigma_N\delta(\eta-N) \\ &= \epsilon\Sigma_N\delta(\eta-N)\,|\eta+\beta\rangle \\ &= \epsilon^{\frac{3}{2}}\Sigma_N\Sigma_{N'}\delta(\eta-N)\delta(\eta+\beta-N')\,|N'\rangle \\ &= \epsilon^{\frac{3}{2}}\Sigma_N\Sigma_{N'}\delta_{N',N+\beta}\delta(\eta+\beta-N')\,|N'\rangle. \end{aligned} \tag{3.9}$$

If $v^\dagger(\beta)$ is applied to one of the kets in the discrete set $\{|N\rangle\}$, then

$$v^\dagger(\beta)\,|N\rangle = \int d\eta v^\dagger(\beta)\,|\eta\rangle\langle\eta\mid N\rangle = \Sigma_{N'}\delta_{N',N+\beta}\,|N'\rangle, \tag{3.10}$$

which equals $|N+\beta\rangle$ if $N+\beta$ is a member of the discrete set of points $\{0,1,2,\cdots\}$ and vanishes otherwise. Thus,

$$v^\dagger(\beta) = \Sigma_N\Sigma_{N'}\delta_{N',N+\beta}\,|N'\rangle\langle N|. \tag{3.11}$$

The only nonvanishing cases occur for $\beta = \pm 1$. Accordingly, define the raising operator ($\beta = 1$)

$$\begin{aligned}v^\dagger &= \int dW\,|W\rangle \exp(2\pi iW)\,\langle W| \\ &= \int d\eta\,|\eta\rangle \exp\left(\frac{-\partial}{\partial\eta}\right)\langle\eta| = \int d\eta\,|\eta\rangle\langle\eta-1| \end{aligned} \tag{3.12}$$

and the lowering operator ($\beta = -1$)

$$\begin{aligned}v &= \int dW\mid W\rangle \exp(-2\pi iW)\,\langle W| \\ &= \int d\eta\,|\eta\rangle \exp\left(\frac{\partial}{\partial\eta}\right)\langle\eta| = \int d\eta\,|\eta-1\rangle\langle\eta|. \end{aligned} \tag{1.6′}$$

According to (3.11),

$$\begin{aligned}v^\dagger &= \Sigma_N\Sigma_{N'}\delta_{N',N+1}\,|N'\rangle\langle N|, \\ v &= \Sigma_N\Sigma_{N'}\delta_{N',N-1}\,|N'\rangle\langle N|. \end{aligned} \tag{3.13}$$

Since

$$vv^\dagger = \Sigma_{N=0}^\infty\,|N\rangle\langle N| = \mathbf{1}, \quad v^\dagger v = \Sigma_{N=1}^\infty\,|N\rangle\langle N|, \tag{1.8′}$$

we have that v is not unitary, but one-sided unitary. From (3.10),

$$\begin{aligned}v^\dagger\,|N\rangle &= |N+1\rangle, \quad v\,|N\rangle = |N-1\rangle \quad \text{for} \quad N>0, \\ v\,|N=0\rangle &= 0. \end{aligned} \tag{1.7′}$$

$$\begin{aligned}v^\dagger n v &= \Sigma_{N=0}^\infty\,|N+1\rangle\, N\,\langle N+1| \\ &= \Sigma_{N=1}^\infty\,|N\rangle\,(N-1)\,\langle N| = n - \Sigma_{N=1}^\infty\,|N\rangle\langle N|, \end{aligned} \tag{3.14}$$

$$\begin{aligned}vnv^\dagger &= \Sigma_{N=1}^\infty\,|N-1\rangle\, N\,\langle N-1| \\ &= \Sigma_{N=0}^\infty\,|N\rangle\,(N+1)\,\langle N| = n+\mathbf{1}, \end{aligned} \tag{3.15}$$

$$[n, v] = -v, \quad [n, v^\dagger] = v^\dagger. \tag{3.16}$$

4. CREATION AND DESTRUCTION OPERATORS

From (2.1) and (2.2), the operator n, written as

$$n = \tfrac{1}{2}[p^2/m\hbar\omega + m\omega q^2/\hbar - 1],$$

can be factored into

$$n = a^\dagger a, \tag{4.1}$$

where creation operator $a^\dagger$ and destruction operator a are the Hermitian adjoint operators:

$$\begin{aligned} a^\dagger &= 2^{-\frac{1}{2}}(-ip/b\hbar + bq) \\ &= 2^{-\frac{1}{2}}\int dQ\,|Q\rangle\left(-b^{-1}\frac{\partial}{\partial Q} + bQ\right)\langle Q|, \\ a &= 2^{-\frac{1}{2}}(ip/b\hbar + bq) \\ &= 2^{-\frac{1}{2}}\int dQ\,|Q\rangle\left(b^{-1}\frac{\partial}{\partial Q} + bQ\right)\langle Q|. \end{aligned} \tag{4.2}$$

From these expressions it is readily verified that

$$[a, a^\dagger] = \mathbf{1}, \tag{4.3}$$

$$[n, a] = -a, \quad [n, a^\dagger] = a^\dagger. \tag{4.4}$$

These properties of a, $a^\dagger$ are consistent with the identification

$$a = vn^{\frac{1}{2}}, \quad a^\dagger = n^{\frac{1}{2}}v^\dagger. \tag{4.5}$$

From (3.15),

$$aa^\dagger = vnv^\dagger = n + \mathbf{1} \tag{4.6}$$

and, from (1.8′),

$$a^\dagger a = n^{\frac{1}{2}}v^\dagger v n^{\frac{1}{2}} = \sum_{N=1}^{\infty}|N\rangle\, N\,\langle N| = n. \tag{4.1′}$$

Subtracting (4.1′) from (4.6) gives back (4.3). Also, (4.4) is a consequence of the identification (4.5), since (3.16) gives

$$\begin{aligned} [n, a] &= [n, vn^{\frac{1}{2}}] = [n, v]n^{\frac{1}{2}} = -vn^{\frac{1}{2}} = -a, \\ [n, a^\dagger] &= [n, n^{\frac{1}{2}}v^\dagger] = n^{\frac{1}{2}}[n, v^\dagger] = n^{\frac{1}{2}}v^\dagger = a^\dagger. \end{aligned}$$

From (1.6′) and (3.13),

$$\begin{aligned} a &= \int_{-\infty}^{\infty} d\eta\,|\eta - 1\rangle\,\eta^{\frac{1}{2}}\langle\eta| \\ &= \sum_{N=0}^{\infty}|N - 1\rangle\,N^{\frac{1}{2}}\langle N|, \\ a^\dagger &= \int_{-\infty}^{\infty} d\eta\,|\eta + 1\rangle\,(\eta + 1)^{\frac{1}{2}}\langle\eta| \\ &= \sum_{N=0}^{\infty}|N + 1\rangle\,(N + 1)^{\frac{1}{2}}\langle N|. \end{aligned} \tag{4.7}$$

Accordingly, we find that

$$a\,|N\rangle = N^{\frac{1}{2}}\,|N - 1\rangle, \quad a^\dagger\,|N\rangle = (N + 1)^{\frac{1}{2}}\,|N + 1\rangle, \tag{4.8}$$

the expressions from which $a^\dagger$ and a obtain their designations as creation and destruction operators.[8]

[8] A. Messiah, *Quantum Mechanics* (John Wiley & Sons, Inc., New York, and North-Holland Publ. Co., Amsterdam, 1961), Chap. XII.

The operator j in (2.2) can be written as

$$j = 2\pi\hbar(n + \tfrac{1}{2}\mathbf{1}) = 2\pi\hbar\cdot\tfrac{1}{2}(aa^\dagger + a^\dagger a). \tag{4.9}$$

5. WEYL TRANSFORMS AND COHERENT STATES

A convenient way to examine the classical limit of quantum operators is to obtain their Weyl transforms.[5] The Weyl transform of an operator A is defined for a one-dimensional system as

$$A(Q, K) = \mathrm{Tr}\,[A\Delta(Q, K)]. \tag{5.1}$$

The inverse transform is

$$A = \int dQ\int dK A(Q, K)\Delta(Q, K). \tag{5.2}$$

The operator $\Delta(Q, K)$, defined as

$$\Delta(Q, K) = \int du\int dv\,\exp\{2\pi i[u(q - Q) + v(k - K)]\}, \tag{5.3}$$

can also be written as

$$\begin{aligned} \Delta(Q, K) &= \int dv\,|Q + \tfrac{1}{2}v\rangle\langle Q - \tfrac{1}{2}v|\,\exp(2\pi ivK) \\ &= \int du\,|K + \tfrac{1}{2}u\rangle\langle K - \tfrac{1}{2}u|\,\exp(-2\pi iuQ). \end{aligned} \tag{5.4}$$

The Weyl transform of any operator $f(q)$, a function of q alone, is $f(Q)$; the transform of $f(k)$, where $k = p/2\pi\hbar$, is $f(K)$. Accordingly, the transforms of many of the operators of the simple harmonic oscillator can be written down by inspection. From (2.1) and (2.2),

$$H(Q, K) = P^2/2m + \tfrac{1}{2}m\omega^2Q^2 \equiv E \tag{5.5}$$

($P = 2\pi\hbar K$), so that $H(Q, K)$ is the energy of the oscillator E. Also,

$$j(Q, K) = 2\pi E/\omega = 2\pi\hbar[n(Q, K) + \tfrac{1}{2}] \equiv J, \tag{5.6}$$

where $j(Q, K)$, the Weyl transform of j, is the classical action variable J. From (4.2), the Weyl transforms of the creation and destruction operators are

$$\begin{aligned} a^\dagger(Q, K) &= 2^{-\frac{1}{2}}[-iP/b\hbar + bQ] \equiv \alpha^*, \\ a(Q, K) &= 2^{-\frac{1}{2}}[iP/b\hbar + bQ] \equiv \alpha. \end{aligned} \tag{5.7}$$

Accordingly,

$$\begin{aligned} \alpha^*\alpha &= \tfrac{1}{2}(P^2/b^2\hbar^2 + b^2Q^2) \\ &= E/\omega\hbar = n(Q, K) + \tfrac{1}{2} = J/2\pi\hbar. \end{aligned} \tag{5.8}$$

In polar form,

$$\alpha = (\alpha^*\alpha)^{\frac{1}{2}}\exp(-2\pi i\phi) = (J/2\pi\hbar)^{\frac{1}{2}}\exp(-2\pi i\phi), \tag{5.9}$$

where

$$\cos 2\pi\phi = \frac{bQ}{(J/\pi\hbar)^{\frac{1}{2}}}, \quad \sin 2\pi\phi = \frac{-P/b\hbar}{(J/\pi\hbar)^{\frac{1}{2}}}. \tag{5.10}$$

Equation (5.10) is equivalent to

$$\begin{aligned} Q &= (J/\pi m\omega)^{\frac{1}{2}} \cos 2\pi\phi, \\ P &= -(Jm\omega/\pi)^{\frac{1}{2}} \sin 2\pi\phi, \end{aligned} \tag{5.11}$$

which are the equations for the classical canonical transformation from coordinate and momentum variables (Q, P) to angle and action variable (J, ϕ). Thus, ϕ is the classical angle or phase variable canonically conjugate to J.

In fact, ϕ is the Weyl transform of the operator w canonically conjugate to j. In order to show this, start with the Weyl transform of the commutation relation (2.3). If the Weyl transform of w is $w(Q, K)$, then it must satisfy the condition[5]

$$\begin{aligned} \int dQ' \int dK' \delta(Q' - Q)\delta(K' - K) 2i \\ \times \sin\left\{(4\pi)^{-1}\left(\frac{\partial}{\partial Q}\frac{\partial}{\partial K'} - \frac{\partial}{\partial K}\frac{\partial}{\partial Q'}\right)\right\} \\ \times w(Q, K) j(Q', K') = i\hbar. \end{aligned} \tag{5.12}$$

Only derivatives of odd order contribute to this expression; from (5.6),

$$\frac{\partial j(QK)}{\partial Q} = 2\pi\hbar b^2 Q, \quad \frac{\partial j(QK)}{\partial K} = 8\pi^3 \hbar \frac{K}{b^2},$$

so that (5.12) becomes the Poisson bracket relation:

$$\begin{aligned} &[w(QK), j(QK)]_{\text{P.B.}} \\ &\quad = \frac{4\pi^2\hbar K}{m\omega} \cdot \frac{\partial w(Q, K)}{\partial Q} - \frac{m\omega Q}{\hbar} \cdot \frac{\partial w(Q, K)}{\partial K} = 1. \end{aligned} \tag{5.13}$$

The solution of this equation is readily verified to be

$$2\pi w(Q, K) = -\tan^{-1}(2\pi\hbar K/m\omega Q), \tag{5.14}$$

so that, according to (5.10), with $P = 2\pi\hbar K$,

$$w(Q, K) = \phi. \tag{5.15}$$

The inverse Weyl transformation (5.2) gives the angle operator as

$$w = -(2\pi)^{-1} \int dQ \int dK \tan^{-1}(2\pi\hbar K/m\omega Q) \Delta(QK). \tag{5.16}$$

The operator $\Delta(QK)$ given in (5.3) can be rewritten in terms of $a^\dagger$, a in (4.2) and their Weyl transforms in (5.7).

Since

$$\begin{aligned} q &= (a + a^\dagger)/\sqrt{2}b, \quad k = (a - a^\dagger)b/2\pi i\sqrt{2}, \\ Q &= (\alpha + \alpha^*)/\sqrt{2}b, \quad K = (\alpha - \alpha^*)b/2\pi i\sqrt{2}, \end{aligned} \tag{5.17}$$

therefore

$$\begin{aligned} \Delta(\alpha^*, \alpha) = \int du \int dv \\ \times \exp\{2\pi i[(u/b\sqrt{2})(a + a^\dagger - \alpha - \alpha^*) \\ + (vb/2\pi i\sqrt{2})(a - a^\dagger - \alpha + \alpha^*)]\}. \end{aligned}$$

With change in variables $(2\pi u/b\sqrt{2}, vb/\sqrt{2}) \to (-u, v)$, this becomes

$$\begin{aligned} \Delta(\alpha^*, \alpha) = \pi^{-1} \int du \int dv \exp[(v - iu)(a - \alpha) \\ - (v + iu)(a^\dagger - \alpha^*)]. \end{aligned} \tag{5.18}$$

Write $\beta = v + iu$, and use the notation

$$d^2\beta = d(\text{Re } \beta)\, d(\text{Im } \beta) = dv\, du. \tag{5.19}$$

Then

$$\begin{aligned} \Delta(\alpha^*, \alpha) &= \pi^{-1} \iint d^2\beta \exp[\beta^*(a - \alpha) - \beta(a^\dagger - \alpha^*)] \\ &= \pi^{-1} \iint d^2\beta\, D^\dagger(\beta) \exp(\beta\alpha^* - \beta^*\alpha) \\ &= \pi^{-1} \iint d^2\beta\, D(\beta) \exp(\beta^*\alpha - \beta\alpha^*). \end{aligned} \tag{5.20}$$

The operator $D(\beta)$ is defined as

$$D(\beta) = \exp(\beta a^\dagger - \beta^* a). \tag{5.21}$$

$D(\beta)$ is a unitary operator for which[9]

$$\begin{aligned} D^\dagger(\beta) a D(\beta) &= a + \beta, \\ D^\dagger(\beta) a^\dagger D(\beta) &= a^\dagger + \beta^*. \end{aligned} \tag{5.22}$$

For every complex number α there is an eigenvector $|\alpha\rangle$ of the destructor operator a,

$$a|\alpha\rangle = \alpha|\alpha\rangle, \tag{5.23}$$

the coherent states.[4,6] They are complete,

$$\pi^{-1} \iint d^2\alpha\, |\alpha\rangle\langle\alpha| = \mathbf{1}, \tag{5.24}$$

but not orthogonal,

$$\langle\beta|\alpha\rangle = \exp(\beta^*\alpha - \tfrac{1}{2}|\beta|^2 - \tfrac{1}{2}|\alpha|^2). \tag{5.25}$$

From (5.22) and (5.23),

$$D^\dagger(\beta) a |\alpha\rangle = (a + \beta) D^\dagger(\beta) |\alpha\rangle = \alpha D^\dagger(\beta) |\alpha\rangle,$$

so that

$$a D^\dagger(\beta) |\alpha\rangle = (\alpha - \beta) D^\dagger(\beta) |\alpha\rangle.$$

[9] See Ref. 6, p. 158.

Therefore, $D^\dagger(\beta)\,|\alpha\rangle$ is an eigenvector of a belonging to $\alpha - \beta$. Consistency with (5.25) requires

$$D^\dagger(\beta)\,|\alpha\rangle = \exp\,[\tfrac{1}{2}(\beta^*\alpha - \beta\alpha^*)]\,|\alpha - \beta\rangle,$$
$$D(\beta)\,|\alpha\rangle = \exp\,[\tfrac{1}{2}(\beta\alpha^* - \beta^*\alpha)]\,|\alpha + \beta\rangle. \quad (5.26)$$

In particular, for $\alpha = 0$,

$$D(\beta)\,|0\rangle = |\beta\rangle, \quad (5.27)$$

which is a well-known result.[9]

Now, from (5.20), the effect of $\Delta(\alpha^*, \alpha)$ upon a coherent state $|\gamma\rangle$ is

$$\Delta(\alpha^*, \alpha)\,|\gamma\rangle$$
$$= \pi^{-1}\iint d^2\beta \exp\,[\beta^*(\alpha - \tfrac{1}{2}\gamma) - \beta(\alpha^* - \tfrac{1}{2}\gamma^*)]\,|\gamma + \beta\rangle. \quad (5.28)$$

From (5.25),

$$\langle\delta|\,\Delta(\alpha^*, \alpha)\,|\gamma\rangle = \pi^{-1}\iint d^2\beta$$
$$\times \exp\,[-\tfrac{1}{2}\,|\beta|^2 + \beta^*(\alpha - \gamma) - \beta(\alpha^* - \delta^*)]\langle\delta\,|\,\gamma\rangle$$
$$= 2\langle\delta\,|\,\gamma\rangle \exp\,[-2(\alpha^* - \delta^*)(\alpha - \gamma)]. \quad (5.29)$$

This result also follows directly from the expression for $\Delta(\alpha^*\alpha)$ obtained from (5.20):

$$\Delta(\alpha^*\alpha) = \pi^{-1}\iint d^2\beta \exp\,[-\tfrac{1}{2}\,|\beta|^2]$$
$$\times \exp\,[-\beta(a^\dagger - \alpha^*)] \exp\,[\beta^*(a - \alpha)]. \quad (5.30)$$

6. TIME OPERATOR FOR THE SIMPLE HARMONIC OSCILLATOR

From (2.2) and (2.3), a time operator t, canonically conjugate to the Hamiltonian, can be defined as

$$t = 2\pi w/\omega, \quad (6.1)$$

so that the canonical commutation relation holds, i.e.,

$$[t, H] = i\hbar\mathbf{1}. \quad (6.2)$$

The operator t is diagonal in the angle representation:

$$t = \int dW\,|W\rangle\,(2\pi W/\omega)\,\langle W|. \quad (6.3)$$

Let

$$T = 2\pi W/\omega. \quad (6.4)$$

Then

$$t = \int dT(\omega/2\pi)\,|W\rangle\,T\,\langle W| = \int dT\,|T\rangle\,T\,\langle T|, \quad (6.5)$$

where

$$|T\rangle = (\omega/2\pi)^{\frac{1}{2}}\,|W\rangle. \quad (6.6)$$

The set of vectors $\{|T\rangle\}$ is complete according to (2.23),

$$\int dT\,|T\rangle\langle T| = \mathbf{1}, \quad (6.7)$$

but not orthogonal according to (2.24),

$$\langle T'\,|\,T\rangle = \epsilon(\omega/2\pi)\Sigma_N \exp\,[iN\omega(T' - T)]. \quad (6.8)$$

The kets of the set $\{|T\rangle\}$ are not eigenvectors of the Hermitian operator t. The time operator is an internal property[7] of the oscillator determined by its phase. According to (5.15), the Weyl transform of t is

$$t(Q, K) = (2\pi/\omega)w(Q, K) = 2\pi\phi/\omega. \quad (6.9)$$

To this point the operators and their kets have been expressed in the Schrödinger picture. Operators H, j, n all commute with the Hamiltonian and remain the same in the Heisenberg picture. In order to obtain the Heisenberg angle and time operators, we use (3.6). In terms of the parameter (c number) T, the physical time, the Heisenberg angle operator $w(T)$ is

$$w(T) = \exp\,(iTH/\hbar)w \exp\,(-iTH/\hbar)$$
$$= \exp\,(i\omega Tn)w \exp\,(-i\omega Tn)$$
$$= w + \mathbf{1}T\omega/2\pi, \quad w(0) = w. \quad (6.10)$$

If the Schrödinger operator w is diagonal in the kets $\{|W_0\rangle\}$, then

$$w(T) = \int dW_0\,|W_0\rangle\,W\,\langle W_0|, \quad (6.11)$$

where

$$W = W_0 + T\omega/2\pi. \quad (6.12)$$

Similarly, the Weyl transform of (6.10) is

$$\phi(T) = \phi + T\omega/2\pi. \quad (6.13)$$

From (6.12) and (6.13) we see that the parameter W is identical with the phase angle, the Weyl transform of $w(T)$,

$$W = \phi(T), \quad W_0 = \phi. \quad (6.14)$$

Similarly, the Heisenberg time operator is

$$t(T) = t + T\mathbf{1} \quad (6.15)$$
$$= \int dT_0\,|T_0\rangle\,(T_0 + T)\,\langle T_0|. \quad (6.16)$$

The Weyl transform of $t(T)$ is

$$t(T, Q, K) = t(QK) + T, \quad (6.17)$$

with

$$t(Q, K) = T_0. \quad (6.18)$$

T is the time elapsed from the initial moment T_0.

The Heisenberg equation for the time derivative of $t(T)$ is

$$\frac{dt(T)}{dT} = (i\hbar)^{-1}[t(T), H] = (i\hbar)^{-1}[t, H] = \mathbf{1}. \quad (6.19)$$

Despite the appearance in (6.15), the Heisenberg operator $t(T)$ has no explicit dependence on T; the Schrödinger operator t is independent of T, and

$$t(T) = \exp(iTH/\hbar)t \exp(-iTH/\hbar). \quad (6.20)$$

In the Heisenberg picture, according to (6.10), (4.5), and (3.12),

$$\begin{aligned} a^\dagger(T) &= \exp(iTH/\hbar)n^{\frac{1}{2}} \\ &\quad \times \int dW\, |W\rangle \exp(2\pi iW) \langle W| \exp(-iTH/\hbar) \\ &= n^{\frac{1}{2}} \int dW\, |W - \omega T/2\pi\rangle \\ &\quad \times \exp(2\pi iW) \langle W - \omega T/2\pi|, \end{aligned}$$

so that

$$\begin{aligned} a^\dagger(T) &= a^\dagger(0) \exp(i\omega T), \\ a(T) &= a(0) \exp(-i\omega T). \end{aligned} \quad (6.21)$$

Therefore, from (4.2),

$$\begin{aligned} p(T) &= p(0) \cos(\omega T) - m\omega q(0) \sin \omega T, \\ q(T) &= q(0) \cos(\omega T) + (m\omega)^{-1} p(0) \sin \omega t, \end{aligned} \quad (6.22)$$

in agreement with the classical limit,[5,8]

$$\frac{dq(T)}{dT} = \frac{p(T)}{m}, \quad \frac{dp(t)}{dT} = -m\omega^2 q(T). \quad (6.23)$$

For any one-dimensional system whose Hamiltonian is time independent, the canonically conjugate time operator can be obtained as follows: Solve the eigenvalue problem for the Hamiltonian and construct its spectrum of eigenvectors defined for every real value of the eigenvalue E. Then

$$H = \int dE\, |E\rangle\, E\, \langle E| \quad (6.24)$$

and

$$t = \int dE\, |E\rangle\, i\hbar \frac{\partial}{\partial E} \langle E|. \quad (6.25)$$

In the Heisenberg picture,

$$\begin{aligned} t(T) &= \exp(iTH/\hbar)t \exp(-iTH/\hbar) \\ &= \int dE \exp\left(\frac{iET}{\hbar}\right) |E\rangle \left(i\hbar \frac{\partial}{\partial E}\right) \langle E| \exp\left(\frac{-iTE}{\hbar}\right) \\ &= \int dE\, |E\rangle \left(i\hbar \frac{\partial}{\partial E}\right) \langle E| + \int dE\, |E\rangle\, T\, \langle E| \\ &= t + T\mathbf{1}, \end{aligned}$$

in agreement with (6.15). Classically a canonical transformation is possible[10] for a system whose Hamiltonian is independent of time T, from the coordinate and momentum variables (Q, P) to the canonical variables (S, R) where

$$S = H, \quad R = R_0 + T. \quad (6.26)$$

The Poisson bracket condition for canonicity is satisfied since

$$1 = \frac{dR}{dT} = [R, H]_{\text{P.B.}} = [R_0, H]_{\text{P.B.}}. \quad (6.27)$$

These equations correspond precisely to (6.15) and (6.19). The classical canonical transformation is generated by Hamilton's characteristic function. The quantum-mechanical transformation is generated by an operator U:

$$H = U^\dagger k U, \quad t = U^\dagger q U. \quad (6.28)$$

Just as in the similar transformation (2.25) and (2.26), which generates n and w for the simple harmonic oscillator, the operator U in (6.28) is a one-sided unitary operator since the spectrum of H is restricted to nonnegative values, continuous or discrete.

[10] H. Goldstein, *Classical Mechanics* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1950), p. 280.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Bounds for Effective Electrical, Thermal, and Magnetic Properties of Heterogeneous Materials

MELVIN N. MILLER
P.O. Box 134, *Chalfont, Pennsylvania*

(Received 1 October 1968)

Determining the effective dielectric constant is typical of a broad class of problems that includes effective magnetic permeability, electrical and thermal conductivity, and diffusion. Bounds for these effective properties for statistically isotropic and homogeneous materials have been developed in terms of statistical information, i.e., one-point and three-point correlation functions, from variational principles. Aside from the one-point correlation function, i.e., the volume fraction, this statistical information is difficult or impossible to obtain for real materials. For a broad class of heterogeneous materials (which we shall call cell materials) the functions of the three-point correlation function that appear in the bounds of effective dielectric constant are simply a number for each phase. Furthermore, this number has a range of values $\frac{1}{9}$ to $\frac{1}{3}$ and a simple geometric significance. The number $\frac{1}{9}$ implies a spherical shape, the number $\frac{1}{3}$ a cell of platelike shape, and all other cell shapes, no matter how irregular, have a corresponding number between. Each value of this number determines a new set of bounds which are substantially narrower and always within the best bounds in terms of volume fraction alone (i.e., Hashin–Shtrikman bounds). For dilute suspensions the new bounds are so narrow in most cases as to be essentially an exact solution. There is a substantial improvement over previous bounds for a finite suspension and yet greater improvement for multiphase material where the geometric characteristics of each phase are known.

1. INTRODUCTION

From the continuum point of view, the property of a material is often described by a linear isotropic constitutive relation, e.g., Hooke's law in elasticity, $\mathbf{D} = \epsilon\mathbf{E}$ in dielectrics, Fourier's law in heat conduction, Fick's law in diffusion. In practice many materials are heterogeneous on a macroscopic scale, and it is common to replace the material property specified for a homogeneous material by an effective or over-all property. This effective property, in a sense, replaces the heterogeneous material by a hypothetical homogeneous material and is a useful concept for a broad class of problems of interest. We shall be concerned with determining what must be known about the heterogeneous material in order to determine its effective property.

Complete knowledge of the functional $P[E_i(x), \epsilon_{lm}(x)]$, where $P[E_i(x), \epsilon_{lm}(x)]\, dE_1(x) \cdots dE_3(x) \times d\epsilon_{11}(x) \cdots d\epsilon_{33}(x)$ is defined as the probability of the realization of the particular joint field $[E_i(x), \epsilon_{lm}(x)]$, is required to determine the effective property of the material. Fortunately, bounds can be obtained for the effective property in terms of the volume fraction which is the simplest statistical information one can obtain for a material. In order to obtain better bounds, additional statistical information about the material is necessary. This additional statistical information may be introduced through the n-point correlation function, and bounds have been derived in terms of these functions.

Since virtually nothing is known about the higher-order correlation functions of heterogeneous material, this result has had limited utility. Hence, the problem is to determine how to extend the utility of the above result without having detailed knowledge of the higher-order correlation functions which are, in general, difficult or impossible to obtain.

Beran[1] derived bounds for the effective dielectric constant from a classical variational principle which included additional statistical information, viz., the three-point correlation function.

The development of improved bounds on effective properties depends on a knowledge of the three-point correlation function of the material. Several attempts have been made to approximate the form of the three-point correlation function with little success.

For a broad class of two-phase heterogeneous materials (which we shall call cell materials) we shall show that this problem is obviated, because the functions of the three-point correlation function that appear in the effective dielectric constant bounds[1] are simply a number for each phase. Furthermore, this number has a range of values $\frac{1}{9}$ to $\frac{1}{3}$ and a simple geometric significance. The number $\frac{1}{9}$ implies a cell of spherical shape, the number $\frac{1}{3}$ a cell of platelike shape, and all other cell shapes, no matter how irregular, have a corresponding number between.

Also, each value of this number determines a new set of bounds which is substantially narrower and always within the Hashin–Shtrikman bounds[2] (the best bounds in terms of volume fraction).

The new bounds for dilute suspensions are in most

[1] M. Beran, Nuovo Cimento **38**, 771 (1965).
[2] Z. Hashin and S. Shtrikman, J. Appl. Phys. **33**, 3125 (1962).

cases so narrow as to be essentially an exact solution. For other mixtures, we shall find new bounds which give substantial improvement in the knowledge of effective property bounds for two-phase media where the geometrical characteristics of each phase are known. Further, the above results can be obtained for all statistically isotropic and homogeneous dilute suspension independent of the assumption of a cell material.

The concepts developed here are applicable to multiphase media.

2. EFFECTIVE PERMITTIVITY BOUNDS

A. Introduction

In Beran,[1] by use of a perturbation expansion of the electric field **E** and the electrical displacement **D** as trial functions in two standard variational principles, the following bounds on the effective permittivity ϵ^* were found:

$$\epsilon^* \leq \langle\epsilon\rangle - \frac{1}{3}\frac{\langle\epsilon'^2\rangle}{\langle\epsilon\rangle}\Big/\left(1 + \frac{3\langle\epsilon\rangle I}{\langle\epsilon'^2\rangle}\right) \tag{2.1}$$

and

$$\epsilon^* \geq \left[\left\langle\frac{1}{\epsilon}\right\rangle - \left(\frac{4}{3}\left\langle\frac{\epsilon'}{\epsilon}\right\rangle\right)^2\frac{1}{4\langle\epsilon\rangle^2}\Big/\left(\tfrac{1}{3}\langle\epsilon\rangle^{-2}\left\langle\frac{\epsilon'^2}{\epsilon}\right\rangle + J\right)\right]^{-1}, \tag{2.2}$$

where

$$I = \frac{1}{16\pi^2}\langle\epsilon\rangle^{-2}\int_V\int_{V^1}\frac{\partial^2}{\partial r_3\partial s_3}\langle\epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}, \tag{2.3}$$

$$J = \frac{1}{16\pi^2}\langle\epsilon\rangle^{-2}\int_V\int_{V^1}\frac{\partial^2}{\partial r_3\partial s_3}\left\langle\frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)}\right\rangle\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}. \tag{2.4}$$

ϵ' is the fluctuating part of the permittivity and the bracket denotes an ensemble average, which is assumed to be equal to the spatial average for a statistically homogeneous medium. The above results are subject to the conditions of a medium that is statistically homogeneous and isotropic with a constant electric field.

From the above equations we see that the bounds on the effective permittivity are known in terms of certain averages of the ϵ field and functions of the three-point correlation functions $\langle\epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle$ and $\langle\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})/\epsilon(0)\rangle$. This same procedure can be used to express the bounds in terms of higher-order correlation functions. Beran[1] shows that the bounds in terms of the n-point correlation function will be at least as good as those in terms of the $(n-1)$-point correlation function with the expectation that, as $n \to \infty$, the upper and lower bounds will converge.

The same procedure is valid for determining effective properties for processes treating heat conductivity, electrical conductivity, magnetic permeability, and all other processes which obey the same type of equations.

The usefulness of the above-mentioned approach depends on the determination of I and J and similar derivatives and integrals of higher-order correlation functions. In subsequent sections we shall define a broad class of materials for which these functions are particularly simple and have obvious physical significance.

B. Determination of n-Point Correlation Functions for N-Phase Random Media

The n-point correlation function γ_n is defined as

$$\gamma_n = \langle\epsilon(\mathbf{r}_1)\epsilon(\mathbf{r}_2)\cdots\epsilon(\mathbf{r}_n)\rangle = \int\epsilon(\mathbf{r}_1)\epsilon(\mathbf{r}_2)\cdots\epsilon(\mathbf{r}_n)\,dF_{1\ldots n}(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_n), \tag{2.5}$$

where $F_{1\ldots n}$ is the n-point distribution function.

For a special class of random media, which we shall call an N-phase random medium, composed of N phases $D_1, D_2, \cdots, D_N$ of uniform density and of volume fractions $\varphi_1, \varphi_2, \cdots, \varphi_N$, respectively, a property function $\epsilon(\mathbf{r})$ may be defined for each phase as

$$\epsilon(\mathbf{r}) = \epsilon_i \quad \text{if} \quad \mathbf{r} \in D_i, \quad i = 1, 2, \cdots, N. \tag{2.6}$$

The n-point correlation function $\gamma_n(\mathbf{r}_1, \cdots, \mathbf{r}_n)$ for this process is related to the probability that n points $\mathbf{r}_1, \cdots, \mathbf{r}_n$, thrown at random into the medium, all lie in the same phase D_i. This is a generalization of the famous Buffon needle game (Kendall and Moran[3]). This analogy can be extended to the N-phase material by considering the n-point correlation function as the sum of the probabilities of the n points falling in all possible combinations of D_i $(i = 1, \cdots, N)$. Each term of the sum is multiplied by a weighting factor which is the product of the property function of the phase in which the points lie.

For brevity and clarity we shall restrict ourselves to the two-phase material, as all the arguments developed for it apply equally to the N-phase material. For the two-phase random medium the n-point correlation function is

$$\gamma_n(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_n) = \sum_{\alpha_n=1}^{2}\cdots\sum_{\alpha_1=1}^{2}P_{\alpha_1\cdots\alpha_n}(\mathbf{r}_1, \cdots, \mathbf{r}_n)\epsilon(\mathbf{r}_1)\cdots\epsilon(\mathbf{r}_n), \tag{2.7}$$

[3] M. G. Kendall and P. A. P. Moran, *Geometrical Probability* (Charles Griffin, London, 1963).

where $P_{1,\cdots,1,1}$ is the probability of n points being in phase D_1, $P_{1,\cdots,1,2}$ is the probability of $n-1$ points being in phase D_1 and point $\mathbf{r}_n$ being in phase D_2, $\cdots$, and $P_{2,\cdots,2,2}$ is the probability of n points in phase D_2.

If we consider each domain of our two-phase random medium to be composed of cells, where a cell is defined as being a mathematically closed surface containing a portion of the random medium of uniform property ϵ, we can write the n-point correlation function in an alternate manner. Instead of summing terms which are the probabilities of n points being in different combinations of the phase D_i $(i = 1, 2)$, multiplied by appropriate weighting factors, we shall sum over terms which are the probabilities of n points being in different combinations of cells, multiplied by weighting factors which are the products of the $\epsilon(r)$'s in whose cells the points lie. If we account for all possible combinations, the n-point correlation function becomes

$$\gamma_n(r_1, \cdots, r_n) = \epsilon_1^n \times (\text{probability of } n \text{ points being in the same cell with property } \epsilon_1) + \epsilon_1^{n-1}\epsilon_2 \times (\text{probability of } n-1 \text{ points being in the same cell with property } \epsilon_1 \text{ and 1 point being in a different cell with property } \epsilon_2) + \epsilon_1^n \times (\text{probability of } n-1 \text{ points being in the same cell with property } \epsilon_1 \text{ and 1 point being in a different cell with property } \epsilon_1) + \epsilon_1^{n-2}\epsilon_2^2 \times (\text{probability of } n-2 \text{ points being in the same cell with property } \epsilon_1 \text{ and 2 points being in another cell with property } \epsilon_2) + \cdots + \epsilon_2^n \times (\text{probability of } n \text{ points being in the same cell of property } \epsilon_2). \tag{2.8}$$

For $n = 1$ Eq. (2.8) becomes

$$\gamma_1(\mathbf{r}) = \epsilon_1 \times (\text{probability of 1 point being in a cell of property } \epsilon_1) + \epsilon_2 \times (\text{probability of 1 point being in a cell of property } \epsilon_2). \tag{2.9}$$

If we define φ as the volume fraction of material with property ϵ_1, Eq. (2.9) becomes

$$\gamma_1(\mathbf{r}) = \epsilon_1\varphi + \epsilon_2(1 - \varphi) = \langle\epsilon\rangle. \tag{2.10}$$

For $n = 2$, Eq. (2.8) becomes

$$\gamma_2(\mathbf{r}_1, \mathbf{r}_2) = \epsilon_1^2 f_1(\mathbf{r}_1, \mathbf{r}_2) + \epsilon_2^2 f_2(\mathbf{r}_1, \mathbf{r}_2) + \epsilon_1^2 Z_{11}(\mathbf{r}_1, \mathbf{r}_2) + \epsilon_2\epsilon_1[Z_{12}(\mathbf{r}_1\mathbf{r}_2) + Z_{21}(\mathbf{r}_1\mathbf{r}_2)] + \epsilon_2^2 Z_{22}(\mathbf{r}_1, \mathbf{r}_2), \tag{2.11}$$

where $f_n(\mathbf{r}_1, \mathbf{r}_2)$ is the probability of two points $(\mathbf{r}_1, \mathbf{r}_2)$ being in the same cell of property ϵ_n, $Z_{nm}(\mathbf{r}_1, \mathbf{r}_2)$ is the probability of one point $(\mathbf{r}_1)$ being in a cell of property ϵ_n and the other point $(\mathbf{r}_2)$ being in a different cell of property ϵ_m.

This method of writing the n-point correlation function can be extended to all n and the apparent complexity will vanish when we consider a slightly restricted class of materials in subsequent sections. We shall make particular use of the three-point correlation function expressed in this manner.

We are going to consider a broad class of heterogeneous materials called asymmetric cell materials, but for convenience we shall first consider a restricted class called symmetric cell materials.

C. Symmetric Cell Material

Consider a space to be subdivided by a large number of closed surfaces; these closed regions shall be called cells. The subdivision of the space is arbitrary except for fulfilling the following requirements:

(1) Space is completely covered by cells;

(2) cells are distributed in a manner such that the material is statistically homogeneous and isotropic;

(3) the material property ϵ of a cell is statistically independent of the material property of any other cell;

(4) the conditional probabilities of n points being and m points not being in the same cell of a particular material, given that one point is in a cell of that material, are the same for each material.

An example of such a material is the Poisson cell material (see Gilbert[4] and Frisch[5]). This material is constructed mathematically by distributing a pattern of points in space so that there is an equal probability of finding any point in an infinitesimal volume and the probability that a point is in an infinitesimal volume is proportional to the volume. Now each point is assigned the property ϵ_1 or ϵ_2 with probability φ or $1 - \varphi$ by an independent random process and the space between the points has the same property as the nearest point. The space is thus divided into convex polyhedron-shaped cells.

An infinite family of symmetric cell materials may be obtained by replacing the initial Poisson pattern of points by any other pattern with arbitrary correlation between points and following the same construction.

A second example of a symmetric cell material is a modification of the above model proposed by Johnson and Mehl.[6] They generalized the Poisson model by considering the distributed points to be nuclei

[4] E. N. Gilbert, Ann. Math. Statistics **33**, 958 (1962).
[5] H. L. Frisch, Trans. Soc. Rheol. **9**, 293 (1965).
[6] W. A. Johnson and R. F. Mehl, Trans. AIME **135**, 416 (1939).

from which a constant rate of cell growth begins at different times. Frisch[5] points out that, in two dimensions, one can visualize this model in terms of circular waves spreading from raindrops that fall at random into a puddle. The cells formed in this model no longer have plane or convex sides; instead they are star-shaped.

A third example of a symmetric cell material is a space packed with spheres of varying diameter. We assume then that if we allow spheres of all diameters, the space can be completely filled. The material property of each spherical cell is determined by an independent random process, where volume fraction φ of the cells have property ϵ_1 and $1-\varphi$ of the cells have property ϵ_2. Such a material fulfills the four requirements of a symmetric cell material and (for small concentrations) will be useful as a model for spherical inclusions in a matrix. We can generalize this model to include distributions of any shape or combination of shape cells which completely fill the space (e.g., ellipsoids, cubes, and tetrahedrons).

In general, any subdivision of the space is admissible which fulfills the first two requirements mentioned and for which the property of each cell is determined by an independent random process where $P(\epsilon_1)=\varphi$ and $P(\epsilon_2)=1-\varphi$. Therefore, the geometry of individual cells can be extremely different in a symmetric cell material. We call a material which satisfies the four requirements listed above a symmetric cell material.

For a symmetric cell material we show in Appendix A, using the approach discussed in Sec. 2.2, that the three-point correlation functions that appear in Eqs. (2.3) and (2.4) are

$$\gamma_{3I}=\langle\epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle=\epsilon_1'\frac{\varphi(1-2\varphi)}{(1-\varphi)^2}g(0,\mathbf{r},\mathbf{s}), \quad (2.12)$$

$$\gamma_{3J}=\left\langle\frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)}\right\rangle$$
$$=\frac{\epsilon_1'^2}{\epsilon_1}\frac{\varphi}{(1-\varphi)}[(2\varphi-1)(\alpha-1)g(0,\mathbf{r},\mathbf{s})$$
$$+(\alpha-\varphi(\alpha-1))f(\mathbf{r},\mathbf{s})], \quad (2.13)$$

where $\alpha=\epsilon_1/\epsilon_2\geq 1$, $g(0,\mathbf{r},\mathbf{s})$ is the conditional probability of a triangle (coordinates 0, $\mathbf{r}$, $\mathbf{s}$) having all three vertices in a single cell given one vertex in the cell, and $f(\mathbf{r},\mathbf{s})$ is the conditional probability of a line segment (coordinates $\mathbf{r}$, $\mathbf{s}$) having both ends in a single cell given one end in the cell.

Substituting Eqs. (2.12) and (2.13) into (2.3) and (2.4), respectively, we obtain

$$I=\frac{\epsilon_1'^3}{\langle\epsilon\rangle^2}\frac{\varphi(1-2\varphi)}{(1-\varphi)^2}G \quad (2.14)$$

and

$$J=\frac{\epsilon_1'^2\varphi}{\epsilon_1(1-\varphi)\langle\epsilon\rangle^2}[(2\varphi-1)(\alpha-1)G$$
$$+[\alpha-\varphi(\alpha-1)]F], \quad (2.15)$$

where

$$G=\frac{1}{16\pi^2}\int_V\int_{V^1}\frac{\partial^2 g(0,\mathbf{r},\mathbf{s})}{\partial r_3\partial s_3}\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}, \quad (2.16)$$

and for $f(\mathbf{r},\mathbf{s})=f(|\mathbf{r}-\mathbf{s}|)$ we can integrate F to obtain

$$F=\frac{1}{16\pi^2}\int_V\int_{V^1}\frac{\partial^2 f(\mathbf{r},\mathbf{s})}{\partial r_3\partial s_3}\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}=\tfrac{1}{3}. \quad (2.17)$$

From Eq. (2.16) we can see that G is a number which depends only on the geometry of the cells.

Substituting Eqs. (2.14) and (2.15) into (2.1) and (2.2), respectively, we obtain the following bounds for ϵ^*:

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}}\leq\frac{[1+\varphi(\alpha-1)]}{\alpha^{\frac{1}{2}}}\left[1-\frac{\varphi(\alpha-1)^2(1-\varphi)}{3[1+\varphi(\alpha-1)][1+\varphi(\alpha-1)+3(\alpha-1)(1-2\varphi)G]}\right] \quad (2.18)$$

and

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}}\geq\alpha^{\frac{1}{2}}\Big/\left[\alpha-\varphi(\alpha-1)-\frac{\frac{4}{3}(\alpha-1)^2(1-\varphi)\varphi}{1+\alpha+3(2\varphi-1)(\alpha-1)G}\right]. \quad (2.19)$$

Therefore the upper and lower bounds on ϵ^* for the symmetric cell material depend on α and a single number G which is characteristic of the average cell geometry.

For finite values of ϵ_1 and ϵ_2, we know that ϵ^* must be positive and finite. Brown[7] proved that ϵ^* must obey the more restrictive condition

$$1/\langle 1/\epsilon\rangle\leq\epsilon^*\leq\langle\epsilon\rangle \quad (2.20)$$

if the permittivity ϵ is considered a random function. Mathematically, we may express the former requirement as "the upper bound on ϵ^* must be positive and the lower bound on ϵ^* finite for all values of φ and α." In Appendix E we show that these restrictions on ϵ^*

[7] W. F. Brown, *Magnetostatic Principles in Ferromagnetism* (North-Holland Publ. Co., Amsterdam, 1962).

lead to the following bounds on G:

$$\tfrac{1}{9} \leq G \leq \tfrac{1}{3}. \tag{2.21}$$

We conclude that any value of G outside this range does not refer to a real cell geometry.

Furthermore, we find that when we substitute the value $G = \frac{1}{9}$ into the bounding equations (2.18) and (2.19), the two equations become asymptotic for small concentrations (i.e., φ approaches zero or unity). In Appendix F we show that this asymptotic solution is exactly the small concentration solution of spheres in a matrix. Similarly, putting $G = \frac{1}{3}$ into the bounding equations (2.18) and (2.19), the two equations are again asymptotic for small concentration; for this case the asymptotic solution is exactly the small concentration solution of randomly oriented plates in a matrix (see Appendix F). We therefore assign the geometric significance of a sphere to a G of $\frac{1}{9}$ and a plate to a G of $\frac{1}{3}$.

Hashin and Shtrikman[2] have shown that the best bounds on ϵ^* for a two-phase statistically homogeneous and isotropic material, when only φ is specified, are

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \leq \alpha^{\frac{1}{2}}\left[1 + \frac{3(1-\varphi)(1-\alpha)}{\varphi + \alpha(3-\varphi)}\right] \tag{2.22}$$

and

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \geq \frac{1}{\alpha^{\frac{1}{2}}}\left[1 + \frac{3\varphi(\alpha-1)}{3 + (1-\varphi)(\alpha-1)}\right]. \tag{2.23}$$

These bounds are the exact solution for a space filled with composite spheres where the upper bound represents a low-permittivity core enclosed in a high-permittivity shell and the lower bound represents the reverse situation. When the volume fraction of ϵ_1 is small, the lower bound is the case of a high-permittivity sphere in a low-permittivity matrix. For this case Eq. (2.23) reduces to Eqs. (2.18) and (2.19) with $G = \frac{1}{9}$, which confirms our conclusion that $G = \frac{1}{9}$ represents a spherical cell shape. Similarly, when the volume fraction of ϵ_1 approaches unity, the Hashin–Shtrikman upper bound is the case of a low-permittivity sphere in a high-permittivity matrix. For this case Eq. (2.21) is equal to Eq. (2.18) and (2.19) with $G = \frac{1}{9}$, reconfirming that $G = \frac{1}{9}$ represents spherical cell shapes. When the volume fraction of the material with property ϵ_1 is small (i.e., φ approaches zero), Eq. (2.21) reduces to Eqs. (2.18) and (2.19) with $G = \frac{1}{3}$. Similarly, when φ approaches unity, Eq. (2.23) reduces to Eqs. (2.18) and (2.19) with $G = \frac{1}{3}$.

In Appendix F we show that at small concentrations the new bounding equations are not restricted to symmetric cell materials, but hold for all two-phase statistically isotropic and homogeneous materials.

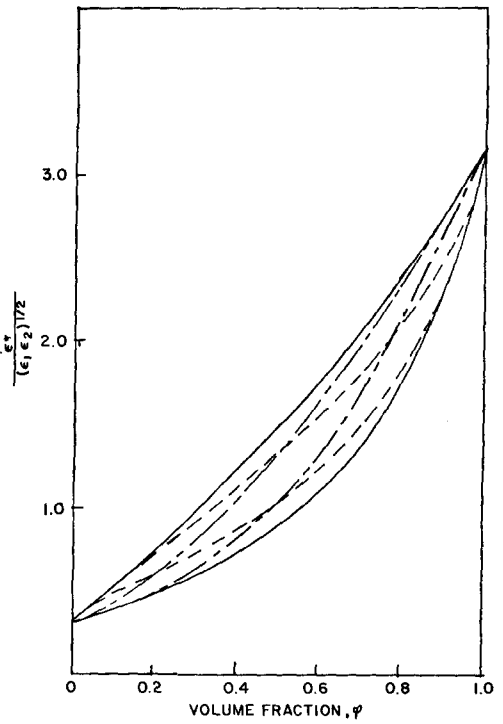

FIG. 1. Bounds on effective permittivity, $\alpha = 10$, symmetric cell material. ———, Hashin–Shtrikman; – – –, $G = \frac{1}{3}$; — — —, $G = \frac{1}{9}$.

For $\varphi = \frac{1}{2}$ both bound equations [viz., Eqs. (2.18) and (2.19)] are independent of G, hence independent of the cell's geometry. The symmetric cell model for $\varphi = \frac{1}{2}$ becomes a symmetric random medium which is defined as a random medium satisfying

$$P_{0\cdots0}(\mathbf{r}_1, \cdots, \mathbf{r}_n) = P_{1\cdots1}(\mathbf{r}_1, \cdots, \mathbf{r}_n) \quad \text{for} \quad n = 1, 2, \cdots.$$

For a symmetric random medium it can be shown that odd-order correlation functions may be expressed in terms of lower-order correlation functions (see Frisch[5]); specifically, for the three-point correlation function, we have

$$\gamma_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \tfrac{1}{2}[\gamma_2(\mathbf{r}_1, \mathbf{r}_2) + \gamma_2(\mathbf{r}_1, \mathbf{r}_3) + \gamma_2(\mathbf{r}_1, \mathbf{r}_3) - \tfrac{1}{2}]. \tag{2.24}$$

This equality for symmetric random media was used; the results were tabulated by Beran and Molyneux[8] as a solution for Eqs. (2.1) and (2.2). These results are equivalent to the result obtained by solving Eqs. (2.18) and (2.19) for $\varphi = \frac{1}{2}$.

In Fig. 1, we plot Eqs. (2.18) and (2.19) for $\alpha = 10$,

[8] M. Beran and J. Molyneux, Quart. Appl. Math. **24**, 107 (1966).

$G = \frac{1}{9}$, $G = \frac{1}{3}$, and the Hashin–Shtrikman bounds. We note that the new bounds offer a substantial improvement in the knowledge of ϵ^* for all values of φ. By "improvement" we mean the reduction in width of the new bounds as compared to the Hashin–Shtrikman bounds. There is an improvement of 50% for $\varphi = \frac{1}{2}$ and greater improvement at other values of φ. For the small-concentration case the uncertainty in ϵ^* is virtually eliminated. The extreme values of bounds for $G = \frac{1}{3}$ and $G = \frac{1}{9}$ represent the extreme bounds for all symmetric cell materials. For all other values of G the upper- and lower-bound curves fall inside these curves except at $\varphi = \frac{1}{2}$, where all upper-bound curves have the same ϵ^* and all lower-bound curves have the same ϵ^*. With the information that we have a symmetric cell material, we have an improvement over the Hashin–Shtrikman bounds. The maximum improvement occurs at $\varphi = \frac{1}{2}$; there is a lesser improvement at other φ's.

In Fig. 1 we see that the bounds on ϵ^* are narrower for spheres ($G = \frac{1}{9}$) than they are for plate ($G = \frac{1}{3}$) for all values of φ except, of course, $\varphi = \frac{1}{2}$, where all values of G have the same bounds. This is due to the fact that there is a single degree of freedom (i.e., the location of the sphere center) associated with $G = \frac{1}{9}$ as compared to the more than two degrees of freedom (i.e., plate center and the rotation of the plate) associated with a $G = \frac{1}{3}$. Consequently, the greater geometric-configuration uncertainty leads to a greater uncertainty in the effective-material property. We are unable to give a physical argument as to why this geometric uncertainty has no effect for $\varphi = \frac{1}{2}$, although we have shown mathematically that the bounds are independent of cell shape for $\varphi = \frac{1}{2}$ [see Eqs. (2.18) and (2.19)]. The contrast in bound spread is more dramatically shown in Fig. 2, where the bounds are plotted for $\alpha = 100$ and there is better than a 20% improvement for $\varphi = 0.5$ and better than a 60% improvement for $\varphi = 0.9$ for a G of $\frac{1}{9}$.

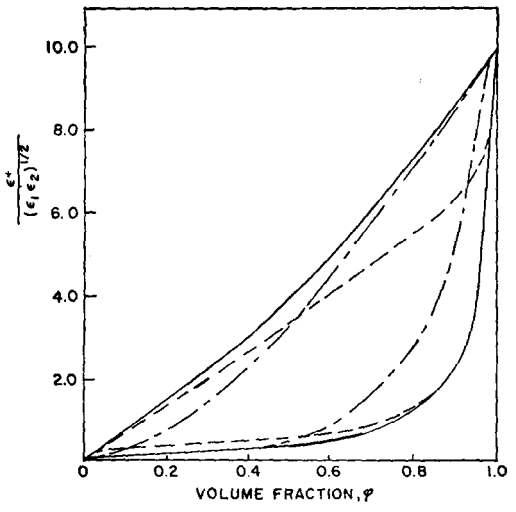

FIG. 2. Bounds on effective permittivity, $\alpha = 100$, symmetric cell material. ———, Hashin–Shtrikman; – – –, $G = \frac{1}{3}$; — — —, $G = \frac{1}{9}$.

For other values of G the bounds always fall within the extreme bounds of $G = \frac{1}{3}$ and $G = \frac{1}{9}$. At small concentrations the bounds for $G = \frac{1}{9}$ and $G = \frac{1}{3}$ are narrower than they are for any other value of G. The physical reason for this is that a G of $\frac{1}{3}$ or $\frac{1}{9}$ refers specifically to a single-cell geometry, while other G values refer to many-cell geometries. In essence, this adds another degree of freedom and a corresponding increase in uncertainty of the effective property. In Appendix F we confirm this by showing that the bounding equations become asymptotic at small concentrations only for G equal to $\frac{1}{9}$ and $\frac{1}{3}$.

D. Asymmetric Cell Material

In this section we shall investigate a class of materials for which the geometry of the cells of the two materials is dissimilar. Consider the space to be subdivided by a large number of closed surfaces, and the enclosed regions shall be called cells. The subdivision of the space is arbitrary except for fulfilling the following requirements:

(1) Space is completely covered by cells;

(2) cells are distributed in a manner such that the material is statistically homogeneous and isotropic;

(3) the material property ϵ of a cell is statistically independent of the material property of any other cell.

We call such a material an asymmetric cell material, and a model of such a material may be obtained by the following construction: Distribute a pattern of black and white points by the Poisson process discussed in Sec. 2.3, where there are $\varphi\rho$ black points and $(1 - \varphi)\rho$ white points per unit volume and ρ is the total density of points. The black points will be seeds from which spherical cells grow, and the white points are seeds from which aspherical cells grow. When cell surfaces from two different seeds touch, growth at that point ceases; this process continues until the entire space is filled. We may consider this model, a generalization of the Poisson cell material, to be one example of an asymmetric cell material. An infinite family of asymmetric cell materials may be obtained by replacing the initial Poisson pattern of points by any other pattern with arbitrary correlation between points and following the same construction. This model can be further generalized by varying the

arrival time of points; this is similar to the Johnson–Mehl model extension mentioned in Sec. 2.3. Since the color of the seed from which the cells grow are *a priori* statistically independent, therefore requirement (3) is satisfied.

Another example of an asymmetric cell material is a space packed with spheres and cubes of varying size. We assume that if we allow all sizes, the space can be completely filled for all volume fractions of spheres and cubes and that the packing can be carried out without preference as to whether a cube or sphere is being inserted at any given location. This model can be generalized to include many combinations of shape cells. Again requirement (3) is satisfied by the *a priori* random determination of cell shape at each location.

The basic difficulty in considering asymmetric, as opposed to symmetric, cell materials is in our ability to determine whether or not the material property is correlated to the cell's geometry. It will not always be possible to tell by observation whether such a correlation exists because contiguous cells may have the same property; therefore the boundary and the cell shape will not be determined by observation.

Using the approach discussed in Sec. 2.2 for the asymmetric cell material, we derive in Appendix A the following three-point correlation functions:

$$\gamma_{3I} = \langle \epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle = \epsilon_1'^3\left[\varphi g_1(0,\mathbf{r},\mathbf{s}) - \frac{\varphi^3}{(1-\varphi)^2}\, g_2(0,\mathbf{r},\mathbf{s})\right] \quad (2.25)$$

and

$$\gamma_{3J} = \left\langle \frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)} \right\rangle = \frac{\epsilon_1'^2}{\epsilon_1}\frac{\varphi}{(1-\varphi)}\Big[(\alpha-1)(\varphi^2 g_2(0,\mathbf{r},\mathbf{s})) + (1-\varphi)^2 g_1(0,\mathbf{r},\mathbf{s}) + (\varphi(1-\alpha)+\alpha)\frac{(1-\varphi)}{\varphi}\bar{f}_1^{11}(\mathbf{r},\mathbf{s}) + \frac{\varphi}{(1-\varphi)}\bar{f}_2^{11}(\mathbf{r},\mathbf{s})\Big], \quad (2.26)$$

where $g_n(0,\mathbf{r},\mathbf{s})$ is the conditional probability of a triangle $(0,\mathbf{r},\mathbf{s})$ having all three vertices in a single cell of material ϵ_n, given that one vertex is in the cell, and $f_n^{-11}(\mathbf{r},\mathbf{s})$ is the probability of a line segment $(\mathbf{r},\mathbf{s})$ having both ends in a single cell of material property ϵ_n.

Substituting Eqs. (2.25) and (2.26) into (2.3) and (2.4), we obtain

$$I = \frac{\varphi\epsilon_1'^3}{\langle\epsilon\rangle^2}\left[G_1 - \frac{\varphi^2}{(1-\varphi)^2}\, G_2\right] \quad (2.27)$$

and

$$J = \frac{\epsilon_1'^2}{\epsilon_1\langle\epsilon\rangle^2}\frac{\varphi}{(1-\varphi)}\Big[(\alpha-1)(\varphi^2 G_2 - (1-\varphi)^2 G_1) + (\alpha+\varphi(1-\alpha))\Big(\frac{(1-\varphi)}{\varphi}F_1 + \frac{\varphi}{(1-\varphi)}F_2\Big)\Big], \quad (2.28)$$

where we define

$$G_n = \frac{1}{16\pi^2}\int_V\int_{V'}\frac{\partial^2 g_n(0,\mathbf{r},\mathbf{s})}{\partial r_3 \partial s_3}\frac{r_i s_i}{r^3 s^3}\, d\mathbf{r}\, d\mathbf{s}, \quad (2.29)$$

$$F_n = \frac{1}{16\pi^2}\int_V\int_{V'}\frac{\partial^2 \bar{f}_n(\mathbf{r},\mathbf{s})}{\partial r_3 \partial s_3}\frac{r_i s_i}{r^3 s^3}\, d\mathbf{r}\, d\mathbf{s}. \quad (2.30)$$

By integration of Eq. (2.30), we find that $F_1 = \varphi/3$ and $F_2 = (1-\varphi)/3$. Following the same argument outlined in Sec. 2.3, we conclude that G_1 and G_2 are numbers that depend only on the geometries of the average cell of material property ϵ_1 and ϵ_2, respectively.

By use of Eqs. (2.27) and (2.28), we may recast (2.1) and (2.2) as

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left[1 + \varphi(\alpha-1) - \frac{\frac{1}{3}\varphi(1-\varphi)(\alpha-1)^2}{1+(\alpha-1)\{\varphi + 3[(1-\varphi)^2 G_1 - \varphi^2 G_2]\}}\right] \quad (2.31)$$

and

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \ge \alpha^{\frac{1}{2}}\Big/\left\{[\alpha - \varphi(\alpha-1)] - \frac{\frac{4}{3}(1-\alpha)^2(1-\varphi)\varphi}{1+\alpha+3(\alpha-1)[\varphi^2 G_2 - (1-\varphi)^2 G_1]}\right\}. \quad (2.32)$$

If G_1 and G_2 are equal, Eqs. (2.31) and (2.32) reduce to the symmetric cell material bounding equation, as is to be expected. In general, we see that the bounding equations now depend on a pair of numbers G_1 and G_2 which characterize the cell geometry of the two materials.

Using the same physical criteria as we did for the symmetric cell material, we can bound G_1 and G_2 so that

$$\tfrac{1}{9} \le G_1 \le \tfrac{1}{3}, \quad (2.33)$$

$$\tfrac{1}{9} \le G_2 \le \tfrac{1}{3}. \quad (2.34)$$

For all possible values of α, φ, G_1, and G_2, the bounding equations are inside the Hashin–Shtrikman bounds. The bounding Eqs. (2.31) and (2.32) attain

their highest upper bound and lowest lower bound for the combinations $G_1 = \frac{1}{3}, G_2 = \frac{1}{9}$ and $G_1 = \frac{1}{9}, G_3 = \frac{1}{3}$, respectively. Hence, these combinations constitute absolute bounds for asymmetric cell materials; since the Hashin–Shtrikman bounds refer to a real physical geometry and fall outside these bounds, an explanation is necessary. The apparent discrepancy is due to the fact that the specific geometry to which the Hashin–Shtrikman bounds refer is concentric spheres, which are part of a class of materials which violate our assumption that the material property of a cell be statistically independent of the material property of any other cell.

For φ approaching zero the new bounding equations converge to the Hashin–Shtrikman upper bound when $G_1 = \frac{1}{3}$ and to the Hashin–Shtrikman lower bound when $G_1 = \frac{1}{9}$. When φ approaches unity, the same convergence occurs with $G_2 = \frac{1}{9}$ and $G_2 = \frac{1}{3}$, respectively. In Appendix F we show that the new bounding equations are not restricted to asymmetric cell materials, but hold for all two-phase statistically homogeneous and isotropic materials.

In Figs. 3 and 4 we plot the new bounds for the combinations $G_1 = \frac{1}{9}, G_2 = \frac{1}{3}$ and $G_2 = \frac{1}{9}, G_1 = \frac{1}{3}$, and the Hashin–Shtrikman bounds for $\alpha = 10$ and $\alpha = 100$. We see that the shape of the bounds found for the symmetric cell material is replaced by a shape

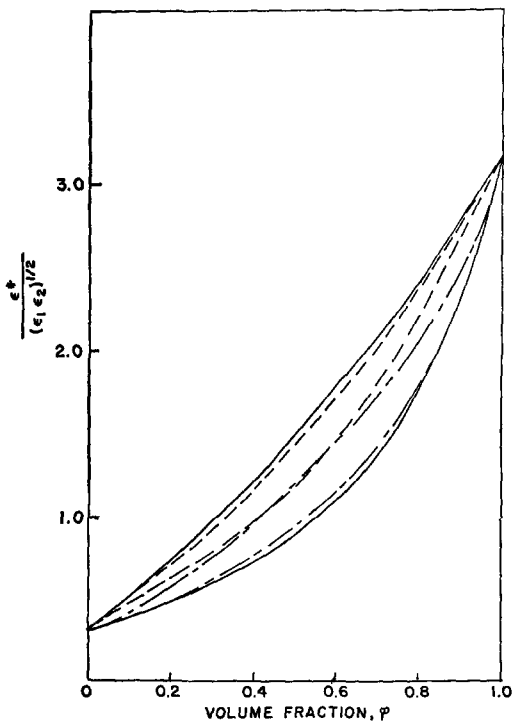

FIG. 3. Bounds on effective permittivity, $\alpha = 10$, asymmetric cell material. ———, Hashin–Shtrikman; – – – $(G_1 = \frac{1}{3}, G_2 = \frac{1}{9})$; — — — $(G_1 = \frac{1}{9}, G_2 = \frac{1}{3})$.

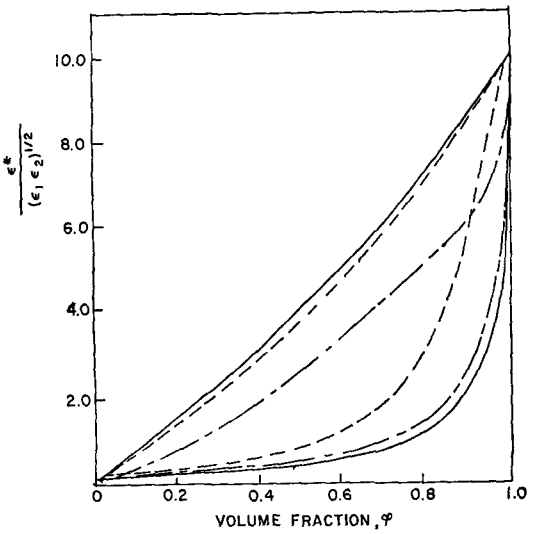

FIG. 4. Bounds on effective permittivity, $\alpha = 100$, asymmetric cell material. ———, Hashin–Shtrikman; – – – $(G_1 = \frac{1}{3}, G_2 = \frac{1}{9})$; — — — $(G_1 = \frac{1}{9}, G_2 = \frac{1}{3})$.

that is similar to the shape of the Hashin–Shtrikman bounds. A comparison of the improvement in resolution shows that the improvement decreases with α and is best at small concentrations (i.e., φ approaches zero or unity). The improvement at $\varphi = 0.5$ for $\alpha = 2$ is 80%, for $\alpha = 10$ is 45%, and for $\alpha = 50$ is 20%.

In the limit α approaches infinity, we compare the absolute bounds for the asymmetric cell material with the Hashin–Shtrikman bounds.

	Upper bound	Lower bound
Asymmetric cell material	$\frac{2\varphi\alpha^{\frac{1}{2}}}{3}\frac{(1-\varphi+\varphi^2)}{(1-\varphi+\frac{2}{3}\varphi^2)}$	$\frac{\alpha^{-\frac{1}{2}}}{(1-\varphi)}\frac{(1+\varphi+\varphi^2)}{(1-\varphi+\varphi^2)}$
Hashin–Shtrikman	$\frac{2\varphi\alpha^{\frac{1}{2}}}{3-\varphi}$	$\frac{\alpha^{-\frac{1}{2}}}{(1-\varphi)}(1+2\varphi)$

For all values of $\alpha > 1$ the asymmetric cell material bounds are inside the Hashin–Shtrikman bounds.

E. Determination of G_1 and G_2

In the preceding sections we have shown that improved bounds for the effective permittivity may be determined if a single number G is known for a symmetric cell material and a pair of numbers G_1 and G_2 is known for an asymmetric cell material. We shall discuss here a number of alternate ways of determining these G's for real materials which fall within our two bound classifications.

First, we already know that spherically shaped cells have a G of $\frac{1}{9}$ and plate-shaped cells have a G of $\frac{1}{3}$. If observation of the material cross section reveals

either or both of these cell shapes, we know the corresponding G. This method of observation may be extended to cell shapes whose G is determined by any of the subsequent methods.

Although we have shown that the bounds converge at small concentrations only for G's of $\frac{1}{3}$ and $\frac{1}{9}$, the bounds also converge when we consider small perturbations (i.e., α approaches unity). This convergence is expected because Beran[1] used the small perturbation solution as the admissible function in the variational principles to determine the bounding equations (2.1) and (2.2). If we let φ approach zero or unity, let α approach unity in the bounding equations, and set the coincident solution equal to the small concentration solution for any shape inclusion, we can solve for G. Using the small concentration solutions summarized by Reynolds and Hough,[9] we find we can solve for the G of all randomly oriented spheroids. We find that for needles $G = \frac{1}{6}$, oblate spheroids $\frac{1}{9} < G < \frac{1}{6}$, and prolate spheroids $\frac{1}{6} < G < \frac{1}{3}$ (see Appendix F).

If it is possible to construct a small-concentration sample of the composite of interest without changing the cell shape, bounds may be found on G by measuring the effective property of the sample and putting this value in the bounding equations. Since the accuracy of this method increases with decreasing α, it might be advantageous under certain circumstances to measure the effective property for which α is the smallest (i.e., although the α for dielectric constant is high, the α for thermal conductivity might be small). This must be tempered by the knowledge of which effective property may be most accurately measured and the difficulty of making measurements for small α. This method is applicable for determining the value of G for irregular cell shapes.

G_1 and G_2 may also be determined from their defining equation (2.29). In order to determine $g_1(0, \mathbf{r}, \mathbf{s})$ and $g_2(0, \mathbf{r}, \mathbf{s})$, the following mathematical experiment must be performed. The experiment consists of measuring the frequency with which the vertices of a triangle of coordinates $(0, \mathbf{r}, \mathbf{s})$, dropped randomly on a cross section of the composite, fall within a single cell. This experiment must be repeated for various triangle coordinates. Corson[10] has performed an experiment similar to this, in which he measures the frequency with which all three vertices of the triangle fall in any cell having the same property, for composites of sintered metals. The experiment for determining $g_1(0, \mathbf{r}, \mathbf{s})$ and $g_2(0, \mathbf{r}, \mathbf{s})$ may also be conducted on a computer by characterizing the cell boundaries by a coordinate address and the location and orientation of the randomly thrown triangle by random numbers. A criterion can be established for determining whether or not the triangle vertices are inside a cell; then the determination of $g_1(0, \mathbf{r}, \mathbf{s})$ and G follows directly.

[9] J. A. Reynolds and J. M. Hough, Proc. Phys. Soc. (London) **B70**, 769 (1957).

[10] Experiment being conducted at the Towne School, University of Pennsylvania, as part of Ph.D. thesis by P. B. Corson.

F. Higher-Order Correlation Functions

We have seen in the preceding sections that the three-point correlation function provides information about the geometry of the cells which compose the material, and this information reduces our uncertainty concerning the effective property of the composite. For a fixed volume fraction we note that platelike cells in a matrix have higher effective properties than spherelike cells. We reason that the higher effective property is related to the greater tendency of a fixed volume fraction of plates to form a continuous or nearly continuous path through the material than spheres. Thus the cell geometry that tends to form a continuous path is more strongly felt and leads to an effective property near the Hashin–Shtrikman upper bound for $\varphi < \frac{1}{2}$. Based on these observations, we expect higher-order correlation functions to contribute geometrical information concerning the tendency of cells to form paths through the composite.

On the above supposition, it might be possible to devise a systematized approach to determine the degree to which a particular cell arrangement tends to form a continuous path. From this determination we could determine whether an estimate of the effective property should be made near the upper or the lower bound for a particular set of G's.

G. Empirical Validation

A large number of experimental studies of the effective properties of two-phase materials are available in the literature.[11–15] A relation which Landauer has given for a mixture of materials where all the regions are spherical is

$$\varphi(\epsilon_1 - \epsilon^*)/(\epsilon_1 + 2\epsilon^*) + (1 - \varphi)(\epsilon_2 - \epsilon^*)/(\epsilon_2 + 2\epsilon^*\varphi) = 0. \quad (2.35)$$

Landauer found this expression to be a good approximation to most of the experimental data he analyzed

[11] R. Landauer, J. Appl. Phys. **23**, 779 (1952).

[12] G. P. DeLoor, "Dielectric Properties of Heterogeneous Mixtures," Ph.D. thesis, University of Leyden, Leyden, 1956.

[13] A. Sugawara and Y. Yoshizawa, J. Appl. Phys. **33**, 3135 (1962).

[14] C. Herring, J. Appl. Phys. **31**, 1939 (1960).

[15] W. Woodside and J. H. Messmer, J. Appl. Phys. **32**, 1688 (1961).

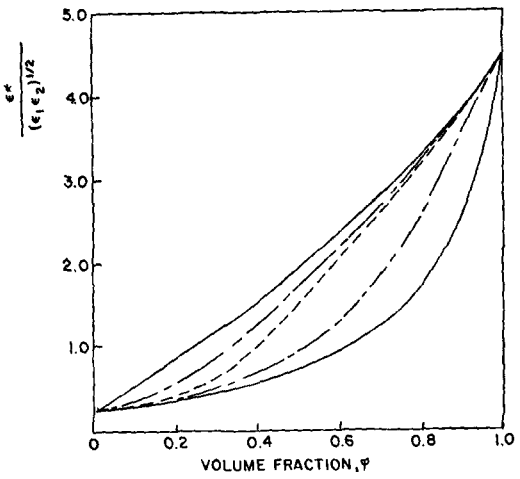

FIG. 5. Comparison of Landauer equation with bounds, $\alpha = 20$. – – –, Landauer equation; ———, Hashin–Shtrikman; — — —, $G = \frac{1}{9}$.

(e.g., Cd-Pb, Matthiesson; Cu-Fe, Ruer and Fick; CuSb-Sb, Stephens; Bi-Bi_2Pb, Herod). In Fig. 5 we see that Eq. (2.35) always falls inside the bounding equation for spheres (i.e., $G_1 = G_2 = \frac{1}{9}$); thus all these experimental data verify the assumption of spherical cells.

For dilute suspensions of regularly shaped particles we have shown that the bounding equations converge to the exact solutions; hence these results are verified by all dilute suspension measurements of regularly shaped particles that agree with the previously derived exact solutions.

For the other experimental studies it is difficult to make comparisons because the shape factors G_1 and G_2 are unknown for the materials tested.

3. SIGNIFICANCE OF G

Improved bounds for the effective permittivity are determined if a single number G is known for a symmetric cell material and a pair of numbers G_1 and G_2 are known for an asymmetric cell material. The concept of this number G is linked to the cell material from which it was derived; hence, to understand its usefulness, we must examine the limitations of the cell-material model. The fundamental assumption required to define the cell-material model used in this paper was that the property of a cell be independent of the property of any other cell. The question then is, How does this requirement limit the physical models which can be represented by the cell material?

First, let us consider the case of a dilute suspension. For this case, there is no problem since we can consider the suspension materials to be cells of a particular distribution of sizes and shapes and the matrix material to be made up of cells of all sizes and shapes. For such a combination of materials, we can pack a space without preference regarding the type of cell which is inserted at any location. The cells of the suspension are far apart so their shape will not interfere with the packing procedure, and there is so much freedom in the size and shape of the matrix material cells that we are certain that the space can be completely filled without violating the fundamental assumption. Since the bounds on effective properties for the dilute suspension are independent of the G value of the matrix cells, the cell-material model is applicable to dilute suspensions without difficulty. This even applies when the actual construction of the material violates the fundamental assumption, i.e., suspended particles tend to cluster together, since we can consider the clustered particles to be a single cell of a different geometry. Since we have left so much freedom in the cell construction of the matrix, we will not violate the assumption of the cell-material model. This will be true even when the clustered particles are not contiguous since we can make them mathematically contiguous with appropriate imaginary slices through the matrix material. This is further verified by the fact that we can derive the bound equations for dilute suspensions without the cell-model assumption.

Let us attempt to generalize this model to the finite suspension. Since the suspension particles are close to each other, we see that a problem can arise when we attempt to pack an empty space without regard as to which type cell is inserted; i.e., assume the suspension is made up of spheres of the same diameter which will not fit at certain locations in the packing procedure. On the other hand, if we allow the suspension spheres to be any diameter, we can still fill the space without violating the fundamental assumption. Thus we see that, for the finite suspension, there are certain restrictions on the size and shape of the distribution of cells which will satisfy the assumptions of the cell-material model. What happens when we have a distribution of suspended particle sizes and shapes that violates the fundamental assumption? It was never assumed in the development of the cell-material model that the cell shape and size distributions be the same as particle shape and size distribution; therefore, the cell-material model is so general that cells of some value G will be a good approximation to any finite suspension. The problem, however, is that in this case observation might lead to a false conclusion as to which G is appropriate, although it is doubtful that the new G will vary much from the

G that represents the particle shape. Since the bounds on effective properties of finite suspensions depend on the value of G_2 of the matrix cells (whose shape and size were left arbitrary to assure filling the space without violating the fundamental assumption), the bounds will be those defined by G_1 of the suspended particles and the values of G_2 (in the range $\frac{1}{9}$ to $\frac{1}{3}$) which give the highest upper and lowest lower bound. These bounds will still result in a substantial improvement over the Hashin–Shtrikman bounds for volume fractions of up to approximately 0.6.

For cell-like materials (e.g., sintered materials and eutectic alloys), the cell model is a natural representation, and the G values will represent the geometries of the component cells.

We can not represent every mixture by a cell-material model, as evidenced by the fact that the Hashin–Shtrikman bounds which refer to concentric spheres fall outside the range of our bounds. But even in this case, which clearly violates the fundamental assumption of the cell-material model, the Hashin–Shtrikman bounds only deviate slightly from the extreme bounds ($G_1 = \frac{1}{3}$, $G_2 = \frac{1}{9}$ and $G_1 = \frac{1}{9}$, $G_2 = \frac{1}{3}$) for all values of volume fraction and the cell model is a fair approximation.

APPENDIX A: DERIVATION OF THREE-POINT CORRELATION FUNCTIONS

The two three-point correlation functions defined in Sec. 2 are

$$\gamma_{3I} \equiv \langle \epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s}) \rangle \tag{A1}$$

and

$$\gamma_{3J} \equiv \left\langle \frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)} \right\rangle. \tag{A2}$$

For a two-phase cell-like material we can write Eq. (A1) in the alternate form

$$\begin{aligned}\gamma_{3I} = {} & \epsilon_1'^3\bar{g}_1 + \epsilon_2'^3\bar{g}_2 + \epsilon_1'^3[\bar{h}_{11} + \bar{h}_{11}^1 + \bar{h}_{11}^1] \\ & + \epsilon_1'^2\epsilon_2'[\bar{h}_{12} + \bar{h}_{12}^1 + \bar{h}_{12}^{11}] \\ & + \epsilon_1'\epsilon_2'^2[\bar{h}_{21}^{11} + \bar{h}_{21}^1 + \bar{h}_{21}] + \epsilon_2'^3[\bar{h}_{22} + \bar{h}_{22}^1 + \bar{h}_{22}^{11}] \\ & + \epsilon_1'^3Z_{111} + \epsilon_1'^2\epsilon_2'[Z_{112} + Z_{121} + Z_{211}] \\ & + \epsilon_1'\epsilon_2'^2[Z_{112} + Z_{212} + Z_{221}] + \epsilon_2'^3Z_{222},\end{aligned} \tag{A3}$$

where

$$\epsilon_1' = \epsilon_1 - \langle\epsilon\rangle, \quad \epsilon_2' = \epsilon_2 - \langle\epsilon\rangle,$$

$\bar{g}_n = \bar{g}_n(0, \mathbf{r}, \mathbf{s})$ is the probability that three points $(0, \mathbf{r}, \mathbf{s})$ are in the same cell with material property ϵ_n;

$\bar{h}_{nm} = \bar{h}_{nm}(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(0, \mathbf{r})$ are in the same cell with material property ϵ_n and a point $(\mathbf{s})$ is in another cell with material property ϵ_m;

$\bar{h}_{nm}^1 = \bar{h}_{nm}^1(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(0, \mathbf{s})$ are in the same cell with material property ϵ_n and a point $(\mathbf{r})$ is in another cell with material property ϵ_m;

$\bar{h}_{nm}^{11} = \bar{h}_{nm}^{11}(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(\mathbf{r}, \mathbf{s})$ are in the same cell with material property ϵ_n and a point (0) is in another cell with material property ϵ_m;

$Z_{nmp} = Z_{nmp}(0, \mathbf{r}, \mathbf{s})$ is the probability that three points are in three different cells, where points 0, $\mathbf{r}$, and $\mathbf{s}$ are in cells with material properties ϵ_n, ϵ_m, and ϵ_p, respectively.

Also

$$\bar{g}_1 = \varphi g_1, \quad \bar{g}_2 = (1 - \varphi)g_2, \tag{A4}$$

where

$g_n = g_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that all three points are in the same cell of material property ϵ_n, given that one of the points is in a cell with material property ϵ_n.

We assume

$$\begin{aligned}
&\bar{h}_{11} = \varphi^2 h_1, && \bar{h}_{12} = \varphi(1 - \varphi)h_1, \\
&\bar{h}_{21} = (1 - \varphi)\varphi h_2, && \bar{h}_{22} = (1 - \varphi)^2 h_2, \\
&\bar{h}_{11}^1 = \varphi^2 h_1^1, && \bar{h}_{12}^1 = \varphi(1 - \varphi)h_1^1, \\
&\bar{h}_{21}^1 = (1 - \varphi)\varphi h_2^1, && \bar{h}_{22}^1 = (1 - \varphi)^2 h_2^1, \\
&\bar{h}_{11}^{11} = \varphi^2 h_1^{11}, && \bar{h}_{12}^{11} = \varphi(1 - \varphi)h_1^{11}, \\
&\bar{h}_{21}^1 = (1 - \varphi)\varphi h_2^{11}, && \bar{h}_{22}^1 = (1 - \varphi)^2 h_2^{11}.
\end{aligned} \tag{A5}$$

This assumption implies that the material property of a cell is statistically independent of the material property of any other cell.

Also we assume

$$\begin{aligned}
&Z_{111} = \varphi^3 Z, \quad Z_{112} = Z_{211} = Z_{121} = \varphi^2(1 - \varphi)Z, \\
&Z_{122} = Z_{212} = Z_{221} = \varphi(1 - \varphi)^2 Z, \quad Z_{222} = (1 - \varphi)^3 Z,
\end{aligned} \tag{A6}$$

where

$h_n = h_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(0, \mathbf{r})$ are in the same cell and point $(\mathbf{s})$ is in another cell, given that one point (0 or $\mathbf{r}$) is in a cell of property ϵ_n;

$h_n^1 = h_n^1(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(0, \mathbf{r})$ are in the same cell and point $(\mathbf{r})$ is in another cell, given that one point (0 or $\mathbf{s}$) is in a cell of property ϵ_n;

$h_n^{11} = h_n^{11}(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(\mathbf{r}, \mathbf{s})$ are in the same cell and point (0) is in another cell, given that one point ($\mathbf{r}$ or $\mathbf{s}$) is in a cell of property ϵ_n;

$Z = Z(0, \mathbf{r}, \mathbf{s})$ is the probability that three points are in different cells.

By taking into account all possible ways that three points can fall into a cell material, we have

$$Z = 1 - \bar{g}_1 - \bar{g}_2 - \bar{h}_{11} - \bar{h}_{11}^1 - \bar{h}_{11}^{11} - \bar{h}_{12} - \bar{h}_{12}^1 - \bar{h}_{12}^{11} - \bar{h}_{21} - \bar{h}_{21}^1 - \bar{h}_{21}^{11} - \bar{h}_{22} - \bar{h}_{22}^1 - \bar{h}_{22}^{11} \quad \text{(A7)}$$

or

$$Z = 1 - \varphi g_1 - (1 - \varphi)g_2 - \varphi(h_1 + h_1^1 + h_1^{11}) - (1 - \varphi)(h_2 + h_2^1 + h_2^{11}). \quad \text{(A8)}$$

Substituting into Eq. (A3), we obtain

$$\begin{aligned}\gamma_{3I} = {} & \epsilon_1'^3 \varphi g_1 + \epsilon_2'^3(1 - \varphi)g_2 \\ & + (\epsilon_1'^3\varphi^2 + \epsilon_1'^2\epsilon_2'\varphi(1 - \varphi))(h_1 + h_1^1 + h_1^{11}) \\ & + (\epsilon_1'\epsilon_2'^2\varphi(1 - \varphi) + \epsilon_2'^3(1 - \varphi)^2)(h_2 + h_2^1 + h_2^{11}) \\ & + (\epsilon_1'^3\varphi^3 + 3\epsilon_1'^2\epsilon_2'\varphi^2(1 - \varphi) + 3\epsilon_1'\epsilon_2'^2\varphi(1 - \varphi)^2 \\ & + \epsilon_2'^3(1 - \varphi)^3)(1 - \varphi g_1 - (1 - \varphi)g_2 \\ & - \varphi(h_1 + h_1^1 + h_1^{11}) - (1 - \varphi)(h_2 + h_2^1 + h_2^{11})).\end{aligned} \quad \text{(A9)}$$

Now

$$\begin{aligned} f_n(\mathbf{r}) &= g_n + h_n, \\ f_n^1(\mathbf{s}) &= g_n + h_n^1, \\ f_n^{11}(\mathbf{r}, \mathbf{s}) &= g_n + h_n^{11}, \end{aligned} \quad \text{(A10)}$$

where $f_n(\mathbf{r})$ is the conditional probability that 2 points $(0, \mathbf{r})$ are in the same cell with material property ϵ_n, given that one point is in a cell with material property ϵ_n; $f_n^1(\mathbf{r})$ and $f_n^{11}(|\mathbf{r} \cdot \mathbf{s}|)$ are defined in a similar way, except they refer to points $(0, \mathbf{s})$ and $(\mathbf{r}, \mathbf{s})$, respectively.

Substituting for the h's in Eq. (A9) and using the equality $\varphi\epsilon_1' = -(1 - \varphi)\epsilon_2'$, we obtain for any cell material

$$\gamma_{3I} = \epsilon_1'^3\varphi\left[g_1 - \frac{\varphi^2}{(1 - \varphi)^2} g_2\right]. \quad \text{(A11)}$$

If we assume a symmetric cell material ($g_1 = g_2 = g$), we obtain

$$\gamma_{3I} = \epsilon_1'^3 \frac{\varphi(1 - 2\varphi)}{(1 - \varphi)^2} g. \quad \text{(A12)}$$

Using the same approach to derive γ_{3J}, we can write Eq. (A2) in the alternate form

$$\begin{aligned}\gamma_{3J} = {} & \frac{\epsilon_1'^2}{\epsilon_1}\bar{g}_1 + \frac{\epsilon_2'^2}{\epsilon_2}\bar{g}_2 + \frac{\epsilon_1'^2}{\epsilon_1}(\bar{h}_{11} + \bar{h}_{11} + \bar{h}_{11}) \\ & + \frac{\epsilon_1'\epsilon_2'}{\epsilon_1}(\bar{h}_{12} + \bar{h}_{12}) + \frac{\epsilon_1'\epsilon_1'}{\epsilon_2}\bar{h}_{12}^{11} + \frac{\epsilon_2'\epsilon_1'}{\epsilon_2}(\bar{h}_{21} + \bar{h}_{21}^1) \\ & + \frac{\epsilon_2'\epsilon_2'}{\epsilon_1}\bar{h}_{21} + \frac{\epsilon_2'^2}{\epsilon_2}(\bar{h}_{22} + \bar{h}_{22}^1 + \bar{h}_{22}^{11}) \\ & + \frac{\epsilon_1'^2}{\epsilon_1}\bar{Z}_{111} + \frac{\epsilon_1'^2}{\epsilon_2}\bar{Z}_{211} + \frac{\epsilon_1'\epsilon_2'}{\epsilon_1}(\bar{Z}_{121} + \bar{Z}_{112}) \\ & + \frac{\epsilon_1'\epsilon_2'}{\epsilon_2}(\bar{Z}_{212} + \bar{Z}_{221}) + \frac{\epsilon_2'^2}{\epsilon_1}\bar{Z}_{122} + \frac{\epsilon_2'^2}{\epsilon_2}\bar{Z}_{222}.\end{aligned} \quad \text{(A13)}$$

By use of the relations developed before, we rewrite Eq. (A13) as

$$\begin{aligned}\gamma_{3J} = {} & \frac{\epsilon_1'^2}{\epsilon_1}\varphi g_1 + \frac{\epsilon_2'^2}{\epsilon_2}(1 - \varphi)g_2 \\ & + \left(\frac{\epsilon_1'^2}{\epsilon_1}\varphi^2 + \frac{\epsilon_1'\epsilon_2'}{\epsilon_1}\varphi(1 - \varphi)\right)(h_1 + h_1') \\ & + \left(\frac{\epsilon_1'^2}{\epsilon_1}\varphi^2 + \frac{\epsilon_1'^2}{\epsilon_1}\varphi(1 - \varphi)\right)h_1^{11} \\ & + \left(\frac{\epsilon_2'\epsilon_1'}{\epsilon_2}\varphi(1 - \varphi) + \frac{\epsilon_2'^2}{\epsilon_2}(1 - \varphi)^2\right)(h_2 + h_2') \\ & + \left(\frac{\epsilon_2'^2}{\epsilon_1}\varphi(1 - \varphi) + \frac{\epsilon_2'^2}{\epsilon_2}(1 - \varphi)^2\right)h_2^{11} \\ & + \left(\frac{\epsilon_1'^2}{\epsilon_1}\varphi^3 + \frac{\epsilon_1'^2}{\epsilon_2}\varphi^2(1 - \varphi) + 2\frac{\epsilon_1'\epsilon_2'}{\epsilon_1}\varphi^2(1 - \varphi)\right. \\ & + 2\frac{\epsilon_1'\epsilon_2'}{\epsilon_2}\varphi(1 - \varphi)^2 + \frac{\epsilon_2'^2}{\epsilon_1}\varphi(1 - \varphi)^2 \\ & \left. + \frac{\epsilon_2'^2}{\epsilon_2}(1 - \varphi)^3\right)(1 - \varphi g_1 - (1 - \varphi)g_2 \\ & - \varphi(h_1 + h_1^1 + h_1^{11}) - (1 - \varphi)(h_2 + h_2^1 + h_2^{11})).\end{aligned} \quad \text{(A14)}$$

Substituting Eq. (A10) into Eq. (A14) and using the equality $\epsilon_1'\varphi = -\epsilon_2'(1 - \varphi)$, we obtain

$$\gamma_{3J} = \frac{\epsilon_1'^2}{\epsilon_1}\frac{\varphi}{(1 - \varphi)}[(\alpha - 1)(\varphi^2 g_2 - (1 - \varphi)^2 g_1) + (\alpha + \varphi(1 - \alpha))((1 - \varphi)f_1^{11} + f_2^{11})]. \quad \text{(A15)}$$

For a symmetric cell material, i.e., $g_1 = g_2 = g$ and $f_1^{11} = f_2^{11} = f$, Eq. (A15) reduces to

$$\gamma_{3J} = \frac{\epsilon_1'^2}{\epsilon_1} \frac{\varphi}{(1-\varphi)} \times [(\alpha - 1)(2\varphi - 1)g + (\alpha + \varphi(1-\alpha))f^{11}]. \tag{A16}$$

APPENDIX B: PROOF $F = \frac{1}{3}$

F is defined as

$$F = \frac{1}{16\pi^2} \int_V \int_{V'} \frac{\partial^2 f(y)}{\partial r_3 s_3} \frac{r_i s_i}{r^3 s^3} \, d\mathbf{r} \, d\mathbf{s}, \tag{B1}$$

where

$$y^2 = r^2 + s^2 + 2rsu,$$
$$u = \cos\theta,$$
$$f(0) = 1, \quad f(\infty) = 0.$$

Choose s_i to lie along x_3 axis; then

$$F = \frac{1}{16\pi^2} \int_V \int_{V'} \frac{\partial}{\partial s} \left(\frac{df(y)}{dy}\right) \left(\frac{ru - s}{y}\right) \frac{u}{r^2 s^2} \, d\mathbf{r} \, d\mathbf{s}. \tag{B2}$$

Change from u and r coordinates to u' and y coordinates:

$$u = \frac{s + yu'}{r}, \quad r^2 = y^2 + s^2 + 2syu'.$$

Then

$$F = \frac{1}{16\pi^2} \iiint \frac{1}{s^2} \frac{d}{dy} \left(\frac{\partial f(y)}{\partial s}\right) K y^2 \, dy \, d\phi \, d\mathbf{s}, \tag{B3}$$

where

$$K(s, y) = \int_{+1}^{-1} \frac{u'(s + yu')}{(y^2 + s^2 + 2ysu')^{\frac{3}{2}}} \, du' = -\frac{2}{3y^2}.$$

Integration with respect to ϕ and y yields

$$F = \frac{1}{3(4\pi)} \int_V \frac{1}{s^2} \frac{df(s)}{ds} \, d\mathbf{s}. \tag{B4}$$

Finally,

$$F = \tfrac{1}{3}. \tag{B5}$$

Similarly, if in Eq. (B1) $f(y) = f_n(y)$, where $f_1(0) = \varphi$ and $f_2(0) = 1 - \varphi$, we obtain

$$F_1 = \tfrac{1}{3}\varphi, \tag{B6}$$

$$F_2 = \tfrac{1}{3}(1 - \varphi). \tag{B7}$$

APPENDIX C: PROOF G DEPENDS ON CELL GEOMETRY, BUT NOT THE DISTRIBUTION OF CELL DIMENSIONS OF EACH GEOMETRY

By definition,

$$G = \frac{1}{16\pi^2} \int_v \int_{v^1} \frac{\partial^2 g(0, \mathbf{r}, \mathbf{s})}{\partial r_3 \partial s_3} \frac{r_m s_m}{r^3 s^3} \, d\mathbf{r} \, d\mathbf{s}. \tag{C1}$$

For a discrete number N of cell shapes,

$$g(0, \mathbf{r}, \mathbf{s}) = \sum_{i=1}^{N} g_i(0, \mathbf{r}, \mathbf{s}) k_i, \tag{C2}$$

where k_i is the frequency of the ith cell shape; g_i is the probability of three points being in the same cell of the ith shape, given that one of the points is in a cell of the ith shape.

Also

$$G = \sum_{i=1}^{N} G_i k_i, \tag{C3}$$

where

$$G_i = \frac{1}{16\pi^2} \int_v \int_{v^1} \frac{\partial^2}{\partial r_3 \partial s_3} g_i(0, \mathbf{r}, \mathbf{s}) \frac{r_m s_m}{r^3 s^3} \, d\mathbf{r} \, d\mathbf{s}. \tag{C4}$$

Now consider that for each cell shape a distribution of sizes exists, so that

$$g_i(0, \mathbf{r}, \mathbf{s}) = \sum_{j=1}^{M} l_{ij} g_j^i(0, \mathbf{r}, \mathbf{s}), \tag{C5}$$

where l_{ij} is the frequency distribution of sizes of the ith cell geometry; $g_j^i(0, \mathbf{r}, \mathbf{s})$ is the probability of three points $(0, \mathbf{r}, \mathbf{s})$ being in the same cell of ith shape and jth size.

Since all cells of the ith shape differ only in size,

$$g_j^i(0, \mathbf{r}, \mathbf{s}) = \bar{g}^i(0, n_j\mathbf{r}, n_j\mathbf{s}), \tag{C6}$$

where $\bar{g}^i(0, n_j\mathbf{r}, n_j\mathbf{s})$ is the probability of three points $(0, n_j\mathbf{r}, n_j\mathbf{s})$ being in the ith cell shape of average size, and n_j = scale factor.

Put Eq. (C6) into integral (C4), change coordinates, and note that

$$\sum_{j=1}^{M} l_{ij} = 1. \tag{C7}$$

We note that G_i is independent of n_j.

Thus G_i depends only on the average size of cells of the ith geometry and not on the particular distribution of sizes for that geometry. G depends on the cell shapes which make up the distribution, multiplied by a distribution-function weighting factor; in this sense it is dependent on the average geometry.

APPENDIX D: DERIVATION OF BOUNDS FOR TWO-PHASE CELL MATERIAL

The upper bound on ϵ^* is given in Eq. (2.1) as

$$\epsilon^* \leq \langle\epsilon\rangle - \frac{1}{3} \frac{\langle\epsilon'^2\rangle/\langle\epsilon\rangle}{1 + 3\dfrac{\langle\epsilon\rangle I}{\langle\epsilon'^2\rangle}}, \tag{D1}$$

where

$$I = \frac{1}{16\pi^2\langle\epsilon\rangle^2}\int_v\int_{v^1}\frac{\partial^2}{\partial r_3\partial s_3}\langle\epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}. \tag{D2}$$

In Appendix A we derived

$$\langle\epsilon'(0)\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})\rangle = \epsilon_1'^3\varphi\left[g_1 - \frac{\varphi^2}{(1-\varphi)^2}g_2\right]. \tag{D3}$$

Substituting Eq. (D3) into (D2), we obtain

$$I = \frac{\epsilon_1'^3}{\langle\epsilon\rangle^2}\left[\varphi G_1 - \frac{\varphi^3}{(1-\varphi)^2}G_2\right], \tag{D4}$$

where

$$G_n = \frac{1}{16\pi^2}\int_v\int_{v^1}\frac{\partial^2}{\partial r_3\partial s_3}g_n\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}. \tag{D5}$$

Using the relationship

$$\langle\epsilon\rangle = \varphi\epsilon_1 + (1-\varphi)\epsilon_2 \tag{D6}$$

in Eq. (D1), we obtain

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left[1 + \varphi(\alpha-1) - \frac{\frac{1}{3}\varphi(1-\varphi)(\alpha-1)^2}{1 + (\alpha-1)[\varphi + 3((1-\varphi)^2G_1 - \varphi^2G_2)]}\right], \tag{D7}$$

where

$$\alpha = \epsilon_1/\epsilon_2.$$

If we have a symmetric cell material, Eq. (D7) reduces to

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left[1 + \varphi(\alpha-1) - \frac{\frac{1}{3}\varphi(1-\varphi)(\alpha-1)^2}{1 + (\alpha-1)(\varphi + 3(1-2\varphi)G)}\right]. \tag{D8}$$

The lower bound for ϵ^* is given by Eq. (2.2):

$$\epsilon^* \ge \left[\langle 1/\epsilon\rangle - \frac{\left(\frac{4}{3}\left\langle\frac{\epsilon'}{\epsilon}\right\rangle\right)^2\frac{1}{4\langle\epsilon\rangle^2}}{\frac{1}{3\langle\epsilon\rangle^2}\left\langle\frac{\epsilon'^2}{\epsilon}\right\rangle + J}\right]^{-1}, \tag{D9}$$

where

$$J = \frac{1}{16\pi^2\langle\epsilon\rangle^2}\int_v\int_{v^1}\frac{\partial^2}{\partial r_3\partial s_3}\left\langle\frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)}\right\rangle\frac{r_i s_i}{r^3 s^3}\,d\mathbf{r}\,d\mathbf{s}. \tag{D10}$$

In Appendix A we have shown that

$$\left\langle\frac{\epsilon'(\mathbf{r})\epsilon'(\mathbf{s})}{\epsilon(0)}\right\rangle = \frac{\epsilon_1^2}{\epsilon_1}\frac{\varphi}{(1-\varphi)}\left[(\alpha-1)(\varphi^2g_2 - (1-\varphi)^2g_1) + (\alpha + \varphi(1-\alpha))\left(\frac{(1-\varphi)}{\varphi}\bar{f}_1^{11} + \frac{\varphi}{(1-\varphi)}\bar{f}_2^{11}\right)\right]. \tag{D11}$$

Solving Eqs. (D10) and (D11), we obtain

$$J = \frac{\epsilon_1'^2}{\langle\epsilon\rangle^2\epsilon_1}\frac{\varphi}{(1-\varphi)}\left[(\alpha-1)(\varphi^2G_2 - (1-\varphi)^2G_1) + (\alpha + \varphi(1-\alpha))\left(\frac{(1-\varphi)}{\varphi}F_1 + \frac{\varphi}{(1-\varphi)}F_2\right)\right]. \tag{D12}$$

Now $F_1 = \varphi/3$ and $F_2 = (1-\varphi)/3$; hence Eq. (D9) may be rewritten as

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \ge \alpha^{\frac{1}{2}}\Big/\left[(\alpha - \varphi(\alpha-1)) - \frac{4}{3}\frac{(1-\alpha)^2(1-\varphi)\varphi}{1 + \alpha + 3(\alpha-1)(\varphi^2G_1 - (1-\varphi)^2G_2)}\right]. \tag{D13}$$

For a symmetric cell material this reduces to

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \ge \alpha^{\frac{1}{2}}\Big/\left[(\alpha - \varphi(\alpha-1)) - \frac{(1-\alpha)^2(1-\varphi)\varphi}{1 + \alpha + 3(\alpha-1)(2\varphi-1)G}\right]. \tag{D14}$$

APPENDIX E: DETERMINATION OF BOUNDS ON G

The upper bound on ϵ^* is given by

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left[1 + \varphi(\alpha-1) - \frac{\frac{1}{3}\varphi(1-\varphi)(\alpha-1)^2}{1 + (\alpha-1)(\varphi + 3((1-\varphi)^2G_1 - \varphi^2G_2))}\right]. \tag{E1}$$

From physical considerations ϵ_1, ϵ_2, and ϵ^* must all be positive-finite quantities; hence G_1 and G_2 must have values which do not violate this constraint. Therefore, the right-hand side of Eq. (E1) must be positive or, alternately,

$$1 + \varphi(\alpha - 1) \geq \frac{\frac{1}{3}\varphi(1 - \varphi)(\alpha - 1)^2}{1 + (\alpha - 1)(\varphi + 3((1 - \varphi)^2 G_1 - \varphi^2 G_2))} \quad \text{(E2)}$$

for all values of $\alpha \geq 1$.

First we shall consider the symmetric-cell case, for which Eq. (E2) reduces to

$$1 + \varphi(\alpha - 1) \geq \frac{\frac{1}{3}\varphi(1 - \varphi)(\alpha - 1)^2}{1 + (\alpha - 1)(\varphi + 3(1 - 2\varphi)G)}. \quad \text{(E3)}$$

In this case the inequality must hold for all values of φ, since the value of φ does not change the geometric structure.

When $\alpha = 1$, Eq. (E3) is satisfied. As α becomes large compared to unity, Eq. (E3) becomes

$$1 + \varphi\alpha \geq \frac{\frac{1}{3}\varphi(1 - \varphi)\alpha^2}{1 + \alpha(\varphi + 3(1 - 2\varphi)G)}. \quad \text{(E4)}$$

If $\varphi\alpha \gg 1$, Eq. (E4) becomes

$$1 \geq \frac{\frac{1}{3}(1 - \varphi)}{\varphi + 3(1 - 2\varphi)G}. \quad \text{(E5)}$$

When φ equals unity, the numerator becomes zero and the inequality holds for all values of G. However, for $\varphi = 1 - \eta$ we obtain the inequality

$$G \neq (1 - \eta)/3(1 - 2\eta). \quad \text{(E6)}$$

When $0 < \eta \leq \frac{1}{2}$, we find that G cannot take on values $\frac{1}{3} < G \leq \infty$. For $\frac{1}{2} \leq \eta \leq 1$, we find G cannot take on values $-\infty \leq G < 0$. Thus we have bounds on G which obey inequality (E5):

$$0 \leq G \leq \tfrac{1}{3}. \quad \text{(E7)}$$

Now consider the case $\varphi \ll 1$; Eq. (E4) becomes

$$(1 + \varphi\alpha)\varphi\alpha \geq 1/9G. \quad \text{(E8)}$$

If $\varphi\alpha \ll 1$, we obtain inequality $G \neq 0$; as α increases in value, G cannot equal values greater than zero. In the limit $\varphi\alpha \gg 1$ we obtain the bound on G:

$$G \geq \tfrac{1}{9}. \quad \text{(E9)}$$

Since negative values of G violate the condition of Eq. (E7), we have bounds on G:

$$\tfrac{1}{9} \leq G \leq \tfrac{1}{3}. \quad \text{(E10)}$$

For the asymmetric cell material we find that Eq. (E2) is satisfied when $\alpha = 1$. We first consider the case when φ is $\ll 1$ or $1 - \varphi \ll 1$. When α becomes large compared to unity, Eq. (E2) becomes

$$1 + \varphi\alpha \geq \frac{\frac{1}{3}\varphi(1 - \varphi)\alpha^2}{1 + \alpha(\varphi + 3((1 - \varphi)^2 G_1 - \varphi^2 G_2))}. \quad \text{(E11)}$$

When $\varphi = 0$, the inequality holds, and for $\varphi \ll 1$ Eq. (E2) becomes

$$1 + \varphi\alpha \geq \frac{\frac{1}{3}\varphi(\alpha - 1)^2}{1 + (\alpha - 1)(\varphi + 3G_1)}. \quad \text{(E12)}$$

When $\alpha = 1$, the inequality holds for all values of G_1. For $\varphi\alpha \ll 1$ Eq. (E12) becomes

$$1 \geq \frac{\frac{1}{3}\varphi(\alpha - 1)^2}{1 + 3(\alpha - 1)G_1}. \quad \text{(E13)}$$

Equation (E13) holds only for positive values of G_1. As α increases so that $\varphi\alpha \gg 1$, Eq. (E12) reduces to

$$1 \geq 1/3(\varphi + 3G_1). \quad \text{(E14)}$$

Since G_1 is positive, we obtain the bound

$$G_1 \geq \tfrac{1}{9} \quad \text{(E15)}$$

in order that Eq. (E14) be satisfied.

When $\varphi = 1$, Eq. (E12) is satisfied for all values of G_1 and G_2. For $1 - \varphi = \eta \ll 1$, Eq. (E12) is

$$\alpha \geq \frac{\frac{1}{3}\eta(\alpha - 1)^2}{1 + (\alpha - 1)(1 - 3G_2)}. \quad \text{(E16)}$$

As α increases in value, Eq. (E16) is satisfied if G_2 is positive and

$$G_2 \neq \tfrac{1}{3} + [1/3(\alpha - 1)] \quad \text{(E17)}$$

or G_2 is negative.

Varying α, we obtain the bound

$$G_2 \leq \tfrac{1}{3}. \quad \text{(E18)}$$

To obtain the other bounds on G_1 and G_2, we must examine the lower-bound equation

$$\frac{\epsilon^*}{(\epsilon_1\epsilon_2)^{\frac{1}{2}}} \geq \alpha^{\frac{1}{2}} \Big/ \left[\alpha + \varphi(1 - \alpha) - \frac{\frac{4}{3}(\alpha - 1)^2(1 - \varphi)\varphi}{1 + \alpha + 3(\alpha - 1)(\varphi^2 G_2 - (1 - \varphi)^2 G_1)}\right]. \quad \text{(E19)}$$

Physical considerations require that the right-hand side of Eq. (E19) be positive and finite; therefore,

$$\alpha + \varphi(1-\alpha) \geq \frac{\frac{4}{3}\varphi(\alpha-1)^2(1-\varphi)\varphi}{1+\alpha+3(\alpha-1)(\varphi^2G_2-(1-\varphi)^2G_1)}. \quad \text{(E20)}$$

When $\varphi = 0$, Eq. (E20) is satisfied, and for $\varphi \ll 1$ we obtain

$$\alpha \geq \frac{\frac{4}{3}(\alpha-1)^2\varphi}{1+\alpha-3(\alpha-1)G_1}. \quad \text{(E21)}$$

For $\alpha = 1$, Eq. (E21) is satisfied, and as α increases in value we obtain the requirement that

$$G_1 \neq (1+\alpha)/3(\alpha-1) \quad \text{(E22)}$$

or the bound

$$G_1 \leq \tfrac{1}{3}. \quad \text{(E23)}$$

When $\varphi = 1$, Eq. (E20) is satisfied and for $\eta = 1 - \varphi \ll 1$, we obtain

$$\eta\alpha + 1 \geq \frac{\frac{4}{3}(\alpha-1)^2\eta}{1+\alpha+3(\alpha-1)G_2}. \quad \text{(E24)}$$

When $\eta\alpha \ll 1$, this reduces to

$$1 \geq \frac{\frac{4}{3}(\alpha-1)^2\eta}{1+\alpha+3(\alpha-1)G_2}. \quad \text{(E25)}$$

This inequality is satisfied for

$$G_2 \neq -(1+\alpha)/3(\alpha-1). \quad \text{(E26)}$$

Hence G_2 must be positive.

Now consider the case $\eta\alpha \gg 1$. Equation (E24) reduces to

$$G_2 \geq \tfrac{1}{9}. \quad \text{(E27)}$$

Therefore, we obtain the bounds

$$\tfrac{1}{9} \leq G_1 \leq \tfrac{1}{3}, \quad \text{(E28)}$$

$$\tfrac{1}{9} \leq G_2 \leq \tfrac{1}{3} \quad \text{(E29)}$$

for values of φ near 0 or unity. Consider now the case when φ has arbitrary values between 0 and 1. For this cell structure values of G_1 and G_2 may be determined by consideration of only the shapes of the cells. This determination is made independently of φ. In addition, any cell shape is possible in the case of small concentrations. Thus, for example, the cells associated with G at finite φ may be made the inclusions in another material where $\varphi \ll 1$. Similar arguments for $1 - \varphi \ll 1$ show that Eqs. (E27) and (E28) hold for all values of φ.

APPENDIX F: THE GEOMETRIC SIGNIFICANCE OF G

The upper bound on ϵ^* may be written for an asymmetric cell material:

$$U = 1 + \varphi(\alpha-1) - \frac{\frac{1}{3}(1-\varphi)(\alpha-1)^2}{1+(\alpha-1)(\varphi+3((1-\varphi)^2G_1-\varphi^2G_2))}, \quad \text{(F1)}$$

where U = the upper bound on ϵ^*/ϵ_2.

Similarly, the lower-bound equation may be written as

$$L = \alpha\left[\alpha + \varphi(1-\alpha) - \frac{\frac{4}{3}(1-\alpha)^2(1-\varphi)\varphi}{1+\alpha+3(\alpha-1)(\varphi^2G_2-(1-\varphi)^2G_1)}\right]^{-1}, \quad \text{(F2)}$$

where L = the lower bound on ϵ^*/ϵ_2.

For small concentration Eqs. (F1) and (F2) reduce to

$$U = 1 + \varphi(\alpha-1)\left(1 - \frac{\alpha-1}{3+(\alpha-1)9G_1}\right) \quad \text{(F3)}$$

and

$$L = 1 + \varphi\frac{(\alpha-1)}{\alpha}\left(1 + \frac{4(\alpha-1)}{3(1+\alpha-3(\alpha-1)G_1)}\right). \quad \text{(F4)}$$

If we set $G_1 = \frac{1}{3}$, Eqs. (F3) and (F4) become

$$U = L = 1 + \frac{\varphi(\alpha-1)(2\alpha+1)}{3\alpha}. \quad \text{(F5)}$$

Alternately, if we set $G_1 = \frac{1}{9}$, Eqs. (F3) and (F4) become

$$U = L = 1 + \varphi\frac{3(\alpha-1)}{2+\alpha}. \quad \text{(F6)}$$

The upper- and lower-bound equations become asymptotic at small concentrations only for G_1 equal to $\frac{1}{3}$ or $\frac{1}{9}$. This can be seen by setting $U = L$ in Eqs. (F3) and (F4) and we get a quadratic equation in G_1. Since a quadratic equation can have only two roots and we have already shown two roots are $G_1 = \frac{1}{3}$ and $G_1 = \frac{1}{9}$, no other value of G_1 yield asymptotic upper- and lower-bound equations.

Similarly, when φ approaches unity, we find the roots of G_2 to also be $\frac{1}{3}$ and $\frac{1}{9}$.

Reynolds and Hough[9] have summarized the small-concentration solutions of randomly oriented spheroids and, in particular, have shown that the small-concentration solution of spheres and plates are, respectively,

$$\frac{\epsilon^*}{\epsilon_2} = 1 + \frac{3\varphi(\alpha-1)}{\alpha+2} \quad \text{(spheres)} \quad \text{(F7)}$$

and

$$\frac{\epsilon^*}{\epsilon_2} = 1 + \frac{\varphi(\alpha - 1)(2\alpha + 1)}{3\alpha} \quad \text{(plates)}. \tag{F8}$$

Equation (F8) is equivalent to Eq. (F5), and Eq. (F7) is equivalent to Eq. (F6).

Since G is not a function of φ, we conclude that $G = \frac{1}{9}$ has the geometric significance of a sphere and $G = \frac{1}{3}$ has the geometric significance of a plate.

Although we have shown that U and L are asymptotic at small concentrations for only $G = \frac{1}{3}$ and $\frac{1}{9}$ at small perturbations (i.e., α approaches unity), U and L are asymptotic for all values of G. This is to be expected because the bounding equations U and L were derived from a small perturbation expansion. From this result we can find the geometric significance of any other shape for which a small concentration solution exists. A problem of general interest is the needle-shaped cells. Reynolds and Hough[9] have shown that for random orientation the small-concentration solution is

$$\frac{\epsilon^*}{\epsilon_2} = 1 + \frac{\eta(6 + \eta)}{3(2 + \eta)}\varphi, \tag{F9}$$

where $\eta = \alpha - 1$.

Equations (F3) and (F4) for small perturbation can be written

$$U = L = 1 + \frac{\eta}{3}\frac{(3 + \eta(9G_1 - 1))}{(1 + 3\eta G_1)}\varphi. \tag{F10}$$

Solving Eqs. (F9) and (F10) for G_1, we find

$$G_1 = \tfrac{1}{6} \quad \text{(needle)}. \tag{F11}$$

Similarly, we can show that for oblate spheroids $\frac{1}{9} < G < \frac{1}{6}$, and for prolate spheroids $\frac{1}{6} < G < \frac{1}{3}$.

The above-mentioned results can be derived independently of the assumptions for a symmetric or asymmetric cell material. Assume that for any two phase material the probabilities $\overline{g_1}$, $\overline{g_2}$, $\overline{h_{11}}$, and $\overline{z_{222}}$ are the same order of magnitude in their dependence on volume fraction as the corresponding probabilities for a two-phase asymmetric cell material. This assumption and the relation

$$\varphi\epsilon_1' = -(1 - \varphi)\epsilon_2' \tag{F12}$$

reduce Eq. (A3) for small concentration to

$$\gamma_{3I} = \epsilon_1'^3 \varphi g_1. \tag{F13}$$

Following the procedure outlined in Appendix D, the upper-bound equation reduces to Eq. (F3).

Similarly, the lower-bound equation Eq. (F4) may be derived for a dilute two-phase material independent of the assumption of a symmetric or asymmetric cell material.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Bounds for Effective Bulk Modulus of Heterogeneous Materials

MELVIN N. MILLER
P.O. Box 134, Chalfont, Pennsylvania

(Received 1 October 1968)

Bounds for the effective bulk modulus for statistically isotropic and homogeneous materials have been developed in terms of statistical information, i.e., one-point and three-point correlation function from variational principles. Aside from the one-point correlation function, i.e., the volume fraction, this statistical information is difficult or impossible to obtain for real materials. For a broad class of heterogeneous materials which we shall call cell materials, the functions of the three-point correlation function that appear in the bounds of effective bulk modulus are simply a number for each phase. Furthermore, this number has a range of values $\frac{1}{9}$ to $\frac{1}{8}$ and a simple geometric significance. The number $\frac{1}{9}$ implies a cell of spherical shape, the number $\frac{1}{8}$ a cell of plate-like shape, and all other cell shapes, no matter how irregular, have a corresponding number between. Each value of this number determines a new set of bounds which are substantially narrower and always within the best bounds for two-phase media in terms of volume fraction alone (i.e., Hashin–Shtrikman bounds). For dilute suspensions the new bounds are so narrow in most cases as to be essentially an exact solution. There is a substantial improvement over previous bounds for a finite suspension and yet greater improvement for multiphase materials where the geometric characteristics of each phase are known. The shape factor G is found to have exactly the same range of numerical values and the same geometric significance as was found in the determination of effective dielectric constant bounds. It was found further that under certain conditions the bounds on effective bulk modulus and dielectric constant become numerically identical.

1. INTRODUCTION

It was shown[1] that the bounds on the effective dielectric constant of Beran[2] in terms of statistical information for a large class of heterogeneous materials could be expressed in terms of volume fraction and a shape factor G for each phase. These bounds were shown to converge to an exact solution for dilute suspension and for a class of materials, called cell materials, to be narrower than the best bounds in terms of volume fraction alone established by Hashin and Shtrikman[3] for two-phase media. The shape factor G, which is a function of the three-point correlation function, was shown to have a range of values $\frac{1}{9}$ to $\frac{1}{8}$. The geometric significance of $\frac{1}{9}$ was shown to be spheres, $\frac{1}{8}$ randomly oriented thin plates, and all other geometries, no matter how irregular, were shown to have a value within this range.

The purpose of this paper is to extend these notions to the bulk-modulus property.

2. EFFECTIVE BULK-MODULUS BOUNDS

A. Introduction

It has been shown by Beran and Molyneux[4] that the effective bulk modulus K^* may be bounded by using perturbation expansions as trial functions for the stress and strain fields in the two standard variational principles of elasticity. Including the first-order perturbation effect led then to the following bounds:

$$K^* \leq \langle K \rangle - \frac{\langle K'^2 \rangle}{\langle \lambda \rangle + 2\langle \mu \rangle + (\langle \lambda' K'^2 \rangle / \langle K'^2 \rangle)(2I'/\langle K'^2 \rangle)} \tag{2.1}$$

and

$$K^* \geq \left[\frac{1}{\langle K \rangle} - \frac{\langle K'/K \rangle^2}{\langle K'^2/K \rangle - \frac{3}{8}\langle K'^2/\mu \rangle + \frac{9}{8}J'} \right]^{-1}, \tag{2.2}$$

where

$$I' = \frac{1}{16\pi^2} \int_\zeta \int_n \frac{\partial^2}{\partial \zeta_i \partial \zeta_j} \frac{\partial^2}{\partial \eta_i \partial \eta_j} \times \langle \mu'(0)\mu'(\boldsymbol{\zeta}) K'(\mathbf{n}) \rangle \frac{1}{|\zeta|\,|n|} d\boldsymbol{\zeta}\, d\mathbf{n}, \tag{2.3}$$

$$J' = \frac{1}{16\pi^2} \int_\zeta \int_n \frac{\partial^2}{\partial \zeta_i \partial \zeta_j} \frac{\partial^2}{\partial \eta_i \partial \eta_j} \times \left\langle \frac{K'(\boldsymbol{\zeta}) K'(\mathbf{n})}{\mu(0)} \right\rangle \frac{1}{|\zeta|\,|n|} d\boldsymbol{\zeta}\, d\mathbf{n}. \tag{2.4}$$

λ', μ', and K' are the fluctuating parts of the Lamé, shear, and bulk moduli, respectively, and the bracket denotes an ensemble average, which is assumed to be equal to the space average for a statistically homogeneous medium.

Therefore, these bounds on the effective bulk modulus depend on functions of the three-point correlation functions

$$\langle \mu'(0) K'(\mathbf{r}) K'(\mathbf{s}) \rangle \quad \text{and} \quad \langle K'(\mathbf{r}) K'(\mathbf{s}) / \mu(0) \rangle.$$

Better bounds can be derived using this approach in

[1] M. N. Miller, J. Math. Phys. **10**, 1988 (1969).
[2] M. Beran, Nuovo Cimento **38**, 771 (1965).
[3] Z. Hashin and S. Shtrikman, J. Appl. Phys. **33**, 3125 (1962).
[4] M. Beran and J. Molyneux, Quart. Appl. Math. **24**, 107 (1966).

terms of higher-order correlation functions, and the bounds in terms of the n-point correlation function will be at least as good as those in terms of the $n-1$ point correlation function, with the expectation that as n approaches infinity, the upper and lower bounds will converge. The usefulness of expressing bounds in this fashion depends upon the determination of I' and J' and similar derivatives and integrals of higher-order correlation functions. We shall see that for a broad class of materials, which we shall define, those functions are particularly simple, have simple physical significance, and lead to a significant improvement over previous bounds.

B. Symmetric-Cell Material

As discussed by Miller,[1] the geometry of a symmetric-cell material is defined as any division of the space into cells which fulfills the following requirements:

(1) Space is completely covered by cells.

(2) Cells are distributed in a manner such that the material is statistically homogeneous and isotropic.

(3) The material property of a cell is statistically independent of the material property of any other cell.

(4) The conditional probability of n points being and m points not being in the same cell of a particular material, given that one point is in a cell of that material, is the same for each material.

Using the approach discussed,[1] we show in the Appendix that for a symmetric-cell material the three-point correlation functions which appear in Eqs. (2.3) and (2.4) are

$$\gamma_u = \langle \mu'(0)K'(\mathbf{r})K'(\mathbf{s})\rangle = \mu_1' K_1'^2[\varphi(1-2\varphi)/(1-\varphi)^2]g(0,\mathbf{r},\mathbf{s}) \quad (2.5)$$

and

$$\gamma_l = \langle K'(\mathbf{r})K'(\mathbf{s})/\mu(0)\rangle = (K_1'^2/\mu_1)[\varphi/(1-\varphi)]\{(\beta-1)(2\varphi-1)g(0,\mathbf{r},\mathbf{s}) + [\beta+\varphi(1-\beta)]f^{11}(\mathbf{r},\mathbf{s})\}, \quad (2.6)$$

where

$$\beta = \mu_1/\mu_2, \quad K_1' = K_1 - \langle K\rangle, \quad \mu_1' = \mu_1 - \langle \mu\rangle,$$

where $g(0,\mathbf{r},\mathbf{s})$ is the conditional probability of a triangle (coordinates 0, $\mathbf{r}$, $\mathbf{s}$) having all three vertices in a single cell, given that one point is in the cell, $f''(\mathbf{r},\mathbf{s})$ is the conditional probability of a line segment [and coordinates $(\mathbf{r},\mathbf{s})$] having both ends in a single cell, given that one end is in the cell.

Substituting Eqs. (2.5) and (2.6) into Eqs. (2.3) and (2.4), respectively, we obtain

$$I' = 3\mu_1' K_1'^2[\varphi(1-2\varphi)/(1-\varphi)^2]G \quad (2.7)$$

and

$$J' = (K_1'^2/\mu_1)[\varphi/(1-\varphi)]\{3(\beta-1)(2\varphi-1)G + [\beta+\varphi(1-\beta)]F\}, \quad (2.8)$$

where

$$3G = \frac{1}{16\pi^2}\int_{v^1}\int_{v^{11}} \frac{\partial^2}{\partial\zeta_i\partial\zeta_j}\frac{\partial^2}{\partial\zeta_i\partial\zeta_j} g(0,\boldsymbol{\zeta},\mathbf{n})\frac{1}{|\zeta|\,|n|}\,d\boldsymbol{\zeta}\,d\mathbf{n}, \quad (2.9)$$

$$F = \frac{1}{16\pi^2}\int_{v^1}\int_{v^{11}} \frac{\partial^2}{\partial\zeta_i\partial\zeta_j}\frac{\partial^2}{\partial\zeta_i\partial\zeta_j} f^{11}(\boldsymbol{\zeta},\mathbf{n})\frac{1}{|\zeta|\,|n|}\,d\boldsymbol{\zeta}\,d\mathbf{n}. \quad (2.10)$$

Integrals like F are evaluated[4] by noting that $f^{11}(\boldsymbol{\zeta},\mathbf{n}) = f^{11}(|\boldsymbol{\zeta}-n|)$, and $\partial/\partial\boldsymbol{\zeta}_i = -\partial/\partial\boldsymbol{\zeta}_i$, and $f^{11}(0)=1$, and that

$$F = 1. \quad (2.11)$$

Substituting Eqs. (2.19) and (2.20) into Eqs. (2.1) and (2.2), respectively, we find the following bounds for K^*:

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left(1+(\alpha-1)\varphi - \frac{\varphi(1-\varphi)(\alpha-1)^2}{[\alpha-(\alpha-1)\varphi]+2\gamma[1-\frac{4}{3}\varphi+\frac{1}{3}\beta(4\varphi-1)+3(\beta-1)(1-2\varphi)G]}\right) \quad (2.12)$$

and

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \ge \alpha^{\frac{1}{2}}\left(\varphi(1-\alpha)+\alpha - \frac{\varphi(1-\varphi)(1-\alpha)^2}{1+\varphi(\alpha-1)+\frac{3}{8}(\alpha/\beta\gamma)\{3(\beta-1)[(2\varphi-1)3G-\frac{4}{3}\varphi]+3\beta-1\}}\right)^{-1}, \quad (2.13)$$

where

$$\alpha = K_1/K_2 \ge 1, \quad \beta = \mu_1/\mu_2, \quad \gamma = \mu_2/K_2.$$

We see that the upper and lower bounds on K^* for the symmetric-cell material depend on α, β, γ, φ, and a single number G which is characteristic of the average cell geometry.

Real materials have positive and finite effective bulk moduli, for positive finite values of K_1, K_2, μ_1, μ_2. These physical requirements put limitations on the range of values which G may have. We have shown[1] that imposing these requirements on K^* in Eqs. (2.12) and (2.13) limit the values of G to the range

$$\tfrac{1}{9} \le G \le \tfrac{1}{3}. \quad (2.14)$$

Therefore G has precisely the same range of values as the G defined in Miller,[1] and we shall see that the two G's are indeed identical. If we put the value $G = \frac{1}{9}$ into the bounding equations (2.12) and (2.13), we find that the two equations become asymptotic for small concentration (i.e., $\varphi \to 0$ or 1) and the asymptotic solution is identical to the small concentration solution of spheres in a matrix. Similarly, putting $G = \frac{1}{3}$ in the bounding equations, they again become asymptotic at small concentrations and the asymptotic solution is the small concentration solution for plates in a matrix. Therefore, we assign the geometric significance of a sphere to $G = \frac{1}{9}$ and a plate to $G = \frac{1}{3}$, which is the same significance they had in Miller.[1]

Hashin and Shtrikman[5] have shown that the best bounds on K^* for a two-phase statistically homogeneous and isotropic material when only φ is specified are

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \leq \frac{1}{\alpha^{\frac{1}{2}}} \frac{3\alpha + \beta 4\gamma[1 + \varphi(\alpha - 1)]}{4\mu\gamma + 3[\varphi(1 - \alpha) + \alpha]} \tag{2.15}$$

and

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \geq \frac{1}{\alpha^{\frac{1}{2}}} \frac{3\alpha + 4\gamma[1 + \varphi(\alpha - 1)]}{4\gamma + 3[1 + (\alpha - 1)(1 - \varphi)]}. \tag{2.16}$$

These bounds are the exact solution for a space filled with concentric spheres; the upper bound represents a low-bulk modulus core enclosed in a high-bulk modulus shell and the lower bound represents the reverse situation. When the volume fraction of K_1 is small, the lower bound is the case of a high-bulk modulus sphere in a low-bulk modulus matrix. For this case Eq. (2.16) reduces to Eqs. (2.12) and (2.13) with a $G = \frac{1}{9}$, which confirms our conclusion that $G = \frac{1}{9}$ represents a spherical cell shape. Similarly, when the volume fraction of K_1 approaches unity, the Hashin–Shtrikman upper bound is the case of a low-bulk modulus sphere in a high-bulk modulus matrix. For this case, Eq. (2.15) reduces to Eqs. (2.12) and (2.13), with $G = \frac{1}{9}$ confirming the conclusions of our previous discussion. In contrast, when the volume fraction of K_1 material is small, the Hashin–Shtrikman upper bound represents the case of thin shells of high-bulk modulus material in a low-bulk modulus matrix. In this instance, Eq. (2.15) reduces to Eqs. (2.12) and (2.13) with $G = \frac{1}{3}$. Volume fraction φ approaches unity, Eq. (2.16) represents a thin shell of low-bulk modulus material in a matrix of high-permittivity material, and Eq. (2.16) reduces to Eqs. (2.12) and (2.13) with $G = \frac{1}{3}$. This confirms that the thin-shell solution is identical to the plate solution at low concentrations.

For $\varphi = \frac{1}{2}$, both Eqs. (2.12) and (2.13) become independent of G and consequently independent of the geometry. This is in agreement with the known result for a symmetric random medium that odd-order correlation functions[6] become functions of lower even-ordered correlation functions. Furthermore, Eqs. (2.12) and (2.13) are in agreement with the results obtained by Beran and Molyneux[4] for the case of a two-phase symmetric medium.

Hashin and Shtrikman[5] have found that their bounds converge to an exact solution for the case in which both materials have the same shear modulus, i.e., $\beta = 1$. Bounding equations (2.12) and (2.13) converge to the same exact solution for constant-shear modulus.

When γ, the ratio of μ_2 to K_2, approaches zero, the bounding equations (2.12) and (2.13) converge to an exact solution. Since a γ of zero refers to an incompressible material, this result is not unexpected. As γ increases, the bounds on K^* increases in value and separate until they reach a maximum value and separation for $\gamma = 1.5$, i.e., a Poisson ratio of zero. Interestingly, at this value of γ with $\alpha = \beta$, the bounding equations on the effective bulk modulus K^* reduces to the bounding equations[1] on the effective permittivity ϵ^*. This is true for the bounding equations developed here and the Hashin–Shtrikman bounding equations. This analogy requires further consideration.

For all possible values of α, β, γ, φ, and G, the new bounding equations (2.12) and (2.13) are equal to or inside the Hashin–Shtrikman bounding equations (2.15) and (2.16), and for most cases the width between the new bounding equations is a small fraction of the width between the Hashin–Shtrikman bounding equation.

Figure 1 shows Eqs. (2.12) and (2.13) for $\alpha = \beta = 10$, $\gamma = 1.0$, $G = \frac{1}{9}$, $G = \frac{1}{3}$, and the Hashin–Shtrikman bounding equations. We note a substantial improvement in the knowledge of K^* for all values of φ as compared to the Hashin–Shtrikman bounds, with the least improvement occurring at $\varphi = \frac{1}{2}$. At $\varphi = \frac{1}{2}$ the improvement, defined as the difference in width of the Hashin–Shtrikman bounds and the new bounds divided by the Hashin–Shtrikman bounds, is approximately 50%, while for $\varphi = 0.9$ and $G = \frac{1}{9}$, the improvement is 83%. The extreme values of the $G = \frac{1}{3}$ and the $G = \frac{1}{9}$ curves represent the extremes

[5] Z. Hashin and S. Shtrikman, J. Mech. Phys. Solids **11**, 127 (1963).

[6] H. L. Frisch, Trans. Soc. Rheol. **9**, 293 (1965).

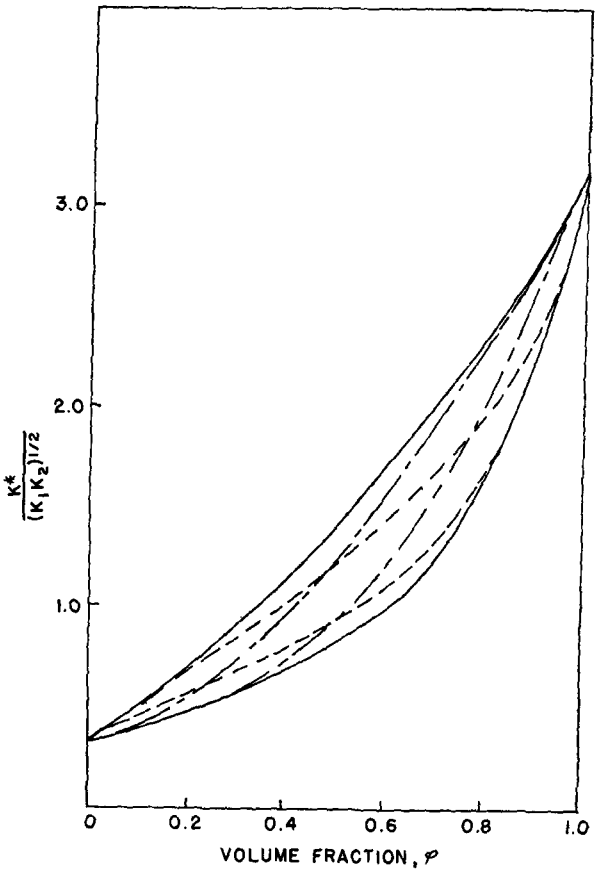

FIG. 1. Bounds on effective bulk modulus $\alpha = \beta = 10$, $\gamma = 1.0$, symmetric-cell material:

——— Hashin–Shtrikman,
– – – $G = \frac{1}{3}$,
— – — $G = \frac{1}{9}$.

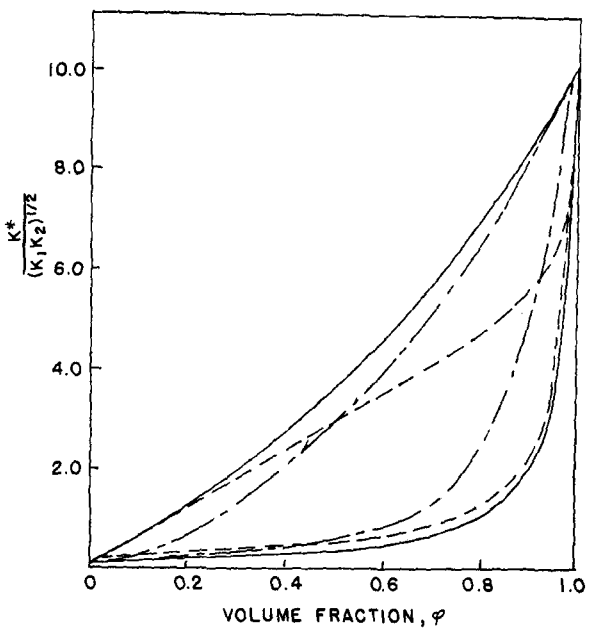

FIG. 2. Bounds on effective bulk modulus $\alpha = \beta = 100$, $\gamma = 1.0$, symmetric-cell material:

——— Hashin–Shtrikman,
– – – $G = \frac{1}{3}$,
— – — $G = \frac{1}{9}$.

for all symmetric-cell materials, i.e., for all other values of G the upper- and lower-bound curves fall inside these curves, except at $\varphi = \frac{1}{2}$, when all upper-bound curves have the same K^* and all lower-bound curves have the same K^*. If G is unknown, there is still an improvement over the Hashin–Shtrikman bounds and the maximum improvement occurs at $\varphi = \frac{1}{2}$.

Figure 1 reveals that the bounds on K^* are narrower for $G = \frac{1}{9}$ (spheres) than for $G = \frac{1}{3}$ (plates) for all values of φ except $\varphi = \frac{1}{2}$, when all G values have the same bounds. This is because there is a single degree of freedom (i.e., the location of the spheres' centers) associated with a distribution of spheres, while there are more than two degrees of freedom (i.e., the location of the plate mass center and the rotational degrees of freedom of the plate) associated with a distribution of plates. There is, therefore, a higher degree of uncertainty in the geometrical configuration when the material has aspherical cells, hence a greater uncertainty in the effective property of the material.

From Figs. 1 and 2, we can see that the percentage improvement over the Hashin–Shtrikman bounds decreases with increasing α; but there is still a substantial improvement at very large α. For example, at $\alpha = \beta = 100$, $\gamma = 1.0$, there is a better than 27% improvement at $\varphi = \frac{1}{2}$ and a better than 52% improvement at $\varphi = 0.9$, $G = \frac{1}{9}$.

As β increases, the percentage improvement over the Hashin–Shtrikman bounds increases as can be seen by comparing Figs. 3 and 4 with Figs. 1 and 2, respectively. For $\alpha = 10$, $\beta = 2$, $\gamma = 10$, $G = \frac{1}{9}$, the percentage improvement over the Hashin–Shtrikman bounds is 83% at $\varphi = 0.5$ and 94% at $\varphi = 0.9$. For $\alpha = 100$, $\beta = 2$, $\gamma = 1.0$, $G = \frac{1}{9}$, the percentage improvement over the Hashin–Shtrikman bounds is 82% at $\varphi = 0.5$ and 92% at $\varphi = 0.9$. For small β, the change in improvement is approximately the same at all values of α. In Fig. 5 we plot $\alpha = 10$, $\beta = 1$ for various values of γ. As we mentioned previously, the new bounds and the Hashin–Shtrikman bounds converge to an exact solution for $\beta = 1$, and this figure shows how the solution increases with increasing γ. Comparing Figs. 6 and 7 with Fig. 1 demonstrates how the bounding equations change with γ. As γ decreases, the percentage improvement decreases so that for $\alpha = \beta = 10$, $\gamma = 0.5$, $G = \frac{1}{9}$, the percentage improvement at $\varphi = 0.9$ is 78%, compared to 83% when $\gamma = 1.0$ and to less than 1%

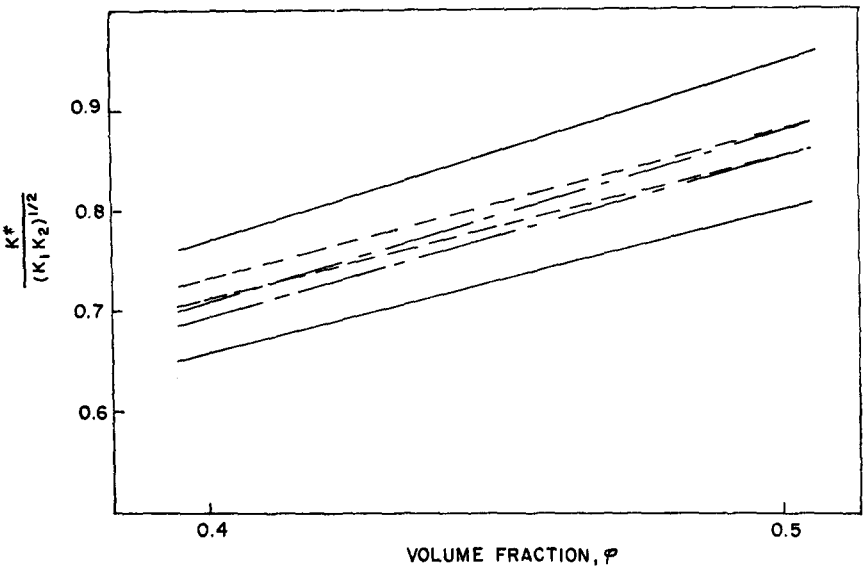

FIG. 3. Bounds on effective bulk modulus $\alpha = 10$, $\beta = 2$, $\gamma = 1.0$, symmetric-cell material:

——— Hashin–Shtrikman,
– – – $G = \frac{1}{3}$,
— – — $G = \frac{1}{9}$.

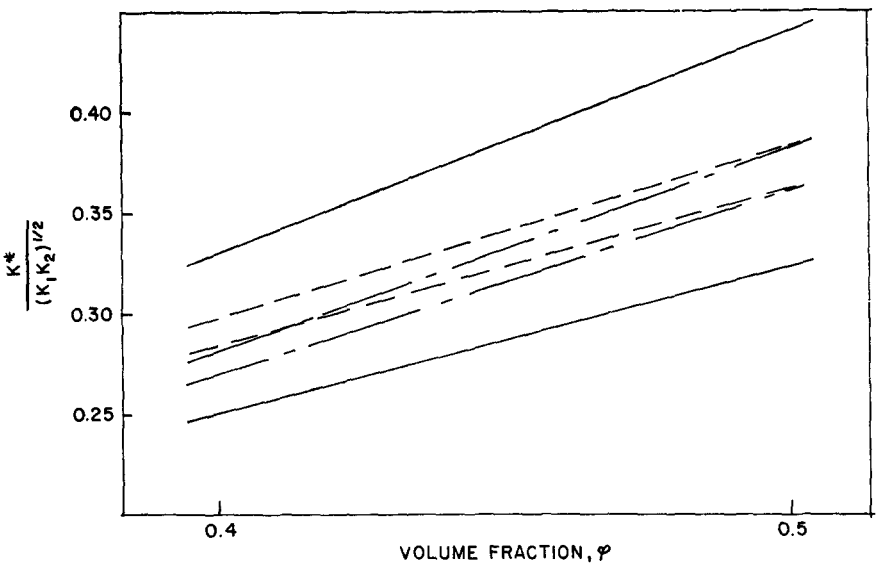

FIG. 4. Bounds on effective bulk modulus $\alpha = 100$, $\beta = 2$, $\gamma = 1.0$, symmetric-cell material:

——— Hashin–Shtrikman,
– – – $G = \frac{1}{3}$,
— – — $G = \frac{1}{9}$.

at $\gamma = 10^{-6}$. However, as γ approaches zero, the bounds approach the exact solution, as may be seen in Fig. 7.

An interesting phenomenon occurs when we consider the mixed case of $\alpha > 1$ and $\beta < 1$. The Hashin–Shtrikman bounds reverse themselves: that is, the upper-bound equation becomes the equation for the lower bound and the lower-bound equation becomes the equation for the upper bound. The bounds also reverse themselves in the sense that for $G = \frac{1}{9}$ (spheres), the new bounds are near the Hashin–Shtrikman upper bound, for β approaching zero, and near the Hashin–Shtrikman lower bound, for β approaching unity, which is the reverse of what occurred for $\beta > 1$. Conversely, the new bounds for $G = \frac{1}{3}$ (plates) are near the Hashin–Shtrikman lower bound for φ approaching zero and near the upper bound for φ approaching unity. The bounds spread and it is seen that the percentage improvement decreases as β approaches zero. For $\alpha = 10$, $\beta = 0.5$, $\gamma = 1.0$, $G = \frac{1}{9}$, there is an 83% improvement at $\varphi = 0.5$ and a 96% at $\varphi = 0.9$. For $\alpha = 10$, $\beta = 0.1$, $\gamma = 1.0$, $G = \frac{1}{9}$, there is a 56% improvement at $\varphi = 0.5$ and an 87% improvement at $\varphi = 0.9$.

Although the mixed case considered here of $\alpha > 1$ and $\beta < 1$ falls outside the assumptions of Hashin–Shtrikman,[5] Walpole[7] has shown that their bounding equations are valid for this case.

It is shown[1] that, at small concentrations, the effective permittivity bounding equations hold for all

[7] L. J. Walpole, J. Mech. Phys. Solids **14**, 151 (1966).

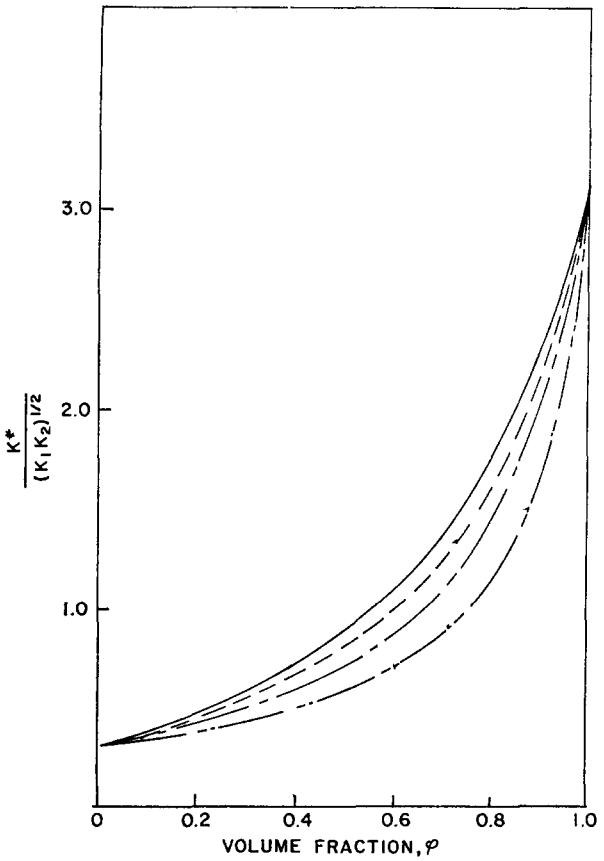

FIG. 5. Bounds on effective bulk modulus $\alpha = 10$, $\beta = 1$, $0 < \gamma \leq 1.5$:

——— $= 1.5$,
– – – $= 1.0$,
— – — $= 0.5$,
— – – — $= 10^{-6}$.

two-phase statistically isotropic and homogeneous materials and are not restricted to symmetric-cell materials. Following the same procedure, the upper- and lower-bound equations (2.12) and (2.13) may be derived for a dilute two-phase material, independent of the assumption of a symmetric- or asymmetric-cell material.

C. Asymmetric-Cell Material

We shall now investigate a class of materials for which the geometry of the cells of the two materials is dissimilar. Consider the space to be subdivided by a large number of closed surfaces and let us call the enclosed regions cells. The subdivision of the space is arbitrary except for fulfilling the following requirements:

(1) Space is completely covered by cells.

(2) Cells are distributed in a manner such that the material is statistically homogeneous and isotropic.

(3) The material property ϵ of a cell is statistically independent of the material property of any other cell.

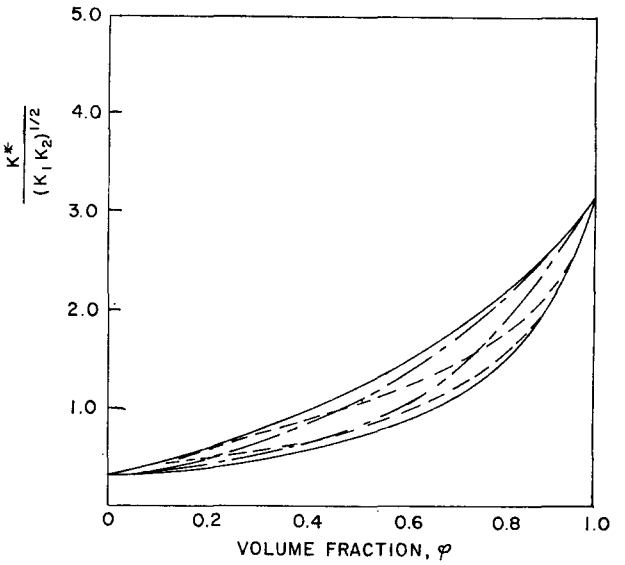

FIG. 6. Bounds on effective bulk modulus symmetric-cell material for $\alpha = \beta = 10$, $\gamma = 0.5$:

——— Hashin–Shtrikman,
– – – $G = \frac{1}{3}$,
— – — $G = \frac{1}{6}$.

Any material which satisfies these requirements shall be called an asymmetric-cell material and models of these materials are described in Miller.[1]

For this material, using the approach formulated in Miller,[1] we find that the three-point correlation functions that appear in Eqs. (2.3) and (2.4) are

$$\gamma_{3I'} = \langle \mu'(0)K'(\mathbf{r})K'(\mathbf{s}) \rangle$$
$$= \mu_1' K_1'^2 \{ \varphi g_1(0, \mathbf{r}, \mathbf{s}) - [\varphi^3/(1-\varphi)^2] g_2(0, \mathbf{r}, \mathbf{s}) \} \tag{2.17}$$

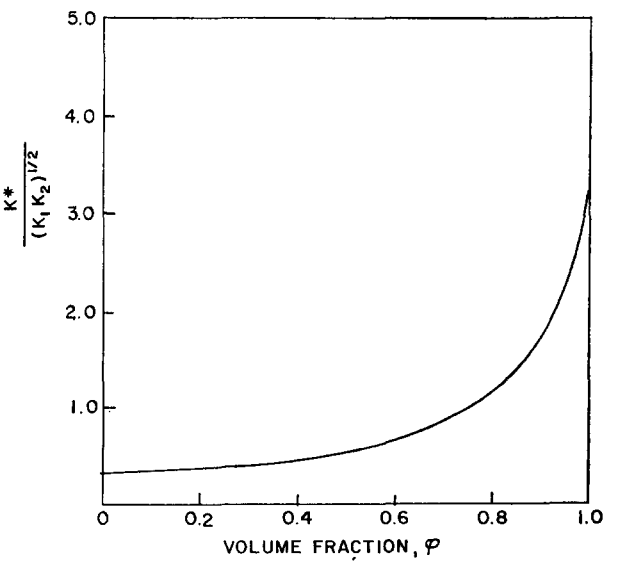

FIG. 7. Effective bulk modulus asymmetric-cell material for $\alpha = \beta = 10$, $\gamma = 10^{-6}$:

——— Hashin–Shtrikman and all combinations of G_1 and G_2.

and

$$\gamma_{3J'} = \left\langle \frac{K'(\mathbf{r})K'(\mathbf{s})}{\mu(0)} \right\rangle = \frac{K_1'^2}{\mu_1}\frac{\varphi}{(1-\varphi)}\Big[(\beta-1)[\varphi^2 g_2(0,\mathbf{r},\mathbf{s}) + (1-\varphi)^2 g_2(0,\mathbf{r},\mathbf{s})] + [\varphi(1-\beta)+\beta]\Big(\frac{(1-\varphi)}{\varphi}\bar{f}_1^{11}(\mathbf{r},\mathbf{s}) + \frac{\varphi}{(1-\varphi)}\bar{f}_2^{11}(\mathbf{r},\mathbf{s})\Big)\Big], \quad (2.18)$$

where

$g_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability of a triangle $(0, \mathbf{r}, \mathbf{s})$ having all three vertices in a single cell of material property K_n and μ_n, given that one point is in the cell;

$\bar{f}_n^{11}(\mathbf{r}, \mathbf{s})$ is the probability of a line segment $(\mathbf{r}, \mathbf{s})$ having both ends in a single cell of material property K_n and μ_n.

Substituting Eqs. (2.17) and (2.18) into Eqs. (2.3) and (2.4), we obtain

$$I' = \mu_1' K_1'^2 3\varphi\{G_1 - [\varphi^2/(1-\varphi)^2]G_2\} \quad (2.19)$$

and

$$J' = \frac{K_1'^2}{\mu_1}\frac{\varphi}{(1-\varphi)}\Big[3(\beta-1)[\varphi^2 G_2 - (1-\varphi)^2 G_1] + [\beta+\varphi(1-\beta)]\Big(\frac{(1-\varphi)}{\varphi}F_1 + \frac{\varphi}{(1-\varphi)}F_2\Big)\Big], \quad (2.20)$$

where we define

$$3G_n = \frac{1}{16\pi^2}\int_v\int_{v^1}\frac{\partial^2}{\partial r_i \partial r_j}\frac{\partial^2}{\partial s_i \partial s_j} g_n(0,\mathbf{r},\mathbf{s})\frac{1}{|\mathbf{r}|\,|\mathbf{s}|}\,d\mathbf{r}\,d\mathbf{s} \quad (2.21)$$

and

$$F_n = \frac{1}{16\pi^2}\int_v\int_{v^1}\frac{\partial^2}{\partial r_i \partial r_j}\frac{\partial^2}{\partial s_i \partial s_j}\bar{f}_n^{11}(\mathbf{r},\mathbf{s})\frac{1}{|\mathbf{r}|\,|\mathbf{s}|}\,d\mathbf{r}\,d\mathbf{s}. \quad (2.22)$$

As shown in Sec. 2.2, G_1 and G_2 are numbers which depend on the geometries of the average cells of material property K_1 and K_2. Integrating Eq. (2.22), we find $F_1 = \varphi$ and $F_2 = 1 - \varphi$.

Substituting Eqs. (2.19) and (2.20) into Eqs. (2.17) and (2.18), we obtain

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \le \frac{1}{\alpha^{\frac{1}{2}}}\left(1+\varphi(\alpha-1) - \frac{\varphi(1-\varphi)\alpha^2}{\alpha-(\alpha-1)\varphi+2\gamma\{1-\frac{4}{3}\varphi+\frac{1}{3}\beta(4\varphi-1)+3(\beta-1)[G_1(1-\varphi)^2-G_2\varphi]^2\}}\right), \quad (2.23)$$

$$\frac{K^*}{(K_1K_2)^{\frac{1}{2}}} \ge \alpha^{\frac{1}{2}}\left(\varphi(1-\alpha)+\alpha - \frac{\varphi(1-\varphi)(\alpha-1)^2}{1+\varphi(\alpha-1)+(3\alpha/8\beta\gamma)(3(\beta-1)\{3[G_2\varphi^2-G_1(1-\varphi)^2]-\frac{4}{3}\varphi+3(\beta-1)\}^2)}\right)^{-1}. \quad (2.24)$$

When $G_1 = G_2$, Eqs. (2.23) and (2.24) reduce as expected to the symmetric-cell material bounding equations.

For all possible values of α, β, γ, φ, G_1, and G_2, the bounding equations (2.23) and (2.24) fall inside the Hashin–Shtrikman bounds. The bounding equations attain their highest upper bound and lowest lower bound for the combinations $G_1 = \frac{1}{3}$, $G_2 = \frac{1}{9}$, and $G_1 = \frac{1}{9}$, $G_2 = \frac{1}{3}$, respectively, for β greater than unity. When β is less than unity, these combinations of G_1 and G_2 reverse themselves, but still give the extreme bounds. Hence, these combinations constitute absolute bounds for asymmetric-cell materials. In Sec. 2.4, we discussed why these absolute bounds fall inside the Hashin–Shtrikman bounds.

For β greater than unity and φ approaching zero, the new bounding equations converge to the Hashin–Shtrikman upper bound when $G_1 = \frac{1}{3}$ and to the Hashin–Shtrikman upper bound when $G_1 = \frac{1}{9}$. Similarly, when φ approaches unity, the same convergence occurs when $G_2 = \frac{1}{9}$ and $G_2 = \frac{1}{3}$, respectively, confirming our conclusion regarding the geometric significance of these two values of G.

In Figs. 8 and 9, we plot the new bounds for the combination $G_1 = \frac{1}{9}$, $G_2 = \frac{1}{3}$, and $G_1 = \frac{1}{3}$, $G_2 = \frac{1}{9}$, and the Hashin–Shtrikman bounds for various values of α. As was the case for the effective permittivity, we see that the shape of the bounds found for the symmetric-cell material are replaced by a shape that is similar to the Hashin–Shtrikman bounds.

As discussed in Sec. 2.2, the bounding equations developed here at small concentrations are not restricted to asymmetric-cell materials, but hold for all two-phase statistically isotropic and homogeneous materials.

As was the case for the symmetric-cell material when

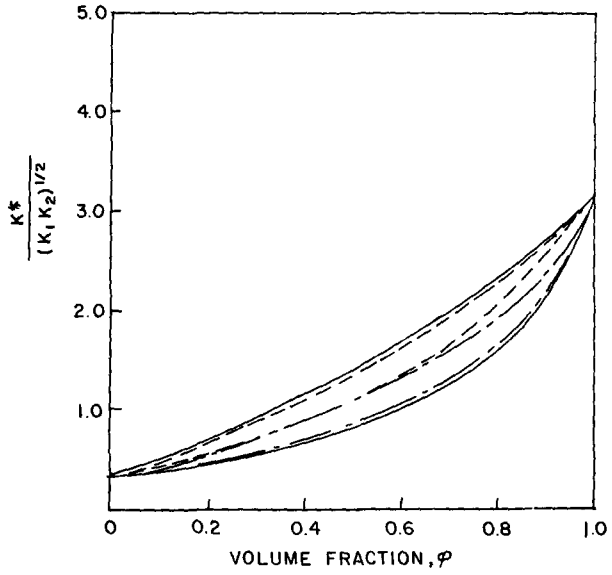

FIG. 8. Bounds on effective bulk modulus asymmetric-cell material for $\alpha = \beta = 10$, $\gamma = 1.0$:

——— Hashin–Shtrikman,
– – – $(G_1 = \frac{1}{3},\ G_2 = \frac{1}{9})$,
— — — $(G_1 = \frac{1}{9},\ G_2 = \frac{1}{3})$.

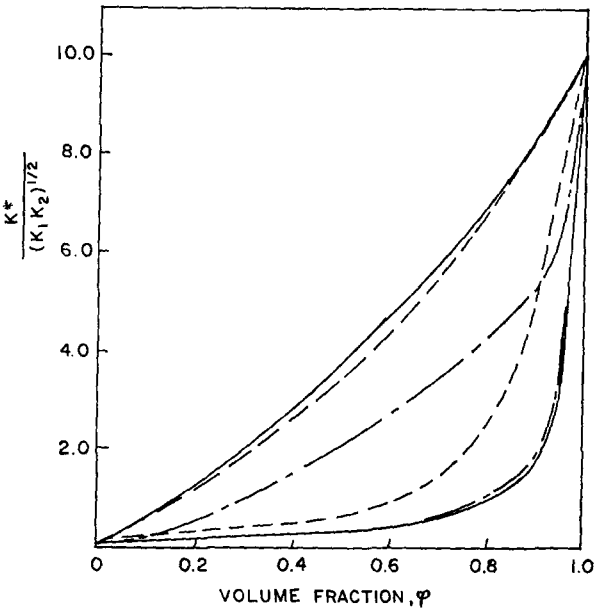

FIG. 9. Bounds on effective bulk modulus asymmetric-cell material for $\alpha = \beta = 100$, $\gamma = 1.0$:

——— Hashin–Shtrikman,
– – – $(G_1 = \frac{1}{3},\ G_2 = \frac{1}{9})$,
— — — $(G_1 = \frac{1}{9},\ G_2 = \frac{1}{3})$.

$\gamma = 1.5$ (equivalent to Poisson ratio of zero) and $\alpha = \beta$, the bounding equations on the effective bulk modulus K^* reduces to the bounding equations on effective permittivity that were previously derived.[1]

All of the notions derived here for a two-phase random media are applicable to an N-phase material. Each phase would have associated with it a G_n which is characteristic of the average shape of the cells with the property of the nth phase.

D. Determination of G_1 and G_2

The following methods of determining G_1 and/or G_2 for the effective permittivity case have been discussed[1]:

(1) Observation (i.e., look at the cross section of the sample).

(2) Small-perturbation analysis.

(3) Empirical measurement at small concentrations.

(4) Determination from defining equations of G where g_1 and g_2 are obtained from a mathematical experiment.

All of these methods are again valid for determining G_1 and G_2. Since the G's have the same geometrical significance, we may find it more convenient to determine the G's by empirical measurements of one of the effective properties discussed[1] (e.g., electric permittivity, thermal conductivity) rather than the effective bulk modulus.

ACKNOWLEDGMENT

The author is indebted to Professor Mark Beran for his guidance and encouragement and for giving so freely of his intuition and creativity which contributed so much to this and the preceding article.

APPENDIX

The two three-point correlation functions defined in Sec. 2 are

$$\Gamma_u \equiv \langle \mu'(0)K'(\mathbf{r})K'(\mathbf{s})\rangle \tag{A1}$$

and

$$\Gamma_l \equiv \langle K'(\mathbf{r})K'(\mathbf{s})/\mu(0)\rangle. \tag{A2}$$

For a two-phase cell-like material we can write Eq. (A1) in the alternate form

$$\begin{aligned}\Gamma_u = {} & \mu_1' K_1'^2 \bar{g}_1 + \mu_2' K_2'^2 \bar{g}_2 + \mu_1' K_1'^2[\bar{h}_{11} + \bar{h}_{11}^1 + \bar{h}_{11}^{11}] \\ & + \mu_1' K_1' K_2'[\bar{h}_{12} + \bar{h}_{12}^1] \\ & + \mu_2' K_1'^2 \bar{h}_{12}^{11} + \mu_2' K_2' K_1'[\bar{h}_{21} + \bar{h}_{21}^1] \\ & + \mu_1' K_2'^2 \bar{h}_{21}^{11} + \mu_2' K_2'^2[\bar{h}_{22} + \bar{h}_{22}^1 + \bar{h}_{22}^{11}] \\ & + \mu_1' K_1'^2 Z_{111} + \mu_1' K_1' K_2'[Z_{112} + Z_{121}] \\ & + \mu_2' K_1'^2 Z_{211} + \mu_2' K_2' K_1'[Z_{221} + Z_{212}] \\ & + \mu_1' K_2'^2 Z_{122} + \mu_2' K_2'^2 Z_{222},\end{aligned} \tag{A3}$$

where

$$\mu_1' = \mu_1 - \langle \mu \rangle, \quad \mu_2' = \mu_2 - \langle \mu \rangle,$$
$$K_1 = K_1' - \langle K \rangle, \quad K_2 = K_2' - \langle K \rangle.$$

$\bar{g}_n = \bar{g}_n(0, \mathbf{r}, \mathbf{s})$ is the probability that three points $(0, \mathbf{r}, \mathbf{s})$ are in the same cell with material property ϵ_n;

$h_{nm} = h_{nm}(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(0, \mathbf{r})$ are in the same cell with material property ϵ_n and a point $(\mathbf{s})$ is in another cell with material property ϵ_m; $h^1_{nm} = h^1_{nm}(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(0, \mathbf{s})$ are in the same cell with material property ϵ_n and a point $(\mathbf{r})$ is in another cell with material property ϵ_m; $h^{11}_{nm} = h^{11}_{nm}(0, \mathbf{r}, \mathbf{s})$ is the probability that two points $(\mathbf{r}, \mathbf{s})$ are in the same cell with material property ϵ_n and a point (0) is in another cell with material property ϵ_m; and $Z_{nmp} = Z_{nmp}(0, \mathbf{r}, \mathbf{s})$ is the probability that three points are in three different cells where points 0, $\mathbf{r}$, and $\mathbf{s}$ are in cells with material properties ϵ_n, ϵ_m, and ϵ_p, respectively.
Also,

$$\begin{aligned} \bar{g}_1 &= \varphi g_1, \\ \bar{g}_2 &= (1 - \varphi) g_2, \end{aligned} \tag{A4}$$

where $g_n = g_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that all three points are in the same cell of material property ϵ_n, given that one of the points is in a cell with material property ϵ_n.
We assume

$$\begin{aligned} h_{11} &= \varphi^2 h_1, & h_{12} &= \varphi(1 - \varphi) h_1, \\ h_{21} &= (1 - \varphi)\varphi h_2, & h_{22} &= (1 - \varphi)^2 h_2, \\ h^1_{11} &= \varphi^2 h^1_1, & h^1_{12} &= \varphi(1 - \varphi) h^1_1, \\ h^1_{21} &= (1 - \varphi)\varphi h^1_2, & h^1_{22} &= (1 - \varphi)^2 h^1_2, \\ h^{11}_{11} &= \varphi^2 h^{11}_1, & h^{11}_{21} &= \varphi(1 - \varphi) h^{11}_1, \\ h^1_{21} &= (1 - \varphi)\varphi h^{11}_2, & h^1_{22} &= (1 - \varphi)^2 h^{11}_2. \end{aligned} \tag{A5}$$

This assumption implies that the material property of a cell is statistically independent of the material property of any other cell.
Also,

$$\begin{aligned} &Z_{111} = \varphi^3 Z, \quad Z_{112} = Z_{211} = Z_{121} = \varphi^2(1 - \varphi) Z, \\ &Z_{122} = Z_{212} = Z_{221} = (1 - \varphi)^2 Z, \quad Z_{222} = (1 - \varphi)^3 Z, \end{aligned} \tag{A6}$$

where $h_n = h_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(0, \mathbf{r})$ are in the same cell and point $(\mathbf{s})$ is in another cell, given that one point (0 or $\mathbf{r}$) is in a cell of property ϵ_n; $h^1_n = h^1_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(0, \mathbf{r})$ are in the same cell and point $(\mathbf{r})$ is in another cell, given that one point (0 or $\mathbf{s}$) is in a cell of property ϵ_n. $h^{11}_n = h^{11}_n(0, \mathbf{r}, \mathbf{s})$ is the conditional probability that two points $(\mathbf{r}, \mathbf{s})$ are in the same cell and point (0) is in another cell, given that one point ($\mathbf{r}$ or $\mathbf{s}$) is in a cell of property ϵ_n; and $Z = Z(0, \mathbf{r}, \mathbf{s})$ is the probability that three points are in different cells.

By taking into account all possible ways that three points can fall into a cell material,

$$\begin{aligned} Z = 1 &- \bar{g}_1 - \bar{g}_2 - h_{11} - h^1_{11} - h^{11}_{11} - h_{12} - h^1_{12} \\ &- h^{11}_{12} - h_{21} - h^1_{21} - h^{11}_{21} - h_{22} - h^1_{22} - h^{11}_{22} \end{aligned} \tag{A7}$$

or

$$\begin{aligned} Z = 1 &- \varphi g_1 - (1 - \varphi) g_2 - \varphi(h_1 + h^1_1 + h^{11}_1) \\ &- (1 - \varphi)(h_2 + h^1_2 + h^{11}_2). \end{aligned} \tag{A8}$$

Making the same substitutions as for γ_{3I} in Miller,[1] we can reduce Eq. (A3) to

$$\Gamma_u = \mu'_1 K'^2_1 \varphi\{g_1 - [\varphi^2/(1 - \varphi)^2] g_2\}. \tag{A9}$$

For a symmetric-cell material this reduces further to

$$\Gamma_u = \mu'_1 K'^2_1 \frac{\varphi(1 - 2\varphi)}{(1 - \varphi)^2} g. \tag{A10}$$

Similarly, Eq. (A2) may be written in the form

$$\begin{aligned} \Gamma_l = &\frac{K'^2_1}{\mu_1} \bar{g}_1 + \frac{K'^2_2}{\mu_2} \bar{g}_2 + \frac{K'^2_1}{\mu_1}(h_{11} + h^1_{11} + h^{11}_{11}) \\ &+ \frac{K'_1 K'_2}{\mu_1}(h_{12} + h^1_{12}) + \frac{K'^2_1}{\mu_2} h^{11}_2 \\ &+ \frac{K'_2 K'_1}{\mu_2}(h_{21} + h^1_{21}) + \frac{K'^2_2}{\mu_1} h^{11}_2 \\ &+ \frac{K'^2_2}{\mu_2}(h_{22} + h^1_{22} + h^{11}_{22}) + \frac{K'^2_1}{\mu_1} Z_{111} + \frac{K'^2_1}{\mu_2} Z_{211} \\ &+ \frac{K'_1 K'_2}{\mu_1}(Z_{121} + Z_{112}) + \frac{K'_1 K'_2}{\mu_2}(Z_{212} + Z_{221}) \\ &+ \frac{K'^2_2}{\mu_1} Z_{122} + \frac{K'^2_2}{\mu_2} Z_{222}. \end{aligned} \tag{A11}$$

Again making substitutions as for γ_{3J} in Miller,[1] we can reduce Eq. (A11) to

$$\begin{aligned} \Gamma_l = &\frac{K'^2_1}{\mu_1} \frac{\varphi}{(1 - \varphi)} \{(\beta - 1)[\varphi^2 g_2 - (1 - \varphi)^2 g_1] \\ &+ [\beta + \varphi(1 - \beta)][(1 - \varphi) f^{11}_1 + \varphi f^{11}_2]\}, \end{aligned} \tag{A12}$$

where

$$\beta = \mu_1/\mu_2.$$

For a symmetric-cell material we can further reduce Eq. (A12) to

$$\begin{aligned} \Gamma_l = &\frac{K'^2_1}{\mu_1} \frac{\varphi}{(1 - \varphi)} \{(\beta - 1)(2\varphi - 1) g \\ &+ [\beta + \varphi(1 - \beta)] f^{11}\}. \end{aligned} \tag{A13}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Spectral Representation of the Covariant Two-Point Function and Infinite-Component Fields with Arbitrary Mass Spectrum

I. T. TODOROV*
International Center for Theoretical Physics, Trieste, Italy
AND
R. P. ZAIKOV
Joint Institute for Nuclear Research, Dubna, USSR

(Received 3 March 1969)

The general form of a Lorentz covariant two-point function is written down in momentum space as an expansion in terms of the total spin eigenfunctions. It leads naturally to a local expression for finite-component fields, but incorporates nonlocal infinite-component fields. Explicit examples of such fields with increasing mass spectrum are constructed. The two-point function is shown to fall exponentially for large spacelike separations provided that the lowest mass in the theory is positive.

INTRODUCTION

Usually, when one is dealing with fields describing particles of definite spin, the spectral representation of the two-point function is given by an integral in the total mass.[1,2] If, however, a field contains more than one spin, then it is necessary to decompose the two-point function with respect also to the spin variable (i.e., with respect to the second Casimir operator of the Poincaré group). This is always the case for infinite-component fields. We find the form of such a representation for a field transforming under an arbitrary irreducible representation of $SL(2, C)$ (the result applies also to more general fields which can be decomposed into irreducible components). The clue to the solution of the problem is the use of the formalism of homogeneous functions of two complex variables (instead of the tensor formalism) in the description of the irreducible representations of the Lorentz group (see, e.g., Ref. 3 or Ref. 4, Appendix A).

For finite-component fields, Poincaré invariance and spectral conditions imply the equivalence between *TCP* and weak local commutativity.[5] For infinite-component fields this is not the case.[6] In Sec. 3 we construct explicit examples of nonlocal *TCP*-invariant generalized free infinite-component fields with an increasing mass spectrum (with respect to spin). If the lowest mass in the theory is positive, then the two-point function (and hence both the commutator and the anticommutator) decrease exponentially for large spacelike separations.

1. GENERAL FORM OF THE COVARIANT TWO-POINT FUNCTION

A field $\psi(x)$ transforming according to the irreducible representation[7] $[l_0, l_1]$ of $SL(2, C)$ can be written down as a homogeneous function $\psi(x, z)$ of the complex (Lorentz) spinor $z = (z_1, z_2)$ (cf. Ref. 8):

$$\psi(x, \lambda z) = \lambda^{\nu_1}\bar{\lambda}^{\nu_2}\psi(x, z), \tag{1.1}$$

where the degree of homogeneity (ν_1, ν_2) is related to the number of the representation $[l_0, l_1]$ by

$$\nu_1 = l_1 + l_0 - 1, \quad \nu_2 = l_1 - l_0 - 1 \tag{1.2}$$

(we recall that the single-valuedness of the ψ implies that $\nu_1 - \nu_2 = 2l_0$ is an integer). For the special case of finite-dimensional representations (for ν_1, ν_2 non-negative integers) $\psi(x, z)$ is by definition a polynomial of z and $\bar{z}$, its coefficients being the ordinary (spinor or tensor) field components. In particular, the Pauli two-component spinor φ and the vector field A

* Present address: Institute for Advanced Study, Princeton; on leave of absence from Joint Institute for Nuclear Research, Dubna, USSR, and from Physical Institute of the Bulgarian Academy of Sciences, Sofia, Bulgaria.

[1] S. Kamefuchi and H. Umezawa, Progr. Theoret. Phys. (Kyoto) **6**, 543 (1951); G. Källén, Helv. Phys. Acta **25**, 417 (1952); H. Lehmann, Nuovo Cimento **11**, 342 (1954); O. Steinmann, J. Math. Phys. **4**, 583 (1963).

[2] I. Rasziller and D. H. Schiller, Nuovo Cimento **A48**, 617, 635, 645 (1967).

[3] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Vol. 5: Integral Geometry and Representation Theory* (Academic Press Inc., New York, 1966); *Generalized Functions, Properties and Operations* (Academic Press Inc., New York, 1964), Appendix B.

[4] D. Tz. Stoyanov and I. T. Todorov, J. Math. Phys. **9**, 2146 (1968).

[5] R. Jost, Helv. Phys. Acta **30**, 409 (1957); see also R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (W. A. Benjamin and Co., New York, 1964).

[6] A. I. Oksak and I. T. Todorov, Commun. Math. Phys. **11**, 125 (1968).

[7] We are using the notation of I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro [*Representations of the Rotation and Lorentz Group and Their Applications* (Pergamon Press Ltd., London, 1963)] for the irreducible representations of $SL(2, C)$. M. A. Naimark [*Linear representations of the Lorentz Group* (Pergamon Press Ltd., London, 1964)], denotes them by $[k_0, c]$.

[8] Dao Vong Dyc and Nguyen Van Hieu, Yad. Fiz. **6**, 186 (1967) [Sov. J. Nucl. Phys. **6**, 137 (1968)].

correspond in our notation to the polynomials

$$\varphi(x, z) = \sum_{\alpha=1}^{2} \varphi^{\alpha}(x) z_{\alpha},$$

$$A(x, z) = \sum_{\mu=0}^{3} A^{\mu}(x) z \sigma_{\mu} \bar{z}$$

(further, the summation sign in similar expressions with repeated upper and lower indices will be omitted).

The relativistic transformation law for ψ has the form

$$U(a, A)\psi(x, z)U^{-1}(a, A) = \psi(\Lambda x + a; zA^{-1}), \quad (1.3)$$

where $A \in SL(2, C)$ and $\Lambda = \Lambda(A)$ is the proper Lorentz transformation defined by

$$A\sigma_{\nu}A^{*} = \sigma_{\mu}\Lambda^{\mu}_{\nu}$$

(σ_0 is the 2×2 unit matrix, the $\sigma_j, j = 1, 2, 3$ are the Pauli matrices).

Consider the two-point function

$$F_{\varphi\psi}(x - y; z, w) = \langle 0 | \varphi(x, z)\psi(y, w) | 0 \rangle, \quad (1.4)$$

where φ and ψ are transforming under the representations $[l_0, l_1]$ and $[l_0', l_1']$, respectively. The spectral conditions allow us to write F in the form

$$F_{\varphi\psi}(x; z, w) = \int \vartheta(p) K(\underset{\sim}{p}; z, w) l^{-ipx} d^4p, \quad (1.5)$$

where $\vartheta(p) \equiv \theta(p^0)\theta(p^2)$ is the characteristic function of the forward cone and

$$\underset{\sim}{p} = \sigma_{\mu}p^{\mu} = \begin{pmatrix} p^0 + p^3 & p^1 - ip^2 \\ p^1 + ip^2 & p^0 - p^3 \end{pmatrix}. \quad (1.6)$$

Lorentz invariance implies that, for each $A \in SL(2, C)$,

$$K(\underset{\sim}{ApA}^{*}; zA^{-1}, wA^{-1}) = K(\underset{\sim}{p}; z, w). \quad (1.7)$$

Furthermore, because of (1.1), K is a homogeneous function of z and w of degree (ν_1, ν_2) and (ν_1', ν_2'), respectively.

To satisfy (1.7) we require[9] that the kernel K is a function of the invariants, which can be formed by the 4-vector p and the Lorentz spinors z and w.

First we find a complete set of independent invariants.

One can construct four zero-length vectors in terms of z and w; one obtains two real lightlike vectors

$$\xi_{\mu} = z\sigma_{\mu}\bar{z} \quad \text{and} \quad \eta_{\mu} = w\sigma_{\mu}\bar{w} \quad (1.8)$$

and two complex conjugate vectors

$$\chi_{\mu} = z\sigma_{\mu}\bar{w} \quad \text{and} \quad \bar{\chi}_{\mu} = w\sigma_{\mu}\bar{z}. \quad (1.9)$$

[9] Actually, this requirement is not independent. We shall prove elsewhere that it follows rigorously from the invariance assumption.

The equalities $\xi^2 = \eta^2 = \chi^2 = 0$ are a consequence of the identity

$$g^{\mu\mu}(\sigma_{\mu})^{a\dot{b}}(\sigma_{\mu})^{c\dot{d}} = 2\epsilon^{ac}\epsilon^{\dot{b}\dot{d}}, \quad \epsilon = i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (1.10)$$

All nonvanishing scalar products of these vectors are proportional to the product of the complex conjugate invariants $\kappa = z\epsilon w$ and $\bar{\kappa}$:

$$\xi\eta = -\chi\bar{\chi} = 2|\kappa|^2, \quad \xi\chi = \eta\chi = \xi\bar{\chi} = \eta\bar{\chi} = 0. \quad (1.11)$$

More generally, the tensor identity holds:

$$\chi_{\mu}\bar{\chi}_{\nu} + \bar{\chi}_{\mu}\chi_{\nu} = \xi_{\mu}\eta_{\nu} + \eta_{\mu}\xi_{\nu} - g_{\mu\nu}\xi\eta; \quad (1.12)$$

it implies, in particular, that

$$(\xi p)(\eta p) - (\chi p)(\bar{\chi} p) = \tfrac{1}{2}p^2\xi\eta = p^2|\kappa|^2. \quad (1.13)$$

Assuming $l_0 \geq |l_0'|$ (which can be achieved without loss of generality by a possible interchange $z \leftrightarrow w$) and taking into account the above identities, we find the following general form for the invariant kernel K (see Appendix A):

$$K(p; z, w) = \kappa^{l_0+l_0'}(\chi p)^{l_0-l_0'}(p\xi)^{l_1-l_0-1} \times (p\eta)^{l_1'-l_0-1}h(\cos\theta; p^2), \quad (1.14)$$

where

$$\cos\theta = 1 - p^2\xi\eta/(p\xi)(p\eta) \quad (1.15)$$

(θ is the angle between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ in the rest frame of p). The kernel (1.14) is single-valued if and only if $l_0 \pm l_0'$ are integers, which will be always assumed.

2. DECOMPOSITION OF THE INVARIANT KERNEL WITH RESPECT TO SPIN

The kernel (1.14) describes, in general, propagation of particles of different spins (provided that the representation of $SL(2, C)$ for the field is not of the form $[l_0, |l_0| + 1]$). We shall now find the invariant kernels which are eigenfunctions of the spin-square operator

$$\mathbf{S}^2 = \tfrac{1}{2}M_{\sigma\rho}M^{\sigma\rho} - M_{\sigma\mu}M^{\sigma\nu}p^{\mu}p_{\nu}p^{-2} = l_0^2 + l_1^2 - 1 - p^{-2}M_{\sigma\mu}p^{\mu}M^{\sigma\nu}p_{\nu}. \quad (2.1)$$

Here $M^{\mu\nu}$ are the generators of one of the representations $[l_0, l_1]$ or $[l_0', l_1']$ under consideration. In terms of z, $M^{\mu\nu}$ can be written as

$$M^{\mu\nu} = \left(z_1, z_2, -\frac{\partial}{\partial\bar{z}_1}, -\frac{\partial}{\partial\bar{z}_2}\right)\frac{i}{4}[\gamma^{\mu}, \gamma^{\nu}]\begin{pmatrix} \partial/\partial z_1 \\ \partial/\partial z_2 \\ \bar{z}_1 \\ \bar{z}_2 \end{pmatrix}, \quad (2.2)$$

where γ^μ are the Dirac matrices in a basis in which

$$\gamma^\mu = \begin{pmatrix} 0 & g^{\mu\mu}\sigma_\mu \\ \sigma_\mu & 0 \end{pmatrix}, \quad \gamma^5 = -i\begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}$$

(see Ref. 4). As far as we are looking for Lorentz-invariant solutions [of the form (1.14)] of the equation

$$[\mathbf{S}^2 - s(s+1)]K(\underset{\sim}{p}; z, w) = 0, \tag{2.3}$$

it is convenient to go to the rest frame ($\underset{\sim}{p} = 0$) in which

$$\begin{aligned}\mathbf{S}^2 = \mathbf{M}^2 &= \tfrac{1}{2}M_{ij}M^{ij} = \frac{1}{4}\left(z\sigma_i\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar z}\sigma_i\bar z\right)^2 \\ &= \frac{1}{4}\left(z\frac{\partial}{\partial z} - \bar z\frac{\partial}{\partial \bar z}\right)^2 \\ &\quad - \frac{1}{2}\left\{\left(z\epsilon\frac{\partial}{\partial \bar z}\right)\left(\bar z\epsilon\frac{\partial}{\partial z}\right) + \left(\bar z\epsilon\frac{\partial}{\partial z}\right)\left(z\epsilon\frac{\partial}{\partial \bar z}\right)\right\}.\end{aligned} \tag{2.4}$$

[Acting on a homogeneous function, the first term on the right-hand side of (2.4) gives l_0^2.] Substituting (2.4) and (1.14) (with $\underset{\sim}{p} = 0$) in (2.3), we obtain the following equation for h:

$$\left\{\sin^2\theta\frac{d^2}{(d\cos\theta)^2} - 2[l_0' + (l_0+1)\cos\theta]\frac{d}{d\cos\theta} + [s(s+1) - l_0(l_0+1)]\right\}h(\cos\theta; p^2) = 0. \tag{2.5}$$

The general solution of this equation, regular for $\cos^2\theta = 1$, is

$$h_s(\cos\theta; p^2) = \rho_s(p^2)P_{s-l_0}^{(l_0+l_0', l_0-l_0')}(\cos\theta), \tag{2.6}$$

where the $P_n^{(\alpha,\beta)}(x)$ are the Jacobi polynomials ($l_0 \geq |l_0'|$).

We would arrive at the same result starting with an $\mathbf{S}^2$ in which $M^{\mu\nu}$ are the generators of the representation $[l_0', l_1']$ expressed in terms of w and $\partial/\partial w$ (this is a consequence of the spin conservation and can be checked directly).

In general, $h(\cos\theta, p^2)$ is a superposition of the spin eigenfunctions (2.6); s takes a finite number of values if at least one of the representations $[l_0, l_1]$ and $[l_0', l_1']$ is finite-dimensional; otherwise it assumes all values of the form $l_0 + n$ where n is a nonnegative integer.

In the important special case where $\psi = \varphi^*$ (and, hence, $l_0 = -l_0' \geq 0$, $l_1' = l_1$) the corresponding decomposition of the invariant kernel reads

$$\begin{aligned}K(\underset{\sim}{p}; z, w) &= \sum_{s\geq l_0}\rho_s(p^2)K_s(\underset{\sim}{p}; z, w) \\ &= (z\underset{\sim}{p}\tilde w)^{2l_0}(z\underset{\sim}{p}\bar z)^{l_1-l_0-1}(w\underset{\sim}{p}\tilde w)^{l_1-l_0-1} \\ &\quad\times\sum_{s\geq l_0}\rho_s(p^2)P_{s-l_0}^{(0,2l_0)}(\cos\theta).\end{aligned} \tag{2.7}$$

We point out that the kernels K_s defined by (2.7) are positive-definite in this case because

$$K_s(p; z, w) = A\sum_{\zeta=-s}^{s} u_{s\zeta}(p, z)\overline{u_{s\zeta}(p, w)}, \tag{2.8}$$

where A is a positive constant,

$$\begin{aligned}u_{s\zeta}(p, z) &= f_{s\zeta}(zB_p), \\ B_p &= p^0 + m + \sigma_j p^j/[2m(p^0+m)]^{\frac{1}{2}},\end{aligned} \tag{2.9}$$

and the $f_{s\zeta}$ are the canonical basis vectors (see Appendix B).

For finite-dimensional representations of $SL(2, C)$, $l_1 - l_0$ is a positive integer and K_s is a polynomial with respect to p. In that case, the (weak) locality condition for the two-point function is equivalent to *TCP* invariance,[5] which implies[6]

$$\langle 0|\,\varphi(x, z)\varphi(y, w)^*\,|0\rangle = \langle 0|\,\varphi(x, w)^*\varphi(y, z)\,|0\rangle. \tag{2.10}$$

For instance, for a local tensor field $\psi([l_0, l_1] = [0, n+1])$ Eq. (2.7) reduces to

$$\begin{aligned}K_{\psi\psi^*}(p; \xi, \eta) &= [(p\xi)(p\eta)]^n\sum_{s=0}^{n}\rho_s(p^2)P_s(\cos\theta) \\ &= \sum_{s=0}^{n}\rho_s(p^2)K_s^{\mu_1\cdots\mu_n\nu_1\cdots\nu_n}(p)\xi_{\mu_1}\cdots\xi_{\mu_n}\eta_{\nu_1}\cdots\eta_{\nu_n} \\ &= K_{\psi^*\psi}(p; \eta, \xi),\end{aligned} \tag{2.11}$$

where $K_s^{\mu_1\cdots\nu_n}$ is a homogeneous polynomial of p of degree $2n$ obtained from

$$\begin{aligned}R_s^{\mu_1\cdots\mu_n\nu_1\cdots\nu_n}(p) &= \sum_{k=0}^{s}\frac{(s+k)!}{(k!)^2(s-k)!}\left(-\frac{p^2}{2}\right)^k \\ &\quad\times g^{\mu_1\nu_1}\cdots g^{\mu_k\nu_k}p^{\mu_{k+1}}\cdots p^{\mu_n}p^{\nu_{k+1}}\cdots p^{\nu_n}\end{aligned} \tag{2.12}$$

by a symmetrization with respect to $\mu_1\cdots\mu_n$ and $\nu_1\cdots\nu_n$ and by a symmetric subtraction of traces, in a way such that

$$g_{\mu_1\mu_2}K_s^{\mu_1\mu_2\cdots\mu_n\nu_1\cdots\nu_n} = g_{\nu_1\nu_2}K_s^{\mu_1\cdots\mu_n\nu_1\nu_2\cdots\nu_n} = 0.$$

Taking into account that ξ and η defined by (1.8) are lightlike vectors, we see that both $K_s^{\mu_1\cdots\nu_n}$ and $R_s^{\mu_1\cdots\nu_n}$ lead to the same expression (2.11) for $K_{\psi\psi^*}$.

For infinite-component fields the locality of the two-point function is rather an exception. It implies that K is a polynomial with respect to p^μ with coefficients depending on p^2 only.[10] As an example of a

[10] N. N. Bogoliubov and V. S. Vladimirov, Nauchn. Dokl. Vyschei Shkoly, Fiz. Mat. Nauk. **1**, No. 3, 26 (1958); **2**, No. 2, 179 (1959).

local infinite-component field we take the free field $\psi(x, z)$ of mass m, transforming under some unitary representation $[l_0, i\sigma]$ of $SL(2, C)$ (of the principal series). In this case

$$K_{\psi\psi^*}(p; z, w) = \left(\frac{zp\bar{w}}{w\underset{\sim}{p}\bar{z}}\right)^{l_0}\left(\frac{z\underset{\sim}{p}\bar{z}}{w\underset{\sim}{p}\bar{w}}\right)^{i\sigma}\delta(z\epsilon w)\delta(p^2 - m^2), \tag{2.13}$$

where $\delta(z\epsilon w)$ is a two-dimensional δ function which is a homogeneous function of degree $(-1, -1)$ of its complex argument (see Ref. 3). In this case the first two factors do not depend in fact on p because of the first δ function. For real self-conjugate representations ($l_0 = 0$, l_1 real), the free local field φ of mass m corresponds to

$$K_{\varphi\varphi^*}(p; \xi, \eta) = (\xi\eta)^{l_1-1}\delta(p^2 - m^2). \tag{2.14}$$

Both examples (2.13) and (2.14) correspond to infinite-mass degeneracy with respect to spin. This is not accidental. Grodsky and Streater showed recently[11] that an infinite-component field with polynomially bounded two-point function in momentum space can be local only for infinite-mass degeneracy.

3. EXAMPLES OF NONLOCAL INFINITE-COMPONENT FIELDS WITH INCREASING MASS SPECTRUM

In agreement with the above-mentioned general result,[11] all examples of infinite-component fields with a nontrivial mass spectrum, obtained by specializing the coefficients $\rho_s(p^2)$ in (2.7), correspond to nonlocal theories. However, if the lowest mass of the theory is positive, i.e., if all $\rho_s(p^2)$ vanish for $p^2 < m_0^2$ (> 0), then, just as in conventional theory of finite-component fields, the two-point function (1.5) goes to zero as $r^\lambda e^{-m_0 r}$ (with some real λ) for $r^2 = -(x - y)^2 \to \infty$ (see Appendix C). The same is true (as a consequence) for the vacuum expectation values of both the commutator and the anticommutator of φ and φ^* (it would be interesting to find some more subtle criterion for locality which would permit us to distinguish between those two).

We shall give here a class of examples of generalized free nonlocal fields with increasing mass spectrum in which the sum over s can be taken explicitly.

For that purpose we put

$$\rho_s(p) = \frac{NC^{s-l_0}}{m_s^{l_1+l_1}}\delta(p^2 - m_s^2), \quad m_s = m_0 + a(s - l_0),$$
$$N > 0, \quad 0 < C \leq 1, \quad m_0 > 0, \quad a > 0. \tag{3.1}$$

Substituting (2.7) with this ρ_s in (1.5) and putting in the $(s - l_0)$th term of the sum $p = m_s n$, we obtain

$$\begin{aligned} F_{\varphi\varphi^*}(x; z, w) &= N\int_{n^0=(1+\mathbf{n}^2)^{\frac{1}{2}}} Qe^{-im_0 nx} \\ &\quad \times \sum_{s=l_0}^{\infty}(ce^{-ianx})^{s-l_0}P_{s-l_0}^{(0,2l_0)}\frac{d^3\mathbf{n}}{2n^0} \\ &= N2^{2l_0}\int_{n^0=(1+\mathbf{n}^2)^{\frac{1}{2}}} Qe^{-im_0(nx)} \\ &\quad \times \frac{1}{R}[R + ce^{-ia(nx)} + 1]^{-2l_0}\frac{d^3\mathbf{n}}{2n^0} \\ &= F_{\varphi^*\varphi}(x; w, z), \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} R &= \{1 - 2c\cos\theta e^{-ia(nx)} + c^2\exp[-2ia(nx)]\}^{\frac{1}{2}}, \\ Q &= (z\underset{\sim}{n}\bar{w})^{2l_0}(z\underset{\sim}{n}\bar{z})^{l_1-l_0-1}(w\underset{\sim}{n}\bar{w})^{l_1-l_0-1}, \end{aligned} \tag{3.3}$$

and the range of integration is the upper unit hyperboloid $n^0 = (1 + \mathbf{n}^2)^{\frac{1}{2}}$. All other two-point functions (as well as all truncated Wightman functions) are zero by definition. The field $\varphi(x, z)$ so defined is *TCP*-invariant but not weakly local. For the mass-degenerate limit $a = 0$, if we put $c = 1$, $m_0 = m$, $l_0 = 0$, and $l_1 = \frac{1}{2}$, we find a local field of the type (2.14).

It would be interesting to analyze the implications of locality on the two-point function in the case when it is a Jaffe-type generalized function,[12] and to find whether the Grodsky–Streater "no-go theorem" [11] can be extended to this case also. We hope that a further investigation of the examples provided by the representation (1.5) and (2.7) may give some hint for the solution of this problem.

ACKNOWLEDGMENTS

The authors would like to thank Professor D. I. Blokhintsev and Dr. D. Tz. Stoyanov for interesting discussions, and Professor R. F. Streater for a prepublication copy of his paper.

The first-named author (I. T. T.) is grateful to Professor Abdus Salam and Professor P. Budini and to the IAEA for the hospitality kindly extended to him at the International Center for Theoretical Physics, Trieste, where the present paper was completed.

APPENDIX A: DERIVATION OF THE REPRESENTATION (1.14) FOR THE INVARIANT KERNEL

We first determine the general form of the invariant monomial

$$M = \kappa^{a_1}\bar{\kappa}^{a_2}(\chi p)^{b_1}(\bar{\chi}p)^{b_2}(\xi p)^c(\eta p)^d \tag{A1}$$

[11] I. T. Grodsky and R. F. Streater, Phys. Rev. Letters **20**, 695 (1968); see also H. D. I. Abarbanel and Y. Frishman, Phys. Rev. **171**, 1442 (1968), where a similar conclusion has been drawn.

[12] A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).

of degree of homogeneity (ν_1, ν_2) [Eq. (1.2)] with respect to z and (ν_1', ν_2') with respect to w. Recalling the definition of κ, ξ, η, χ in terms of z and w [see (1.8), (1.9)], we find the following set of equations for the exponents in (A1):

$$\begin{aligned} a_1 + b_1 + c &= \nu_1 = l_1 + l_0 - 1, \\ a_1 + b_2 + d &= \nu_1' = l_1' + l_0' - 1, \\ a_2 + b_2 + c &= \nu_2 = l_1 - l_0 - 1, \\ a_2 + b_1 + d &= \nu_2' = l_1' - l_0' - 1. \end{aligned} \tag{A2}$$

The general solution of the system (A2) may be expressed in terms of two arbitrary parameters k and λ:

$$\begin{aligned} a_1 &= l_0 + l_0' + k, & a_2 &= k, \\ b_1 &= \lambda + l_0 - l_0' - k, & b_2 &= \lambda - k, \\ c &= l_1 - l_0 - 1 - \lambda, & d &= l_1' - l_0 - 1 - \lambda. \end{aligned} \tag{A3}$$

Substituting (A3) in (A1) and using (1.13) and (1.15), we find

$$\begin{aligned} M &= \kappa^{l_0+l_0'}(\chi p)^{l_0-l_0'}(\xi p)^{l_1-l_0-1} \\ &\qquad (\eta p)^{l_1'-l_0-1}\left(\frac{\kappa\bar{\kappa}}{\chi p \cdot \bar{\chi} p}\right)^k\left(\frac{\chi p \cdot \bar{\chi} p}{\xi p \cdot \eta p}\right)^\lambda \\ &= \kappa^{l_0+l_0'}(\chi p)^{l_0-l_0'}(\xi p)^{l_1-l_0-1}(\eta p)^{l_1'-l_0-1} \\ &\quad \times [2^\lambda (p^2)^k]^{-1}(1 - \cos\theta)^k(1 + \cos\theta)^{\lambda-k}. \end{aligned} \tag{A4}$$

The homogeneous invariant kernel K is a superposition of such monomials and hence has the form (1.14).

We mention that the scalar products $p\xi$ and $p\eta$ are always positive (for $\xi_0 = |\boldsymbol{\xi}|$, $\eta_0 = |\boldsymbol{\eta}|$, $p^0 > |\mathbf{p}|$) so that any (complex) power of these products is well defined and regular in the domain of integration in (1.5). This is not the case for the first two factors in (A4) [or (1.14)]. That is the reason why we ask the exponents of κ and χp to be positive and apply (1.14) only for $l_0 \geq |l_0'|$. If we had instead, for instance, $l_0' \geq |l_0|$, we should put $\lambda = l_0' - l_0 + \lambda'$ and rewrite (A4) in the form

$$\begin{aligned} M &= \kappa^{l_0+l_0'}(\chi p)^{l_0'-l_0}(\xi p)^{l_1-l_0'-1}(\eta p)^{l_1'-l_0'-1} \\ &\qquad \times \frac{(1 - \cos\theta)^k}{2^{\lambda'}(p^2)^k}(1 + \cos\theta)^{\lambda'-k}, \end{aligned} \tag{A5}$$

implying a corresponding modification of (1.14).

APPENDIX B: EXPLICIT EXPRESSIONS FOR THE CANONICAL BASIS AND THE SPIN EIGENFUNCTIONS AND PROOF OF THE SUMMATION FORMULA (2.8)

The transition from the continuous variable $z = (z_1, z_2)$ to the discrete indices $s\zeta$ is given by the expressions of the vectors $|s\zeta) = f_{s\zeta}(z)$ of the canonical basis. We shall present them in a form which exhibits their relation to the matrix elements of the irreducible representations of $SU(2)$ and allows automatic summation over the spin projection to be performed [in particular, to prove (2.8)].

To do this we introduce generalized polar coordinates in the two-dimensional complex space putting

$$\begin{aligned} &z_1 = (r)^{\frac{1}{2}}\cos\tfrac{1}{2}\theta e^{\frac{1}{2}i(\alpha+\varphi)}, \quad z_2 = (r)^{\frac{1}{2}}\sin\tfrac{1}{2}\theta e^{\frac{1}{2}i(\alpha-\varphi)}, \\ &r \geq 0, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \varphi < 2\pi, \quad 0 \leq \alpha < 4\pi. \end{aligned} \tag{B1}$$

It is easy to express the generators of $SL(2, C)$ in terms of these variables. The infinitesimal generators of the $SU(2)$ subgroup (the "compact generators") depend only on the angles

$$\begin{aligned} M^3 &= -i\frac{\partial}{\partial\varphi}, \\ M_\pm &= M^1 \pm iM^2 \\ &= e^{\pm i\varphi}\left(\pm\frac{\partial}{\partial\theta} + \frac{i}{\tan\theta}\frac{\partial}{\partial\varphi} - \frac{i}{\sin\theta}\frac{\partial}{\partial\alpha}\right); \end{aligned} \tag{B2}$$

the generators of the pure Lorentz transformations have the form

$$\begin{aligned} N^3 &= i\left(r\cos\theta\frac{\partial}{\partial r} - \sin\theta\frac{\partial}{\partial\theta}\right), \\ N_\pm &= N^1 \pm iN^2 = e^{\pm i\varphi}\left(ir\sin\theta\frac{\partial}{\partial r} + i\cos\theta\frac{\partial}{\partial\theta}\right. \\ &\qquad \left. \mp\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi} \pm \frac{1}{\tan\theta}\frac{\partial}{\partial\alpha}\right). \end{aligned} \tag{B3}$$

The canonical basis for any irreducible representation $[l_0, l_1]$ of $SL(2, C)$ is defined by the set of properly normalized eigenvectors $|s\zeta)$ of $\mathbf{M}^2$ and M^3 (see Refs. 4, 7):

$$\begin{aligned} &\mathbf{M}^2|s\zeta) = s(s+1)|s\zeta), \quad M^3|s\zeta) = \zeta|s\zeta), \\ &s = |l_0|, |l_0|+1, \cdots, \quad -s \leq \zeta \leq s. \end{aligned} \tag{B4}$$

The conventionally normalized[4] solution of Eqs. (B4) has the form

$$|s\zeta) = f^{l_0 l_1}_{s\zeta}(z) = A_s r^{l_1-1} D^{(s)}_{l_0\zeta}(\alpha, \theta, \varphi), \tag{B5}$$

where

$$\begin{aligned} A_s &= A^{l_0 l_1}_s \\ &= \left[(2s+1)\frac{\Gamma(s+1-l_1)\Gamma(|l_0|+1+l_1)}{\Gamma(|l_0|+1-l_1)\Gamma(s+1+l_1)}\right]^{\frac{1}{2}} \end{aligned} \tag{B6}$$

and $D^{(s)}_{l_0\zeta}$ are the matrix elements of the irreducible representation of $SU(2)$ (corresponding to spins) as functions of the Euler angles; they can be expressed in terms of the Jacobi polynomials as

$$\begin{aligned} D^{(s)}_{l_0\zeta}(\alpha, \theta, \varphi) &= \left[\frac{(s+\zeta)!(s-\zeta)!}{(s+l_0)!(s-l_0)!}\right]^{\frac{1}{2}} e^{i(l_0\alpha+\zeta\varphi)}(\tan\tfrac{1}{2}\theta)^{-\varepsilon_\zeta l_0} \\ &\quad \times (\tfrac{1}{2}\sin\theta)^{|\zeta|}P^{(|\zeta|-\varepsilon_\zeta l_0, |\zeta|+\varepsilon_\zeta l_0)}_{s-|\zeta|}(\cos\theta), \\ &\qquad \varepsilon_\zeta = \operatorname{sgn}\zeta = \zeta/|\zeta|. \end{aligned} \tag{B7}$$

Equation (B5) allows sum rules to be obtained for products of two basic vectors f:

$$\sum_{\zeta=-s}^{s} f_{s\zeta}^{l_0 l_1}(z_{(1)})\overline{f_{s\zeta}^{l_0' l_1'}(z_{(2)})} = A_s^{l_0 l_1}\overline{A_s^{l_0' l_1'}}\, r_1^{l_1-1}\overline{r_2^{l_1'-1}} \sum_{\zeta=-s}^{s} D_{l_0\zeta}^{(s)}(\alpha_1, \theta_1, \varphi_1)[D^{(s)}(\alpha_2, \theta_2, \varphi_2)]_{\zeta l 0}^{*}$$
$$= A_s^{l_0 l_1}\overline{A_s^{l_0' l_1'}}\, r_1^{l_1-1}\overline{r_2^{l_1'-1}} D_{l_0 l_0'}^{(s)}(\alpha, \theta, \varphi), \tag{B8}$$

where $(\alpha, \theta, \varphi)$ are the Euler angles corresponding to the product of rotations $(\alpha_1, \theta_1, \varphi_1)(\alpha_2, \theta_2, \varphi_2)^{-1}$:

$$\begin{aligned} \cos\theta &= \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\varphi_1 - \varphi_2) \\ e^{i(\alpha+\varphi)} &= \frac{\cos\frac{1}{2}(\theta_1 - \theta_2) + i\cos\frac{1}{2}(\theta_1 + \theta_2)\tan\frac{1}{2}(\varphi_1 - \varphi_2)}{\cos\frac{1}{2}(\theta_1 - \theta_2) - i\cos\frac{1}{2}(\theta_1 + \theta_2)\tan\frac{1}{2}(\varphi_1 - \varphi_2)} e^{i(\alpha_1-\alpha_2)}, \\ e^{i(\alpha-\varphi)} &= \frac{\sin\frac{1}{2}(\theta_1 - \theta_2) - i\sin\frac{1}{2}(\theta_1 + \theta_2)\tan\frac{1}{2}(\varphi_1 - \varphi_2)}{\sin\frac{1}{2}(\theta_1 - \theta_2) + i\sin\frac{1}{2}(\theta_1 + \theta_2)\tan\frac{1}{2}(\varphi_1 - \varphi_2)} e^{i(\alpha_1+\alpha_2)}. \end{aligned} \tag{B9}$$

Comparing (2.7) with (B8) and (B9), we check (2.8).

APPENDIX C: ASYMPTOTIC BEHAVIOR OF THE TWO-POINT FUNCTION FOR LARGE SPACELIKE SEPARATIONS

We shall only outline the derivation of the exponential decrease of the two-point function (1.15) (in a theory with a lowest positive mass m_0) without going into mathematical details.

To state the problem properly we have to pass from the continuous variables z and w to vector variables f and g belonging to the representation space [the function $F_{\varphi\psi}(x; z, w)$ is, in general, a distribution of z and w—it may not be defined for some special values of these parameters]. Namely, let the field $\psi(x, z)$ be transforming under the irreducible representation $\tau = [l_0, l_1]$, and let $f(z) \in D_\tau$, where D_τ is the set of homogeneous function of degree (1.2), infinitely differentiable in the whole complex space C_2 except the origin. Then, in general, only the linear functional

$$\psi(x, f) = \int f(z_1, 1)\psi(x; z_1, 1)\, d^2 z_1 (f \in D_{-\tau}) \tag{C1}$$

has a meaning [as an operator-valued distribution in $S(R_4)$].

In these terms the two-point function can be written in the following way:

$$\begin{aligned} F_{\varphi\varphi^*}(x; f, g) &= \langle 0|\, \varphi(\tfrac{1}{2}x, f)\varphi(-\tfrac{1}{2}x, g)^*\, |0\rangle \\ &= \sum_{s \geq l_0} \int_{m_0^2}^{\infty} d\sigma_s(m^2) \\ &\quad \times \int_{n_0=(1+\mathbf{n}^2)^{\frac{1}{2}}} K_s(n) e^{-imnx} \frac{d^3\mathbf{n}}{n_0}, \end{aligned} \tag{C2}$$

where, in accordance with (2.7),

$$\begin{aligned} K_s(\underset{\sim}{n}) = \iint & f(z_{(1)})(z_{(1)}\underset{\sim}{n}\bar{w}_{(1)})^{2l_0} \\ & \times (z_{(1)}\underset{\sim}{n}\bar{z}_{(1)})^{l_1-l_0-1}(w_{(1)}\underset{\sim}{n}\bar{w}_{(1)})^{l_1-l_0-1} \\ & \times P_{s-l_0}^{(0,2l_0)}(\cos\theta) g(w_{(1)})\, d^2 z_1\, d^2 w_1, \end{aligned} \tag{C3}$$

and $z_{(1)} = (z_1, 1)$, $w_{(1)} = (w_1, 1)$ [σ_s being related to ρ_s of (2.7)].

Now we put $x^0 = 0$ and choose the third axis along $\mathbf{x}$ [so that $\mathbf{x} = (0, 0, r)$, $r > 0$].

We see from (C3) that, for k sufficiently large,

$$\frac{\partial^k}{\partial n_3^k}\left(\frac{1}{n_0} K_s(n)\right)$$

is absolutely integrable. This permits us (after a repeated integration by parts with respect to n_3 and a subsequent integration in n_1 and n_2) to rewrite (C2) in the form

$$\begin{aligned} &F_{\varphi\varphi^*}(0, 0, 0, r; f, g) \\ &\quad = \sum_{s \geq l_0} \int_{m_0^2}^{\infty} \frac{d\sigma_s(m^2)}{(rm)^k} \int_{-\infty}^{\infty} dn_3 F_s(n_3) e^{imn_3 r} \end{aligned} \tag{C4}$$

with

$$F_s(n_3) = \iint \frac{\partial^k}{\partial n_3^k}\left(\frac{1}{n_0} K_s(n)\right) dn_1\, dn_2.$$

From the explicit expression of K_s (C3), we see that $F_s(n_3)$ is analytic in the strip

$$-i < n_3 < i$$

and may have a power-type singularity for $n_3 = i$, which we shall write in the form

$$C/(n_3 + i)^{\lambda+k+1}$$

(λ being real). This permits shifting the contour of integration in (C4) to $n_3 = i + \nu$; we finally obtain

$$\begin{aligned} &F_{\varphi\varphi^*}(0, 0, 0, r; f, g) \\ &\quad = r^\lambda e^{-m_0 r} \int_{m_0^2}^{\infty} e^{-(m-m_0)r} G_s(m^2)\, d\sigma_s(m^2), \end{aligned} \tag{C5}$$

where G_s is a continuous polynomially bounded function of m^2. This proves that, apart from a power of r, the two-point function decreases like $e^{-m_0 r}$ for large spacelike separations r.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Perturbation Method for a Nonlinear Wave Modulation. II

NARUYOSHI ASANO AND TOSIYA TANIUTI
Department of Physics, Nagoya University, Nagoya, Japan
AND
NOBUO YAJIMA
Research Institute for Fundamental Physics, Kyoto University, Kyoto, Japan

(Received 10 February 1969)

A perturbation method given in a previous paper of this series is applied to two physical examples, the electron plasma wave and a nonlinear Klein–Gordon equation. In these systems, and probably in most physical systems, an assumed condition for a mode of $l = 0$ is not valid. Consequently, the direct application of the method is impossible. In the present paper, we shall illustrate by these examples how this difficulty can be overcome to allow us to use the method. As a result we shall find that, in either case, the original equation can be reduced to the nonlinear Schrödinger equation.

1. INTRODUCTION

In a previous paper[1] of this series, we formulated a perturbation method to account for a modulation of plane waves due to nonlinear self-interactions. A system of equations, which we considered, was

$$\frac{\partial U}{\partial t} + A(U)\frac{\partial U}{\partial x} + B(U) = 0,$$

for a column vector U with n components, where the $n \times n$ matrix A and the column vector B are, in general, vector-valued functions of U, see Eq. (I.1). [(I.1) denotes Eq. (1) in I. In what follows, this notation will be used for any equations in I.] First, the system was linearized about a constant solution U_0 to characterize the plane waves by means of the dispersion relation [Eq. (I.4)]:

$$\det W_{\pm 1} \equiv |\mp i\omega I \pm ikA_0 + \nabla B_0| = 0.$$

Then, for the matrices W_l ($l = 0, \pm 2, \pm 3, \cdots$), we assumed that [Eq. (I.4′)]:

$$\det W_l \neq 0, \quad |l| \neq 1.$$

However, as was noted at the end of I, in physical problems the condition is not, in general, valid for $l = 0$. Consequently, the method cannot be applied directly. In the present paper, we show by examples that, nevertheless, the method is still applicable to such systems.

The first example given in the next section is the electron plasma wave in a background of fixed ions, in which case the use of Poisson's equation as a subsidiary condition enables us to determine successively the components of $l = 0$, i.e., $U_0^{(\alpha)}$'s of Eq. (I.5). For other modes of $l \neq 0$, one can follow the process of I. We are thus able to exhibit a reduction of the original system of equations to a single tractable nonlinear equation. The equation is the nonlinear Schrödinger equation (I.16), which admits a further reduction to the Kortweg–de Vries equation. Special solutions in the lowest order of perturbation will be given in explicit forms.

In Sec. 3 a nonlinear Klein–Gordon equation is discussed as an example. Transformation of the second-order equation into the matrix form (I.1) is straightforward but shows that det W_0 also vanishes. The difficulty can be overcome in a way similar to that for the electron plasma wave, and, furthermore, the original equation is reduced to a nonlinear Schrödinger equation.

2. ELECTRON PLASMA WAVE

The equations governing the one-dimensional motion of an isothermal electron fluid without dissipation[2] are the continuity equation

$$\frac{\partial n}{\partial t} + u\frac{\partial n}{\partial x} + n\frac{\partial u}{\partial x} = 0 \tag{1a}$$

and the momentum balance-equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\kappa T_e}{mn}\frac{\partial n}{\partial x} + \frac{e}{m}E = 0. \tag{1b}$$

Here, n and u are the density and the flow velocity, respectively, m and e are the electron mass and charge, respectively, κ is the Boltzmann constant, T_e is the constant temperature, and E denotes the longitudinal component of the electric field. The hydrodynamic equations (1a), (1b) are coupled with Poisson's equation

$$\frac{\partial E}{\partial x} = 4\pi e(n_0 - n), \tag{1c}$$

[1] T. Taniuti and N. Yajima, J. Math. Phys. **10**, 1369 (1969). This paper will be referred to as I.

[2] L. Spitzer, *Physics of Fully Ionized Gases* (Interscience Publishers, Inc., New York, 1962), Sec. 3.

where n_0 is a constant density of fixed ions. Equations (1a)–(1c) constitute a complete set of equations which determine n, u, and E. However, in order to work with the system of equations written in the form of (I.1), we consider, instead of Eq. (1c), the longitudinal component of the first Maxwell equation,

$$\frac{\partial E}{\partial t} - 4\pi e n u = 0. \tag{1d}$$

Then Eqs. (1a) and (1d) imply that Poisson's equation (1c) perpetuates if it is valid initially. Consequently, it may be considered as a subsidiary condition. Thus, we have the system of equations [Eq. (I.1)]:

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} + B = 0,$$

with the representations

$$U = \begin{pmatrix} n \\ u \\ E \end{pmatrix}, \tag{2a}$$

$$A = \begin{pmatrix} u & n & 0 \\ \kappa T_e/mn & u & 0 \\ 0 & 0 & u \end{pmatrix}, \tag{2b}$$

$$B = \begin{pmatrix} 0 \\ (e/m)E \\ -4\pi e n_0 u \end{pmatrix}. \tag{2c}$$

A constant solution $U^{(0)}$, about which the expansion (5.1) is considered, may be taken as

$$U^{(0)} \equiv \begin{pmatrix} n_0 \\ 0 \\ 0 \end{pmatrix}. \tag{3}$$

Then the matrix W_l, i.e., $-i\omega lI + iklA_0 + \nabla B_0$, becomes

$$\begin{pmatrix} -il\omega & ilkn_0 & 0 \\ ilka^2/n_0 & -il\omega & e/m \\ 0 & -4\pi e n_0 & -il\omega \end{pmatrix} \tag{4a}$$

and $\det W_1 = 0$ yields the dispersion relation

$$\omega^2 = \omega_0^2 + a^2k^2, \tag{4b}$$

where ω_0 is the plasma frequency $(4\pi n_0 e^2/m)^{\frac{1}{2}}$ and a the thermal velocity $(\kappa T_e/m)^{\frac{1}{2}}$. Thus, the vectors R and L introduced through Eqs. (9) and (I.9′) are

$$R = \begin{pmatrix} 1 \\ \omega/kn_0 \\ i4\pi e/k \end{pmatrix}, \tag{5a}$$

$$L = (a^2/n_0,\ \omega/k,\ -ie/mk). \tag{5b}$$

Computations of the coefficients α and β given by Eq. (I.14) are straightforward, and we have

$$\alpha = 2\omega^2/n_0k^2, \tag{6a}$$

$$\beta = -i\omega_0^2a^2/\omega k^2 n_0. \tag{6b}$$

Since $\det W_l$ does not vanish for any l except 0 and ± 1, one can directly calculate $R_2^{(2)}$ of Eq. (I.15a):

$$R_2^{(2)} = \frac{1}{3n_0\omega_0^2}\begin{pmatrix} 4\omega^2 + 2\omega_0^2 \\ \omega(4\omega^2 - \omega_0^2)/n_0k \\ i4\pi e(2\omega^2 + \omega_0^2)/k \end{pmatrix}. \tag{7}$$

However, $\det W_0$ obviously vanishes; hence, $U_0^{(2)}$ and, consequently, $R_0^{(2)}$ can not be given by Eq. (I.15b), and the method of solution of I is not directly applicable. In the subsequent discussions, we show that by means of the subsidiary condition (1c), the difficulty is overcome and the $U_0^{(a)}$'s are determined successively. First, we note that for $l = 0$, the left-hand side of Eq. (1c) is

$$\epsilon^2(\partial E_0^{(1)}/\partial\xi) + \epsilon^3(\partial E_0^{(2)}/\partial\xi) + \cdots.$$

Hence, one has immediately that

$$n_0^{(1)} = 0,$$

whence Eq. (I.7a) for $l = 0$ (i.e., $W_0U_0^{(1)} = 0$) yields

$$u_0^{(1)} = E_0^{(1)} = 0,$$

and, consequently,

$$U_0^{(1)} = 0. \tag{8}$$

Substituting Eq. (8) in Eq. (1c), we find that

$$n_0^{(2)} = 0.$$

On the other hand, for $l = 0$, Eq. (I.7b) becomes

$$W_0U_0^{(2)} + ik[(\nabla A_0 \cdot R^*)R - (\nabla A_0 \cdot R)R^*]\,|\phi^{(1)}|^2 = 0.$$

It is easy to see that this equation determines all the components of $U_0^{(2)}$ except $n_0^{(2)}$. After elementary calculations we thus obtain

$$U_0^{(2)} = \frac{-2\omega}{n_0^2k}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}|\phi^{(1)}|^2 \tag{9a}$$

or

$$R_0^{(2)} = \frac{-2\omega}{n_0^2k}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{9b}$$

In this way, we can proceed to any desired order in ϵ. Substituting Eqs. (7) and (9b) in Eq. (I.14c),

$$\gamma = ia^2\omega(8\omega^2 + \omega_0^2)/3n_0^3\omega_0^2, \tag{10}$$

and we finally obtain Eq. (I.16):

$$i\frac{\partial\phi^{(1)}}{\partial\tau} + p\frac{\partial^2\phi^{(1)}}{\partial\xi^2} + q\,|\phi^{(1)}|^2\,\phi^{(1)} = 0. \tag{11a}$$

Here p and q are given by the equations

$$p = \beta/|\alpha| = \omega_0^2 a^2/2\omega^3, \tag{11b}$$

$$q = \gamma/|\alpha| = -k^2a^2(8\omega^2 + \omega_0^2)/6n_0^2\omega\omega_0^2. \tag{11c}$$

Special solutions can be found easily. Assume, for $\phi^{(1)}$, the form

$$\phi^{(1)} = f(\xi) \exp(-i\nu\tau),$$

where ν is a positive constant.

Let f be a real function of ξ. Then one has

$$\frac{d^2f}{d\xi^2} + \tilde{\nu}f + \tilde{\gamma}f^3 = 0, \tag{12}$$

in which $\tilde{\nu}$ and $\tilde{\gamma}$ are equal to ν/p and q/p, respectively, e.g.,

$$\tilde{\nu} = 2\omega^3\nu/(\omega_0 a)^2, \tag{13a}$$

$$\tilde{\gamma} = -\omega^2k^2(8\omega^2 + \omega_0^2)/3n_0^2\omega_0^4. \tag{13b}$$

Hence, in this case, the system is equivalent to the motion of a mass point under the potential

$$V(f) = \tfrac{1}{2}\tilde{\nu}f^2 + \tfrac{1}{4}\tilde{\gamma}f^4. \tag{14}$$

Figure 1 shows that the motion is oscillatory for amplitudes below the critical amplitude $(-\tilde{\nu}/\tilde{\gamma})^{\frac{1}{2}} = (-\nu/q)^{\frac{1}{2}}$, for which f becomes

$$f = (-\nu/q)^{\frac{1}{2}} \tanh[(\nu/2p)^{\frac{1}{2}}\xi]. \tag{15}$$

If f is complex valued, we have the plane wave (I.18), i.e., Eq. (11a) is satisfied by

$$f = \phi_0 \exp(is\xi), \tag{16a}$$

provided s is given by

$$\nu = ps^2 - q\phi_0^2. \tag{16b}$$

Since $pq < 0$, as was shown in Eq. (I.20) below, the plane waves are stable for modulation, and Eq. (11a) can be reduced, in an asymptotic sense, to the Kortweg–de Vries equation. That is, consider the solution of the form (I.19)

$$\phi^{(1)} = \rho^{\frac{1}{2}} \exp\left[i\int \frac{\sigma\, d\xi}{2p}\right] \tag{17}$$

to obtain the system of equations [(I.20a) and (I.20b)] for ρ and σ:

$$\frac{\partial\rho}{\partial\tau} + \sigma\frac{\partial\rho}{\partial\xi} + \rho\frac{\partial\sigma}{\partial\xi} = 0,$$

$$\frac{\partial\sigma}{\partial\tau} + \sigma\frac{\partial\sigma}{\partial\xi} - 2pq\frac{\partial\rho}{\partial\xi} - p^2\frac{\partial}{\partial\xi}\left\{\left(\frac{1}{\rho}\right)^{\frac{1}{2}}\frac{\partial}{\partial\xi}\left[\left(\frac{1}{\rho}\right)^{\frac{1}{2}}\frac{\partial\rho}{\partial\xi}\right]\right\} = 0.$$

Here, we note that Eqs. (I.20) admit the solitary wave; namely, under the boundary condition for $\xi \to \infty$,

$$\rho = \rho_0,$$
$$\sigma = a^* - \sigma_\infty, \quad \sigma_\infty > 0,$$

one has

$$\sigma = a^* - \sigma_\infty\{1 - (1 - M^2) \times \operatorname{sech}^2[(a^*/2p)(1 - M^2)^{\frac{1}{2}}(\xi - a^*\tau)]\}^{-1}, \tag{18a}$$

$$\rho = \rho_0 - \rho_0(1 - M^2) \times \operatorname{sech}^2[(a^*/2p)(1 - M^2)^{\frac{1}{2}}(\xi - a^*\tau)], \tag{18b}$$

where a^* is the effective sound velocity equal to $(-2pq\rho_0)^{\frac{1}{2}}$, whence M is the Mach number σ_∞/a^*, which is less than unity in order that the solitary wave exist. Then an expansion in terms of a small parameter μ (> 0),

$$\rho = \rho_0 + \mu\rho_1 + \mu^2\rho_2 + \cdots,$$
$$\sigma = \mu\sigma_1 + \mu^2\sigma_2 + \cdots,$$

yields the Kortweg–de Vries equation for ρ_1 and σ_1:

$$\frac{\partial\sigma_1}{\partial\tau'} + \frac{3}{2}\sigma_1\frac{\partial\sigma_1}{\partial\xi'} - \frac{\sigma_{1\infty}}{2}\left(\frac{\partial\sigma_1}{\partial\xi'}\right) - \frac{p^2}{2a^*}\frac{\partial^3\sigma_1}{\partial\xi'^3} = 0, \tag{19a}$$

$$\rho_1 = (\rho_0/a^*)(\sigma_1 - \sigma_{1\infty}), \tag{19b}$$

in which ξ' and τ' are stretched coordinates defined as

$$\xi' = \mu^{\frac{1}{2}}(\xi - a^*\tau),$$
$$\tau' = \mu^{\frac{3}{2}}\tau.$$

We have assumed $\sigma_1 \to \sigma_{1\infty} > 0$, $\rho_1 \to 0$ for $\xi \to \infty$. From Eq. (19a), we get the solitary wave

$$\sigma_1 = \sigma_{1\infty} - 2\sigma_{1\infty} \cdot \operatorname{sech}^2[\tfrac{1}{2}(2\sigma_{1\infty}a^*/p^2)^{\frac{1}{2}}\xi'] \tag{20}$$

which, of course, also results from Eq. (18a) if we put $M^2 = 1 - 2\mu$ and $\sigma_\infty = \sigma_{1\infty}$. Substituting Eq. (20) in Eq. (17) gives

$$\phi^{(1)} \simeq \rho_0^{\frac{1}{2}}(1 + \sigma/2a^*)[\exp(-i\nu\tau)] \times \exp\{i\sigma_\infty[\xi - (2/\delta)\tanh\delta(\xi - a^*\tau)]\}. \tag{21}$$

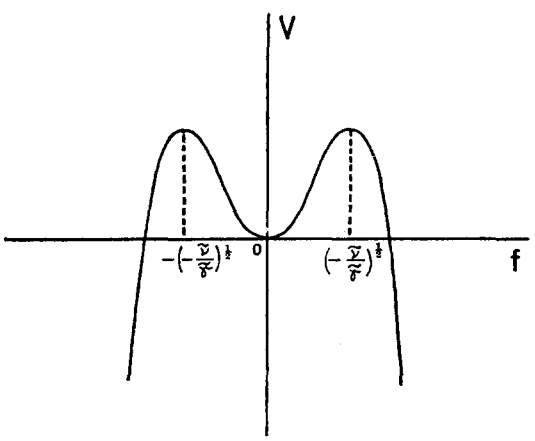

FIG. 1. Potential curve.

Here, σ and σ_∞ stand for $\mu(\sigma_1 - \sigma_{1\infty})$ and $\mu\sigma_{1\infty}$, respectively, and δ designates $\frac{1}{2}(2\sigma_\infty a^*\mu/p^2)^{\frac{1}{2}}$, where ν is $-q\rho_0$.

3. NONLINEAR KLEIN-GORDON EQUATION

Consider the nonlinear Klein-Gordon equation[3] for a real scalar function ψ:

$$\frac{\partial^2\psi}{\partial t^2} - \frac{\partial^2\psi}{\partial x^2} + m^2\psi + \kappa\psi^3 = 0, \tag{22}$$

where m and κ are real constants. Introducing χ and ϕ by the equations

$$\chi - \frac{\partial\psi}{\partial x} = 0, \tag{23a}$$

$$\phi - \frac{\partial\psi}{\partial t} = 0, \tag{23b}$$

and differentiating Eq. (23a) with respect to t, we can bring Eq. (22) into the matrix form (I.1):

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} + B = 0. \tag{24}$$

Here, U, A, and B take the forms

$$U = \begin{pmatrix} \phi \\ \chi \\ \psi \end{pmatrix}, \tag{25a}$$

$$A = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{25b}$$

$$B = \begin{pmatrix} m^2\psi + \kappa\psi^3 \\ 0 \\ -\phi \end{pmatrix}. \tag{25c}$$

In what follows, Eq. (23a) will be regarded as a subsidiary condition which perpetuates if it is valid initially. We now assume the expansion (I.5) about the constant solution

$$U_0 = 0.$$

Then the matrices W_l are expressed by

$$W_l = \begin{pmatrix} -il\omega & -ilk & m^2 \\ -ilk & -il\omega & 0 \\ -1 & 0 & -il\omega \end{pmatrix}. \tag{26}$$

The dispersion relation $\det W_{\pm 1} = 0$ becomes

$$\omega^2 - k^2 - m^2 = 0, \tag{27}$$

which implies

$$\det W_l \neq 0,$$

unless $|l|$ is zero or unity. Hence, the $U_l^{(1)}$, except $U_0^{(1)}$, are given by Eqs. (I.8), i.e.,

$$U_l^{(1)} = 0, \quad \text{for} \quad |l| \geq 2, \tag{28a}$$

$$U_1^{(1)} = \phi^{(1)}R, \tag{28b}$$

where R is the column vector introduced in Eq. (I.9) and takes the form

$$R = \begin{pmatrix} \omega \\ -k \\ i \end{pmatrix}. \tag{29a}$$

The corresponding row vector L may be given by

$$L = (\omega, -k, -im^2). \tag{29b}$$

Since $\det W_0$ vanishes also in this example, a different method is required to account for $U_0^{(\alpha)}$. Consider the component $l = 0$ of Eq. (24). For the first order in ϵ, it yields

$$\phi_0^{(1)} = \psi_0^{(1)} = 0;$$

hence, the subsidiary condition (23a) yields

$$\chi_0^{(1)} = 0$$

and, consequently, we have

$$U_0^{(1)} = 0. \tag{28c}$$

Similarly, by substituting Eqs. (28a) and (28c) in Eqs. (24) and (23a), for the second order in ϵ we have

$$U_0^{(2)} = 0. \tag{30a}$$

Computations of the $U_l^{(2)}$ ($l \neq 0$) are straightforward: Since

$$\nabla A_0 = \nabla\nabla A_0 = \nabla\nabla B_0 = 0,$$

we find from Eqs. (12c), (12d), and (15a) that

$$U_2^{(2)} = U_{-2}^{(2)} = R_2^{(2)} = 0, \tag{30b}$$

and solving Eq. (I.10) for $U_1^{(2)}$ yields

$$U_1^{(2)} = \phi^{(2)}R + i\begin{pmatrix} -k/\omega \\ 1 \\ 0 \end{pmatrix}\frac{\partial\phi^{(1)}}{\partial\xi}. \tag{30c}$$

Introducing Eqs. (29) and (30) into Eqs. (I.14) yields

$$\alpha = 2\omega^2, \tag{31a}$$

$$\beta = -im^2/\omega, \tag{31b}$$

$$\gamma = 3i\omega\kappa. \tag{31c}$$

[3] L. I. Schiff, Phys. Rev. **84**, 1 (1951); W. Thirring, Z. Naturforsch. **7a**, 63 (1952).

Hence, we have Eq. (I.16):

$$i\frac{\partial\phi^{(1)}}{\partial\tau} + p\frac{\partial^2\phi^{(1)}}{\partial\xi^2} + q\,|\phi^{(1)}|^2\,\phi^{(1)} = 0,$$

with

$$p = m^2/2\omega^3,$$

$$q = -3\kappa/2\omega.$$

For a negative κ, pq is positive so that this equation admits the solitary-wave solution, such as that given by Eq. (I.17), while the plane waves are modulationally unstable as was stated before in Eq. (I.18). On the other hand, if κ is positive, we have the solutions such as those given in Sec. 2 for the plasma wave, and the equation can be reduced to the Kortweg–de Vries equation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

On Canonical $SO(4, 1)$ Transformations of the Dirac Equation

A. J. BRACKEN AND H. A. COHEN
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia

(Received 20 August 1968)

Certain matrix transformations of the free-particle Dirac equation are described as momentum-dependent $SO(4, 1)$ transformations. Such of these belonging to any one of five subgroups $G^{(\alpha)}$ ($\alpha = 0, 1, 2, 3, 4$) are canonical, preserving the Lorentz-invariant Dirac scalar product in a corresponding one of five modes of expression. The Dirac equation itself is linear in all five components p_α [p_μ ($\mu = 0, 1, 2, 3$) is the four-momentum operator, and $p_4 = m$] of the "five-vector" $\tilde{p}$, and a transformation in $G^{(\beta)}$ has the additional property that the component p_β appears linearly also in the transformed equation. The Mendlowitz and the Foldy–Wouthuysen–Tani transformation accordingly are in $G^{(0)}$, the $SO(4)$ subgroup; and that proposed by Chakrabarti is in $G^{(4)}$, the $SO(3, 1)$ subgroup associated with homogeneous Lorentz transformations. For any $\tilde{p}'$, obtained from $\tilde{p}$ by a momentum-dependent $SO(4, 1)$ transformation, there is a corresponding transform of the Dirac equation. Where p_α appears in the Dirac equation, p'_α appears in the transformed equation. The ambiguities which arise in the specification of the transformation leading to a given such equation are associated with the existence of a "little group" for any such $\tilde{p}'$.

1. INTRODUCTION

The Dirac equation for the four-component wave-function $\psi^{(D)}(x)$ is

$$(\gamma_\mu p^\mu - m)\psi^{(D)} = 0, \tag{1.1}$$

where

$$p_\mu = \frac{i\partial}{\partial x^\mu}, \quad \mu = 0, 1, 2, 3, \tag{1.2}$$

and the matrices γ_μ form an irreducible representation of the Dirac–Clifford algebra, with

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}. \tag{1.3}$$

[We choose the diagonal metric with $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$; and, with no significant loss of generality, we take γ_0, $i\gamma_1$, $i\gamma_2$, $i\gamma_3$, and $i\gamma_5$ ($= i\gamma_0\gamma_1\gamma_2\gamma_3$) to be Hermitian.]

The transformation properties of the bispinor function $\psi^{(D)}(x)$ with respect to the restricted homogeneous Lorentz group $SO(3, 1)$ are well known. However, it has also been long known that larger groups, in fact certain groups of rotations in five- and six-dimensional spaces, are pertinent to discussions of the Dirac equation, the Lie algebra of the Dirac matrices γ_μ, etc.[1–7] In this paper, the connection between Eq. (1.1) and the group $SO(4, 1)$ in particular is exploited in the development of a unifying group-theoretical description of certain canonical momentum-dependent transformations of the equation.

In a previous publication,[8] henceforth referred to as BC, two features of Eq. (1.1) assume significance: namely, the linearity in all five of the quantities p_μ, m, and the existence of five different ways of expressing one and the same Lorentz-invariant scalar product

[1] A. S. Eddington, Proc. Roy. Soc. (London) **121A**, 524 (1928).
[2] P. A. M. Dirac, Ann. Math. **36**, 657 (1935).
[3] Harish-Chandra, Proc. Indian Acad. Sci. **22A**, 30 (1945).
[4] A. O. Barut, Phys. Rev. **135**, B839 (1964). See also A. J. MacFarlane, Commun. Math. Phys. **2**, 133 (1966); A. ten Kate, J. Math. Phys. **9**, 181 (1968).
[5] J. K. Lubański, Physica **9**, 310 (1942). See also H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945); J. A. de Vos and J. Hilgevoord, Nucl. Phys. **B1**, 494 (1967); M. M. Bakri, J. Math. Phys. **10**, 298 (1969).
[6] See, for example, C. Fronsdal, Proc. Roy. Soc. (London) **288A**, 113 (1965), and references given therein.
[7] A. O. Barut, Phys. Rev. Letters **20**, 893 (1968).
[8] A. J. Bracken and H. A. Cohen, Progr. Theoret. Phys. **41**, 816 (1969).

Hence, we have Eq. (I.16):

$$i\frac{\partial\phi^{(1)}}{\partial\tau} + p\frac{\partial^2\phi^{(1)}}{\partial\xi^2} + q\,|\phi^{(1)}|^2\,\phi^{(1)} = 0,$$

with

$$p = m^2/2\omega^3,$$

$$q = -3\kappa/2\omega.$$

For a negative κ, pq is positive so that this equation admits the solitary-wave solution, such as that given by Eq. (I.17), while the plane waves are modulationally unstable as was stated before in Eq. (I.18). On the other hand, if κ is positive, we have the solutions such as those given in Sec. 2 for the plasma wave, and the equation can be reduced to the Kortweg–de Vries equation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

On Canonical $SO(4, 1)$ Transformations of the Dirac Equation

A. J. BRACKEN AND H. A. COHEN
Department of Mathematical Physics, University of Adelaide, Adelaide, South Australia

(Received 20 August 1968)

Certain matrix transformations of the free-particle Dirac equation are described as momentum-dependent $SO(4, 1)$ transformations. Such of these belonging to any one of five subgroups $G^{(\alpha)}$ ($\alpha = 0, 1, 2, 3, 4$) are canonical, preserving the Lorentz-invariant Dirac scalar product in a corresponding one of five modes of expression. The Dirac equation itself is linear in all five components p_α [p_μ ($\mu = 0, 1, 2, 3$) is the four-momentum operator, and $p_4 = m$] of the "five-vector" $\tilde{p}$, and a transformation in $G^{(\beta)}$ has the additional property that the component p_β appears linearly also in the transformed equation. The Mendlowitz and the Foldy–Wouthuysen–Tani transformation accordingly are in $G^{(0)}$, the $SO(4)$ subgroup; and that proposed by Chakrabarti is in $G^{(4)}$, the $SO(3, 1)$ subgroup associated with homogeneous Lorentz transformations. For any $\tilde{p}'$, obtained from $\tilde{p}$ by a momentum-dependent $SO(4, 1)$ transformation, there is a corresponding transform of the Dirac equation. Where p_α appears in the Dirac equation, p'_α appears in the transformed equation. The ambiguities which arise in the specification of the transformation leading to a given such equation are associated with the existence of a "little group" for any such $\tilde{p}'$.

1. INTRODUCTION

The Dirac equation for the four-component wave-function $\psi^{(D)}(x)$ is

$$(\gamma_\mu p^\mu - m)\psi^{(D)} = 0, \tag{1.1}$$

where

$$p_\mu = \frac{i\partial}{\partial x^\mu}, \quad \mu = 0, 1, 2, 3, \tag{1.2}$$

and the matrices γ_μ form an irreducible representation of the Dirac–Clifford algebra, with

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}. \tag{1.3}$$

[We choose the diagonal metric with $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$; and, with no significant loss of generality, we take γ_0, $i\gamma_1$, $i\gamma_2$, $i\gamma_3$, and $i\gamma_5$ ($= i\gamma_0\gamma_1\gamma_2\gamma_3$) to be Hermitian.]

The transformation properties of the bispinor function $\psi^{(D)}(x)$ with respect to the restricted homogeneous Lorentz group $SO(3, 1)$ are well known. However, it has also been long known that larger groups, in fact certain groups of rotations in five- and six-dimensional spaces, are pertinent to discussions of the Dirac equation, the Lie algebra of the Dirac matrices γ_μ, etc.[1–7] In this paper, the connection between Eq. (1.1) and the group $SO(4, 1)$ in particular is exploited in the development of a unifying group-theoretical description of certain canonical momentum-dependent transformations of the equation.

In a previous publication,[8] henceforth referred to as BC, two features of Eq. (1.1) assume significance: namely, the linearity in all five of the quantities p_μ, m, and the existence of five different ways of expressing one and the same Lorentz-invariant scalar product

[1] A. S. Eddington, Proc. Roy. Soc. (London) **121A**, 524 (1928).
[2] P. A. M. Dirac, Ann. Math. **36**, 657 (1935).
[3] Harish-Chandra, Proc. Indian Acad. Sci. **22A**, 30 (1945).
[4] A. O. Barut, Phys. Rev. **135**, B839 (1964). See also A. J. MacFarlane, Commun. Math. Phys. **2**, 133 (1966); A. ten Kate, J. Math. Phys. **9**, 181 (1968).
[5] J. K. Lubański, Physica **9**, 310 (1942). See also H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945); J. A. de Vos and J. Hilgevoord, Nucl. Phys. **B1**, 494 (1967); M. M. Bakri, J. Math. Phys. **10**, 298 (1969).
[6] See, for example, C. Fronsdal, Proc. Roy. Soc. (London) **288A**, 113 (1965), and references given therein.
[7] A. O. Barut, Phys. Rev. Letters **20**, 893 (1968).
[8] A. J. Bracken and H. A. Cohen, Progr. Theoret. Phys. **41**, 816 (1969).

$(\psi_1^{(D)}, \psi_2^{(D)})$ of any two solutions $\psi_{1,2}^{(D)}(x)$. The five expressions are all simple in the momentum representation, introduced by defining

$$\psi^{(D)}(x_0, \mathbf{x}) = (2\pi)^{-\frac{3}{2}} \int \frac{d^3k}{\omega(\mathbf{k}, m)} e^{i\mathbf{k}\cdot\mathbf{x}} \{e^{-i\omega(\mathbf{k},m)x_0} \chi^{(D)+}(\mathbf{k}) + e^{+i\omega(\mathbf{k},m)x_0} \chi^{(D)-}(\mathbf{k})\}, \quad (1.4)$$

where

$$\omega(\mathbf{k}, m) = (\mathbf{k}^2 + m^2)^{\frac{1}{2}}, \quad (1.5)$$

and with, conversely,

$$\chi^{(D)\pm}(\mathbf{k}) = e^{\pm i\omega(\mathbf{k},m)x_0} \tfrac{1}{2}[\omega(\mathbf{k}, m) \pm p_0](2\pi)^{-\frac{3}{2}} \times \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \psi^{(D)}(x_0, \mathbf{x}). \quad (1.6)$$

Then the well-known coordinate representation form

$$(\psi_1^{(D)}, \psi_2^{(D)}) = \int_{x_0 \text{ const}} d^3x \psi_1^{(D)\dagger}(x_0, \mathbf{x}) \psi_2^{(D)}(x_0, \mathbf{x}), \quad (1.7)$$

where $\psi^{(D)\dagger}$ is the Hermitian conjugate of $\psi^{(D)}$, yields the five expressions

$$(\psi_1^{(D)}, \psi_2^{(D)}) = \int \frac{d^3k}{\omega^2(\mathbf{k}, m)} \{\chi_1^{(D)+\dagger}(\mathbf{k}) \chi_2^{(D)+}(\mathbf{k}) + \chi_1^{(D)-\dagger}(\mathbf{k}) \chi_2^{(D)-}(\mathbf{k})\} \quad (1.8a)$$

$$= \int \frac{d^3k}{m\omega(\mathbf{k}, m)} \{\bar{\chi}_1^{(D)+}(\mathbf{k}) \chi_2^{(D)+}(\mathbf{k}) - \bar{\chi}_1^{(D)-}(\mathbf{k}) \chi_2^{(D)-}(\mathbf{k})\} \quad (1.8b)$$

and

$$= \int \frac{d^3k}{\omega(\mathbf{k}, m) k_i} \{\bar{\chi}_1^{(D)+}(\mathbf{k}) \gamma_i \chi_2^{(D)+}(\mathbf{k}) - \bar{\chi}_1^{(D)-}(\mathbf{k}) \gamma_i \chi_2^{(D)-}(\mathbf{k})\} \quad (i = 1, 2, \text{or } 3; \text{ no summation}); \quad (1.8c)$$

where

$$\bar{\chi}^{(D)\pm} = \chi^{(D)\pm\dagger} \gamma_0. \quad (1.9)$$

The equivalence of these five expressions is established using the identity

$$k_\mu \bar{\chi}_1^{(D)\pm}(\mathbf{k}) \chi_2^{(D)\pm}(\mathbf{k}) = m \bar{\chi}_1^{(D)\pm}(\mathbf{k}) \gamma_\mu \chi_2^{(D)\pm}(\mathbf{k}), \quad (1.10)$$

where we define

$$k_0 \chi^{(D)\pm}(\mathbf{k}) = \pm\omega(\mathbf{k}, m) \chi^{(D)\pm}(\mathbf{k}). \quad (1.11)$$

Equation (1.10) in turn follows[9] from the fact that one has, from (1.1) and (1.6),

$$(\gamma_\mu k^\mu - m) \chi_{1,2}^{(D)\pm} = 0. \quad (1.12)$$

In BC we considered momentum-dependent matrix transformations of the form

$$\psi^{(D)}(x) \to \psi'(x) = V(p, m) \psi^{(D)}(x), \quad (1.13)$$

leading to the equation

$$V(p, m)[\gamma_\mu p^\mu - m] V^{-1}(p, m) \psi' = 0. \quad (1.14)$$

Five special classes of such transformations were presented, with every transformation in a given class having two properties characteristic of that class.

The first of these properties is that the linearity in a corresponding one of the five quantities p_μ, m is maintained in the transformed equation. In this way "p_0-", "p_1-", "p_2-", "p_3-", and "m-linear" equations are obtained.

Amongst the "p_0-linear" forms, one finds the Foldy–Wouthuysen–Tani[10] equation

$$p_0 \psi^{(F)} = \gamma_0 \omega(\mathbf{p}, m) \psi^{(F)}, \quad (1.15)$$

with, in this case,

$$\psi^{(F)}(x) = F(\mathbf{p}, m) \psi^{(D)}(x), \quad (1.16)$$

where

$$F(\mathbf{p}, m) = \exp\left[\frac{\boldsymbol{\gamma}\cdot\mathbf{p}}{2|\mathbf{p}|} \arctan\left(\frac{|\mathbf{p}|}{m}\right)\right]. \quad (1.17)$$

Also of the "p_0-linear" type is the equation proposed by Mendlowitz[11]:

$$p_0 \psi^{(M)} = \omega(\mathbf{p}, m) \gamma_0 \frac{\boldsymbol{\gamma}\cdot\mathbf{p}}{|\mathbf{p}|} \psi^{(M)}, \quad (1.18)$$

where

$$\psi^{(M)}(x) = M(\mathbf{p}, m) \psi^{(D)}(x), \quad (1.19)$$

with

$$M(\mathbf{p}, m) = \exp\left[-\frac{\boldsymbol{\gamma}\cdot\mathbf{p}}{2|\mathbf{p}|} \arctan\left(\frac{m}{|\mathbf{p}|}\right)\right]. \quad (1.20)$$

Amongst the "m-linear" equations is that proposed by Chakrabarti[12]:

$$\epsilon(p_0)(p_\mu p^\mu)^{\frac{1}{2}} \gamma_0 \psi^{(C)} = m \psi^{(C)}, \quad (1.21)$$

where

$$\epsilon(p_0) = p_0 |p_0|^{-1} \quad (1.22)$$

and

$$\psi^{(C)}(x) = C(p) \psi^{(D)}(x), \quad (1.23)$$

[9] The proof is a simple extension of that for the case $\chi_1^{(D)\pm} = \chi_2^{(D)\pm}$ as given, for example, in S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row Peterson & Company, New York, 1961), Chap. 4, preceding equation (129).

[10] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950); S. Tani, Progr. Theoret. Phys. (Kyoto) **6**, 267 (1951). The transformation was in fact earlier proposed by M. H. L. Pryce, Proc. Roy. Soc. (London) **195A**, 62 (1948).

[11] H. Mendlowitz, Phys. Rev. **102**, 527 (1956). The transformation was rediscovered by M. Cini and B. Touschek, Nuovo Cimento **7**, 422 (1958); and independently by S. K. Bose, A. Gamba, and E. C. G. Sudarshan, Phys. Rev. **113**, 1661 (1959).

[12] A. Chakrabarti, J. Math. Phys. **4**, 1215, 1223 (1963).

with

$$C(p) = \exp\left\{-\frac{\epsilon(p_0)\gamma_0\boldsymbol{\gamma}\cdot\mathbf{p}}{2|\mathbf{p}|}\operatorname{arc\,tanh}\left(\frac{|\mathbf{p}|}{|p_0|}\right)\right\}. \quad (1.24)$$

The simplest "p_3-linear" form obtained is

$$p_3\psi' = \epsilon(p_0)\lambda(p_0, p_1, p_2, m)\gamma_0\gamma_3\psi', \quad (1.25)$$

where

$$\lambda(p_0, p_1, p_2, m) = [(p_0)^2 - (p_1)^2 - (p_2)^2 - m^2]^{\frac{1}{2}} \quad (1.26)$$

and

$$\psi'(x) = W(p_0, p_1, p_2, m)\psi^{(D)}(x), \quad (1.27)$$

with

$$W(p_0, p_1, p_2, m) = \exp\left\{-\frac{\epsilon(p_0)\gamma_0(\gamma_1 p_1 + \gamma_2 p_2 + m)}{2[(p_1)^2 + (p_2)^2 + m^2]^{\frac{1}{2}}} \times \operatorname{arc\,tanh}\left(\frac{[(p_1)^2 + (p_1)^2 + m^2]^{\frac{1}{2}}}{|p_0|}\right)\right\}. \quad (1.28)$$

The second property characterizing a given one of the five classes is that every transformation within that class preserves the Dirac scalar product in a corresponding one of the five modes (1.8a)–(1.8c), so that each class consists of canonical transformations. In fact, the transformations leading to "p_0-, "m-, and "p_i-linear" forms preserve the modes (1.8a)–(1.8c), respectively.

In describing a subset of transformations (1.13) as momentum-dependent $SO(4, 1)$ transformations, making use of the connection between Eq. (1.1) and this group, we aim here in particular to interpret the above results of BC in group-theoretical terms. To this end, in Sec. 2, we make explicit this connection to the extent required in what follows.

In Sec. 3, the significance of such a connection in regard to transformations of the form (1.13) is established. We stress in particular the existence of five subgroups of $SO(4, 1)$, labeled by us $G^{(\alpha)}$ ($\alpha = 0, 1, 2, 3, 4$), which have the special property that any momentum-dependent transformation (1.13) within a given $G^{(\alpha)}$ leaves the equation linear in the corresponding p_α (where we write $p_4 = m$).

As might be expected, the five classes of canonical transformations presented in BC fall into these five subgroups and, in fact, every transformation in a given $G^{(\alpha)}$ also preserves the corresponding mode of the scalar product. This we show in Secs. 4, 5, and 6, where the subgroups $G^{(0)}$, $G^{(4)}$, and $G^{(3)}$ (as typical of $G^{(i)}$, $i = 1, 2, 3$) and, correspondingly, "p_0-, "m-, and "p_3-linear" forms of the equation, are discussed in more detail. [It is not shown that an arbitrary momentum-dependent $SO(4, 1)$ transformation is canonical—in fact it is not possible to write the scalar product (1.8a)–(1.8c) in $SO(4, 1)$-invariant form.]

We find that $G^{(0)}$ is the maximal compact subgroup $SO(4)$, and $G^{(4)}$ the $SO(3, 1)$ group relating to the Lorentz transformation properties of the equation. The $G^{(i)}$ are also $SO(3, 1)$ subgroups, distinct from $G^{(4)}$ and from one another.

In Sec. 7, a discussion is given of the "little group" of $SO(4, 1)$ transformations which leave a particular transform of the equation invariant, and the nature of ambiguities which arise when one wishes to transform one equation into another are made explicit.

2. DIRAC EQUATION AND $SO(4, 1)$

The sixteen elements

$$I, \gamma_\mu, \gamma_5, \gamma_5\gamma_\mu, \quad \text{and} \quad [\gamma_\mu, \gamma_\nu] \quad (2.1)$$

of the Dirac–Clifford algebra form a complete set of 4×4 matrices, in terms of which the infinitesimal generators of a four-dimensional representation of any Lie group can be expressed as linear combinations with complex coefficients.

In this connection, one is familiar with the case of $SO(3, 1)$,[13] where the generators are defined as

$$S_{\mu\nu} = (i/4)[\gamma_\mu, \gamma_\nu] \quad (2.2)$$

and satisfy the characteristic Lorentz-group commutation rules

$$[S_{\mu\nu}, S_{\rho\sigma}] = -i(g_{\mu\rho}S_{\nu\sigma} + g_{\nu\sigma}S_{\mu\rho} - g_{\mu\sigma}S_{\nu\rho} - g_{\nu\rho}S_{\mu\sigma}). \quad (2.3)$$

The significance of these operators in regard to the Dirac equation is well known. In essence, the invariance of the Dirac description of free spin-$\frac{1}{2}$ particles under restricted homogeneous Lorentz transformations is expressed in the fact that

$$[\gamma_\mu p^\mu, J_{\rho\sigma}] = 0, \quad (2.4)$$

where

$$J_{\mu\nu} = L_{\mu\nu} + S_{\mu\nu} \quad (2.5)$$

and the $L_{\mu\nu}$, satisfying commutation rules analogous to (2.3), are defined by

$$L_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu. \quad (2.6)$$

Although not always referred to as such explicitly, representations of the Lie algebras of larger groups, such as $SO(4, 1)$ and $SO(4, 2)$, have been given in terms of (2.1) by, for example, Eddington,[1] Dirac,[2]

[13] We have taken some license with notation in referring to the groups $SO(3, 1)$, $SO(4)$, and $SO(4, 1)$, when in fact the covering groups $SL(2, C)$, $SU(2) \otimes SU(2)$, etc., are meant.

Harish-Chandra,[3] and Barut.[4] The connection between the orthogonal groups in five dimensions [such as $SO(4,1)$] and a class of relativistic wave equations, of which the Dirac equation is the simplest, was first discussed in detail by Lubański.[5] More recently $S\tilde{U}(4)$ [of which $SO(4,2)$ may be regarded as a type] has received attention from several authors[6]; and Barut,[7] in particular, has exploited the connection between Eq. (1.1) and $SO(4,2)$ in "reformulating the Dirac theory of the electron."

For our purposes here it will be sufficient to indicate and use the relationship of the group $SO(4,1)$ to the Dirac equation. Multiplying (1.1) by γ_5, one obtains

$$(\gamma_5\gamma_\mu p^\mu - \gamma_5 m)\psi^{(D)} = 0. \tag{2.7}$$

Defining

$$\Gamma_\mu = \gamma_5\gamma_\mu, \quad \Gamma_4 = \gamma_5, \tag{2.8}$$

and

$$p^4 = -m, \tag{2.9}$$

one can write (2.7) as

$$\Gamma_\alpha p^\alpha \psi^{(D)} = 0, \tag{2.10}$$

where the summation is now over $\alpha = 0, 1, 2, 3$, and 4. In the following, note that indices take these values:

$$\begin{gathered}\alpha, \beta, \gamma, \delta, \epsilon: 0, 1, 2, 3, 4,\\ \mu, \nu, \rho, \sigma: 0, 1, 2, 3,\\ \tau, \zeta, \eta: 0, 1, 2, 4,\\ \mathrm{a}, \mathrm{b}, \mathrm{c}: 1, 2, 3, 4,\\ i, j, k: 1, 2, 3.\end{gathered}$$

Introducing $g^{\alpha\beta}$ $(= g_{\alpha\beta})$, with

$$\begin{gathered}g_{00} = -g_{11} = -g_{22} = -g_{33} = -g_{44} = 1,\\ g_{\alpha\beta} = 0, \quad \alpha \neq \beta,\end{gathered} \tag{2.11}$$

we define $\Gamma^\alpha = g^{\alpha\beta}\Gamma_\beta$, etc. Now defining also

$$T_{\alpha\beta} = (i/4)[\Gamma_\alpha, \Gamma_\beta] \tag{2.12}$$

and noting

$$\{\Gamma_\alpha, \Gamma_\beta\} = 2g_{\alpha\beta}, \tag{2.13}$$

we find

$$\begin{aligned}[T_{\alpha\beta}, T_{\gamma\delta}] = -i(&g_{\alpha\gamma}T_{\beta\delta} + g_{\beta\delta}T_{\alpha\gamma}\\ &- g_{\alpha\delta}T_{\beta\gamma} - g_{\beta\gamma}T_{\alpha\delta}),\end{aligned} \tag{2.14}$$

which are the characteristic commutation rules for the Lie algebra of $SO(4,1)$. Since this group is non-compact, the $T_{\alpha\beta}$ are not all Hermitian, but

$$\Gamma_0 T_{\alpha\beta}^\dagger = T_{\alpha\beta}\Gamma_0. \tag{2.15}$$

One also has

$$[\Gamma_\alpha, T_{\beta\gamma}] = i(g_{\alpha\beta}\Gamma_\gamma - g_{\alpha\gamma}\Gamma_\beta). \tag{2.16}$$

Note that the ten different $T_{\alpha\beta}$ consist of the six different $T_{\mu\nu}$ $[\equiv S_{\mu\nu}$ of (2.2)] and four $T_{4\mu}$ $[= -(i/2)\gamma_\mu]$, so that this Lie algebra is as small as any containing scalar multiples of all four Dirac matrices γ_μ.[4] The representation of $SO(4,1)$ generated by these ten operators is irreducible. The Casimir operators[14] $-\frac{1}{2}T_{\alpha\beta}T^{\alpha\beta}$ and $-\omega_\alpha\omega^\alpha$, where

$$\omega_\alpha = \tfrac{1}{8}\epsilon_{\alpha\beta\gamma\delta\epsilon}T^{\beta\gamma}T^{\delta\epsilon} \tag{2.17}$$

$(= \frac{3}{4}\Gamma_\alpha$ in this case), are multiples of the unit matrix by $-\frac{5}{2}$ and $-\frac{45}{16}$, respectively.

It is worth mentioning that there are two inequivalent irreducible representations of the Clifford algebra defined by (2.13), both of four dimensions. By making the choice (2.8) for Γ_α, we fix on one of these. The other representation is obtained if one chooses instead

$$\Gamma_\mu = \gamma_5\gamma_\mu, \quad \Gamma_4 = -\gamma_5 \tag{2.18}$$

(and so necessarily $p^4 = m$). The set of $T_{\alpha\beta}$ one obtains in this case is then also different, again with

$$T_{\mu\nu} = (i/4)[\gamma_\mu, \gamma_\nu], \tag{2.19}$$

but now

$$T_{4\mu} = +(i/2)\gamma_\mu. \tag{2.20}$$

However, these $T_{\alpha\beta}$ generate an *equivalent* representation of $SO(4,1)$. (The invariants take the same values.) This is clear from the fact that this second set of $T_{\alpha\beta}$ is obtained from the first set via the substitution $\gamma_\mu \to -\gamma_\mu$. However, this can be achieved by a similarity transformation, because, under this substitution, a different set of matrices satisfying (1.3) is obtained and, as is well known, all irreducible representations of the Dirac–Clifford algebra are equivalent.

3. $SO(4,1)$ TRANSFORMATIONS OF THE EQUATION

When written in the form (2.10), the Dirac equation has an $SO(4,1)$-invariant appearance. One might hope to find operators $M_{\beta\gamma}$ satisfying commutation rules analogous to (2.14), and such that $\tilde{p}$ transforms[15] as a five-vector operator with respect to transformations generated by them, that is, such that [cf. (2.16)]

$$[p_\alpha, M_{\beta\gamma}] = i(g_{\alpha\beta}p_\gamma - g_{\alpha\gamma}p_\beta). \tag{3.1}$$

One would then have, ensuring $SO(4,1)$ invariance,

$$[\Gamma_\alpha p^\alpha, K_{\beta\gamma}] = 0, \tag{3.2}$$

[14] See, for example, T. D. Newton, Ann. Math. **51**, 730 (1950).

[15] We introduce at this point the notation $\tilde{p}$ for the object with components p_α, distinguishing it from the four- and three-vector operators p and $\mathbf{p}$, respectively.

with

$$K_{\beta\gamma} = M_{\beta\gamma} + T_{\beta\gamma}; \tag{3.3}$$

however, such $M^{\beta\gamma}$ (and $K^{\beta\gamma}$) cannot be found, as is clear from (3.1) and the fact that p_4 is a constant. Nevertheless, one can consider the effect of transformations generated by all ten $T_{\beta\gamma}$ on the equation. Furthermore, despite the above conclusions, we shall see that it is in some ways convenient to regard $\tilde{p}$ as a five-vector quantity and, similarly,

$$p_\alpha p^\alpha \equiv (p_0)^2 - (p_1)^2 - (p_2)^2 - (p_3)^2 - m^2 \tag{3.4}$$

as an $SO(4, 1)$ scalar.

Consider the $SO(4, 1)$ transformation

$$\Gamma_\alpha \to \Gamma'_\alpha = L^\beta_\alpha \Gamma_\beta \; [= (L\Gamma)_\alpha], \tag{3.5}$$

where L^β_α are real and satisfy

$$L^\beta_\alpha g^{\alpha\gamma} L^\delta_\gamma = g^{\beta\delta}, \tag{3.6}$$

$$L^0_0 \geq 1, \tag{3.7}$$

and

$$\det L = 1. \tag{3.8}$$

Then one can write

$$\Gamma'_\alpha = Q\Gamma_\alpha Q^{-1}, \tag{3.9}$$

where

$$Q = \exp\,[(i/2)\omega^{\alpha\beta} T_{\alpha\beta}], \tag{3.10}$$

with the $\omega^{\alpha\beta}$ $(= -\omega^{\beta\alpha})$ real quantities determined by the L^δ_γ. [The converse result also holds: (3.9) and (3.10) $\Rightarrow$ (3.5)–(3.8).] Defining

$$\psi'(x) = Q\psi^{(D)}(x), \tag{3.11}$$

one obtains from (2.10) and (3.9) the equation

$$(L\Gamma)_\alpha p^\alpha \psi' = 0 \tag{3.12}$$

or, equivalently [using (3.6)],

$$\Gamma_\alpha (L^{-1}\tilde{p})^\alpha \psi' = 0. \tag{3.12'}$$

Thus this transformed equation is obtained from (2.10) by replacing therein the "five-vector" $\tilde{p}$ with its transform under the $SO(4, 1)$ transformation inverse to the L of (3.5) [at the same time replacing $\psi^{(D)}$ by ψ' as in (3.11)]. Conversely, if $\tilde{p}'$ is obtained from $\tilde{p}$ by some arbitrary $SO(4, 1)$ transformation, then the equation

$$\Gamma_\alpha p'^\alpha \psi' = 0 \tag{3.13}$$

can be obtained from (2.10) by defining ψ' as in (3.11) with appropriate coefficients $\omega^{\alpha\beta}$ determining Q. Furthermore, since the p_α behave like real numbers to the extent that they commute with one another and with all Γ_β, and have only real eigenvalues on the functions under consideration, a generalization to allow $\omega^{\alpha\beta}$ to be (Hermitian) functions of $\tilde{p}$ is possible.

To summarize: The possible forms (3.13) of Dirac's equation, obtained from (2.10) via transformations of the form (3.11), with $\omega^{\alpha\beta} = \omega^{\alpha\beta}(\tilde{p})$, are determined by the possible transforms $\tilde{p}'$ of $\tilde{p}$,

$$p'_\alpha = L^\beta_\alpha p_\beta, \tag{3.14}$$

where L^β_α are (Hermitian) functions of $\tilde{p}$ satisfying (3.6–3.8). For all such transformations,

$$p'_\alpha p'^\alpha = p_\alpha p^\alpha \equiv p_\mu p^\mu - m^2, \tag{3.15}$$

and

$$\epsilon(p'_0) = \epsilon(p_0). \tag{3.16}$$

At this point we note that knowledge of $\tilde{p}'$ is not sufficient to uniquely determine the transformation L^β_α of (3.14), as for any $\tilde{p}'$ there is a "little group" of such transformations which leave it invariant. Correspondingly, there are transformations of the form (3.10), which, on application to a solution of a given equation (3.13), produce a further solution of the same equation. A further discussion of these questions is given in Sec. 7.

As mentioned in the Introduction, there are five subgroups, which we label $G^{(\alpha)}$, of $SO(4, 1)$, having particular significance when the question is raised of the canonicality of transformations of the form (3.11). $G^{(\alpha)}$ is that subgroup consisting of all $SO(4, 1)$ transformations which leave invariant arbitrary five-vectors whose only nonzero component is the αth. Since every component of the "five-vector" $\tilde{p}$ appears linearly in (2.10), it follows that if $\tilde{p}'$ is obtained from $\tilde{p}$ by an $SO(4, 1)$ transformation (3.14) in $G^{(\beta)}$, the β component of $\tilde{p}$ appears linearly also in (3.13), which we then refer to as "a 'p_β-linear' form of the Dirac equation."

4. $G^{(0)}$ AND "p_0-LINEAR" FORMS

The subgroup $G^{(0)}$ [the maximal compact subgroup $SO(4)$] acts only on the indices 1, 2, 3, and 4. The corresponding generators are T_{ab}, which are Hermitian matrices, and they in fact generate two inequivalent unitary irreducible representations of $SO(4)$, labeled by the two eigenvalues ± 1 of Γ_0 $(= \gamma_5\gamma_0)$, which is effectively a Casimir operator for this subgroup. (Note $[\Gamma_0, T_{ab}] = 0$.) Under the associated transformations (3.14), p_0 and $p_a p^a$ $[= -\omega^2(\mathbf{p}, m)]$ remain separately invariant.

Thus, from

$$\tilde{p} = (p_0, p_1, p_2, p_3, m) \tag{4.1}$$

via $G^{(0)}$ transformations, one can obtain

$$\tilde{p}' = (p_0, p'_1, p'_2, p'_3, p'_4), \tag{4.2}$$

where the $p'_a(p_b)$ are Hermitian and

$$p'_a p'^a = -\omega^2(\mathbf{p}, m). \tag{4.3}$$

The corresponding equation (3.13) is in each case

$$(\Gamma_0 p^0 + \Gamma_a p'^a)\psi' = 0, \tag{4.4}$$

which gives, on multiplication with Γ_0, the general "p_0-linear" form

$$p_0\psi' = \gamma_0(\gamma_1 p'_1 + \gamma_2 p'_2 + \gamma_3 p'_3 + p'_4)\psi'. \tag{4.4'}$$

In considering the canonicality of momentum-dependent transformations in $G^{(0)}$, we note that when

$$\psi^{(D)}(x) \to Q(p, m)\psi^{(D)}(x), \tag{4.5}$$

where $Q(p, m)$ is as in (3.10) [with $\omega^{\alpha\beta} = \omega^{\alpha\beta}(\tilde{p})$], one has, from (1.6),

$$\chi^{(D)\pm}(\mathbf{k}) \to Q(k, m)\chi^{(D)\pm}(\mathbf{k}). \tag{4.6}$$

Furthermore, when in particular Q is in $G^{(0)}$, it follows from the Hermiticity of T_{ab} that $Q(k, m)$ is a unitary matrix. Every such transformation is therefore canonical, preserving the scalar product in the mode (1.8a).

All transformations presented in BC and yielding "p_0-linear" forms are of $G^{(0)}$ type. For example, in the simple cases of the Foldy–Wouthuysen–Tani equation, which corresponds to

$$\tilde{p}' = (p_0, 0, 0, 0, \omega(\mathbf{p}, m)), \tag{4.7}$$

and the Mendlowitz equation, which corresponds to

$$\tilde{p} = \left(p_0, \frac{\omega(\mathbf{p}, m)}{|\mathbf{p}|}p_1, \frac{\omega(\mathbf{p}, m)}{|\mathbf{p}|}p_2, \frac{\omega(\mathbf{p}, m)}{|\mathbf{p}|}p_3, 0\right), \tag{4.8}$$

one sees that the corresponding transformations (1.17), (1.20) are indeed of the form

$$\exp\,[(i/2)\omega^{ab}T_{ab}], \tag{4.9}$$

with the ω^{ab} [$= \omega^{ab}(p_c)$] Hermitian. They are, in fact, the $SO(4, 1)$ transformations (3.10) corresponding to the "little group rotation-free" (l.g.r.f.) transformations $\tilde{p} \to \tilde{p}'$ in the two cases. More generally, corresponding to

$$\tilde{p}' = (p_0, \pm rq_1, \pm rq_2, \pm rq_3, \pm rq_4), \tag{4.10}$$

where the $q_a(p_b)$ are Hermitian,

$$r = \omega(\mathbf{p}, m)[-q_a q^a]^{-\frac{1}{2}}, \tag{4.11}$$

and

$$q_a(p^a - q^a) = 0, \tag{4.12}$$

from (4.4') one has

$$p_0\psi' = \pm\omega(\mathbf{p}, m)[-q_a q^a]^{-\frac{1}{2}}\gamma_0(\gamma_1 q_1 + \gamma_2 q_2 + \gamma_3 q_3 + q_4)\psi', \tag{4.13}$$

which is the general "p_0-linear" form obtained in BC. [There are the $\tilde{p}'$ as in (4.2) which cannot be expressed in the manner of (4.10)–(4.12), viz., those for which $p'_a p^a = 0$. The corresponding transformations and equations were not obtained in BC.] Again, the transformation presented in BC and yielding (4.13) corresponds to the l.g.r.f. transformation of $\tilde{p}$ into $\tilde{p}'$ as in (4.10).

The angles appearing in (1.17) and (1.20) can be regarded as those between the Euclidean "four vectors" p_a and p'_a through which one rotates to obtain p'_a in each case. [The idea of looking upon the Foldy–Wouthuysen–Tani and Mendlowitz transformations as rotations is not new,[16] nor is the use of the group $G^{(0)}$ in discussing them: it is evident in the work of Bollini and Giambiagi,[17] who have not, however, noted the connection with $SO(4, 1)$.]

5. $G^{(4)}$ AND "m-LINEAR" FORMS

$G^{(4)}$ is the $SO(3, 1)$ group associated with homogeneous Lorentz transformations. Thus momentum-dependent $SO(4, 1)$ transformations (3.14) in $G^{(4)}$ leave p_4 ($= m$), $p_\mu p^\mu$, and also $\epsilon(p_0)$ separately invariant. The associated generators $T_{\mu\nu}$ are not all Hermitian, but satisfy

$$T^\dagger_{\mu\nu} = \Gamma_0\Gamma_4 T_{\mu\nu}\Gamma_0\Gamma_4\,(= \gamma_0 T_{\mu\nu}\gamma_0) \tag{5.1}$$

and, as is well known, they generate two inequivalent irreducible representations of $SO(3, 1)$, labeled by the two eigenvalues $\pm i$ of Γ_4 ($= \gamma_5$) (cf. the case of $G^{(0)}$).

From

$$\tilde{p} = (p_0, p_1, p_2, p_3, m), \tag{5.2}$$

via $G^{(4)}$ transformations, one can obtain

$$\tilde{p}' = (p'_0, p'_1, p'_2, p'_3, m), \tag{5.3}$$

where the $p'_\mu(p_\nu)$ are Hermitian,

$$p'_\mu p'^\mu = p_\nu p^\nu, \tag{5.4}$$

and

$$\epsilon(p'_0) = \epsilon(p_0). \tag{5.5}$$

In each case, Eq. (3.13) is

$$(\Gamma_\mu p'^\mu - \Gamma_4 m)\psi' = 0, \tag{5.6}$$

yielding the general "m-linear" form

$$m\psi' = \gamma_\mu p'^\mu \psi'. \tag{5.6'}$$

It should be mentioned at this point that, because of the way it is obtained, p'_μ will not, in general,

[16] See, for example, K. M. Case, Phys. Rev. **95**, 1323 (1954).

[17] C. G. Bollini and J. J. Giambiagi, Nuovo Cimento **21**, 107 (1961). See also Ref. 19. *Note added in proof:* Since the preparation of this paper, E. de Vries [Physica **43**, 45 (1969)] has independently established this connection.

be a Lorentz four-vector operator like p_μ. Our notation is perhaps misleading in this regard: $\tilde{p}$ is better regarded as a numerical five-vector than a five-vector operator for the purposes of this paper.

Each momentum-dependent transformation in $G^{(4)}$ is also canonical. This follows from the fact that (5.1) implies that the corresponding $Q(k, m)$ of (4.6) satisfies

$$Q^{-1}(k, m) = \gamma_0 Q^\dagger(k, m)\gamma_0, \tag{5.7}$$

so that the scalar product in the mode (1.8b) is preserved in every case.

Every transformation presented in BC and leading to an "m-linear" equation is of $G^{(4)}$ type (again corresponding to the l.g.r.f. transformation of $\tilde{p}$ into $\tilde{p}'$ in each case). Thus, for example, in the Chakrabarti case, where

$$\tilde{p}' = (\epsilon(p_0)(p_\mu p^\mu)^{\frac{1}{2}}, 0, 0, 0, m), \tag{5.8}$$

the l.g.r.f. transformation (1.24) is indeed of the form

$$\exp\{(i/2)\omega^{\mu\nu}T_{\mu\nu}\}. \tag{5.9}$$

In fact, *all* $\tilde{p}'$ of the form (5.3) can be written as

$$\tilde{p}' = (rq_0, rq_1, rq_2, rq_3, m), \tag{5.10}$$

with $q_\mu(p_\nu)$ Hermitian, $q_\mu q^\mu$ positive-definite,

$$r = (p_\mu p^\mu)^{\frac{1}{2}}(q_\nu q^\nu)^{-\frac{1}{2}}, \tag{5.11}$$

and

$$q_\mu(p^\mu - q^\mu) = 0. \tag{5.12}$$

[*Proof:* Take $q_\rho = p'_\mu p^\mu (p_\nu p^\nu)^{-1} p'_\rho$, noting that $p'_\mu p^\mu$ is positive-definite because of (5.5).] Then (5.6′) becomes

$$m\psi' = (p_\mu p^\mu)^{\frac{1}{2}}(q_\nu q^\nu)^{-\frac{1}{2}}\gamma_\rho q^\rho \psi', \tag{5.13}$$

which is the general "m-linear" form presented in BC.

Whereas in the case of the compact subgroup $G^{(0)}$ one can talk of an "angle of rotation" for each transformation, here one typically has pseudoangles associated with the anti-Hermitian T_{0i}, generators of "boosts" rather than rotations [cf. (1.24)].

6. $G^{(3)}$ AND "p_3-LINEAR" FORMS

Each of the three subgroups $G^{(i)}$ is again an $SO(3, 1)$ group. Transformations of the form (3.14) within $G^{(3)}$ leave p_3, $p_\tau p^\tau$, and $\epsilon(p_0)$ separately invariant. The corresponding generators $T_{\tau\eta}$ in this case are not all Hermitian, but

$$T^\dagger_{\tau\eta} = \Gamma_0\Gamma_3 T_{\tau\eta}\Gamma_0\Gamma_3 \,(= \gamma_0\gamma_3 T_{\tau\eta}\gamma_0\gamma_3). \tag{6.1}$$

In complete analogy with the case of $G^{(4)}$, the $T_{\tau\eta}$ generate two inequivalent irreducible representations of $SO(3, 1)$, labeled by the two eigenvalues $\pm i$ of Γ_3 $(= \gamma_5\gamma_3)$ in this case.

From

$$\tilde{p} = (p_0, p_1, p_2, p_3, m), \tag{6.2}$$

via $G^{(3)}$ transformations, one can obtain

$$\tilde{p}' = (p'_0, p'_1, p'_2, p_3, p'_4), \tag{6.3}$$

where the $p'_\tau(p_\eta)$ are Hermitian,

$$p'_\tau p'^\tau = p_\eta p^\eta \,[= \lambda^2(p_0, p_1, p_2, m)], \tag{6.4}$$

and

$$\epsilon(p'_0) = \epsilon(p_0). \tag{6.5}$$

(Note that the transformations may become singular as $p_\tau p^\tau \to 0$.) The corresponding equation (3.13) in each case is

$$(\Gamma_\tau p'^\tau - \Gamma_3 p_3)\psi' = 0 \tag{6.6}$$

or, equivalently, the general "p_3-linear" form

$$p_3\psi' = -\gamma_3(\gamma_0 p'_0 - \gamma_1 p'_1 - \gamma_2 p'_2 - p'_4)\psi'. \tag{6.6′}$$

Again in analogy with the $G^{(4)}$ case, one finds that every momentum-dependent transformation in $G^{(3)}$ is canonical, the scalar-product mode [(1.8c); $i = 3$] being preserved in each case as a result of (6.1).

The transformations presented in BC and leading to "p_3-linear" equations are all of $G^{(3)}$ type (in each case corresponding to the l.g.r.f. transformation of $\tilde{p}$ into $\tilde{p}'$). In the simplest case, for example, where the equation is (1.25), corresponding to

$$\tilde{p}' = (\epsilon(p_0)(p_\tau p^\tau)^{\frac{1}{2}}, 0, 0, p_3, 0), \tag{6.7}$$

the l.g.r.f. transformation (1.28) is indeed of the form

$$\exp\{(i/2)\omega^{\tau\eta}T_{\tau\eta}\}. \tag{6.8}$$

Furthermore, all $\tilde{p}'$ as in (6.3) can be written in the form

$$\tilde{p}' = (rq_0, rq_1, rq_2, p_3, rq_4), \tag{6.9}$$

where $q_\tau(p_\eta)$ are Hermitian, $q_\tau q^\tau$ is positive-definite,

$$r = (p_\tau p^\tau)^{\frac{1}{2}}(q_\eta q^\eta)^{-\frac{1}{2}}, \tag{6.10}$$

and

$$q_\tau(p^\tau - q^\tau) = 0. \tag{6.11}$$

[Take $q_\zeta = p'_\tau p^\tau (p_\eta p^\eta)^{-1} p'_\zeta$.] Equation (6.6′) then becomes the general "p_3-linear" form of BC:

$$p_3\psi' = -\lambda(p_0, p_1, p_2, m)(q_\tau q^\tau)^{-\frac{1}{2}} \times \gamma_3(\gamma_0 q_0 - \gamma_1 q_1 - \gamma_2 q_2 - q_4)\psi'. \tag{6.12}$$

Again pseudoangles rather than angles appear in association with the anti-Hermitian generators $T_{0\tau}$ [cf. (1.28)].

7. SIGNIFICANCE OF THE LITTLE GROUP

We have mentioned that for a given "five vector" $\tilde{p}'$ there is a little group of $SO(4, 1)$ transformations which leave it invariant. We are dealing here only with functions on which $p'_\alpha p'^\alpha$ $(= p_\alpha p^\alpha = p_\mu p^\mu - m^2)$ vanishes. In order to identify the little group appropriate in this situation, consider the particular case (corresponding to the Foldy–Wouthuysen–Tani equation)

$$\tilde{p}' = (p_0, 0, 0, 0, \omega(\mathbf{p}, m)). \tag{7.1}$$

It is seen that the matrix generators corresponding to the little group in this case are

$$T_{12}, T_{23}, \text{ and } T_{31}, \tag{7.2}$$

together with

$$T_{10} - \epsilon(p_0)T_{14}, \quad T_{20} - \epsilon(p_0)T_{24}, \quad \text{and} \quad T_{30} - \epsilon(p_0)T_{34}. \tag{7.3}$$

The Lie algebra of (7.2) and (7.3) is isomorphic to that of the three-dimensional Euclidean group, (7.2) being the generators of "rotations," and (7.3) of "translations."

However, from (3.13) and (7.1), we find that $\Gamma_0 - \epsilon(p_0)\Gamma_4$ vanishes on the wavefunctions involved here. Multiplying this by $(i/2)\Gamma_j$ ($j = 1$, 2, or 3), we obtain the result that this is also true of each of the generators (7.3). Thus the little group is effectively reduced to $SU(2)$.[18]

We conclude that, for any given $\tilde{p}'$, the little group consists of an effective part, which is $SU(2)$, and an ineffective part. Any transformation in the ineffective part is unity when applied to a wavefunction satisfying (3.13), while one in the effective part produces a new function satisfying the same equation. Note that $\tilde{p}'$, as in (7.1), can be obtained from $\tilde{p}$ by a transformation in $G^{(0)}$ and that the effective little group generators in this case (7.2) also generate $G^{(0)}$ transformations. This indicates that if (3.13) is obtained from (2.10) by means of a canonical $SO(4, 1)$ transformation (i.e., a transformation in one of the subgroups $G^{(\alpha)}$), a subsequent little-group transformation leaving (3.13) invariant is also canonical.

It is clear that any $\tilde{p}'$ and $\tilde{p}''$ obtained from $\tilde{p}$ by $SO(4, 1)$ tranformations (3.14) must themselves be linked by a further such transformation. Furthermore, by a procedure analogous to that used in obtaining (3.13) from (2.10), it is possible to obtain the equation

$$\Gamma_\alpha p''^\alpha \psi'' = 0 \tag{7.4}$$

directly from

$$\Gamma_\alpha p'^\alpha \psi' = 0. \tag{7.5}$$

Denoting by Q'' and Q' the operators (3.10) used to obtain ψ'' and ψ', respectively, from $\psi^{(D)}$, we have, trivially,

$$\psi'' = Q''Q'^{-1}\psi'. \tag{7.6}$$

While the operator $Q''Q'^{-1}$ certainly corresponds to an $SO(4, 1)$ transformation taking $\tilde{p}'$ into $\tilde{p}''$ and enables one to obtain (7.4) from (7.5), it will not in general correspond to the l.g.r.f. such transformation, even if Q'' and Q' correspond to the l.g.r.f. transformations taking $\tilde{p}$ into $\tilde{p}''$, $\tilde{p}'$, respectively. More precisely, if we denote by $Q(\tilde{p}', \tilde{p})$, $Q(\tilde{p}'', \tilde{p})$, and $Q(\tilde{p}'', \tilde{p}')$ the operators (3.10) corresponding to the l.g.r.f. transformations taking $\tilde{p}$ into $\tilde{p}'$, $\tilde{p}$ into $\tilde{p}''$, and $\tilde{p}'$ into $\tilde{p}''$, respectively, then in general

$$Q(\tilde{p}'', \tilde{p}') = \Lambda Q(\tilde{p}'', \tilde{p})Q^{-1}(\tilde{p}', \tilde{p}), \tag{7.7}$$

where Λ is also of the form (3.10) and corresponds to an $SO(4, 1)$ transformation in the little group defined by $\tilde{p}''$. If Λ is in the effective part of the little group, then $Q(\tilde{p}'', \tilde{p}')$ and $Q(\tilde{p}'', \tilde{p})Q^{-1}(\tilde{p}', \tilde{p})$ will differ on the wavefunctions ψ'; but if it is in the ineffective part, then these two operators, while perhaps differing formally, will produce the same result when applied to any such ψ' satisfying (7.5).

As an example, consider the case when $\tilde{p}'$ is as in (7.1), and

$$\tilde{p}'' = (\epsilon(p_0)(p_\mu p^\mu)^{\frac{1}{2}}, 0, 0, 0, m), \tag{7.8}$$

corresponding to the Chakrabarti equation. It is seen that in this case the l.g.r.f. transformation taking $\tilde{p}'$ into $\tilde{p}''$ is in the 0–4 plane, and correspondingly,

$$Q(\tilde{p}'', \tilde{p}') = \exp(i\varphi T_{04}) \tag{7.9}$$

$$= \cosh(\tfrac{1}{2}\varphi) - \gamma_0 \sinh(\tfrac{1}{2}\varphi), \tag{7.9'}$$

where $\varphi(\tilde{p})$ is Hermitian. A straightforward calculation of the pseudoangle φ involved in this "boost" transformation yields

$$\varphi(\tilde{p}) = \operatorname{arc\,tanh}\{\epsilon(p_0)[\mathcal{E}^2 - \mathcal{M}^2][\mathcal{E}^2 + \mathcal{M}^2]^{-1}\}, \tag{7.10}$$

where

$$\mathcal{E} = \tfrac{1}{2}[|p_0| + \omega(\mathbf{p}, m)] \tag{7.11}$$

and

$$\mathcal{M} = \tfrac{1}{2}[(p_\mu p^\mu)^{\frac{1}{2}} + m]. \tag{7.12}$$

Then

$$\cosh(\tfrac{1}{2}\varphi) = \tfrac{1}{2}\{[\mathcal{E}/\mathcal{M}]^{\frac{1}{2}} + [\mathcal{M}/\mathcal{E}]^{\frac{1}{2}}\} \tag{7.13}$$

and

$$\sinh(\tfrac{1}{2}\varphi) = [\tfrac{1}{2}\epsilon(p_0)]\{[\mathcal{E}/\mathcal{M}]^{\frac{1}{2}} - [\mathcal{M}/\mathcal{E}]^{\frac{1}{2}}\}. \tag{7.14}$$

[18] There is a marked analogy here with the case of the little group (in the usual connection with the Poincaré group now) appropriate to a particle of zero rest mass and nonzero spin, as treated by V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **34**, 211 (1948). See also de Vos and Hilgevoord (Ref. 5) in this connection.

Because $\tilde{p}'$ in this case corresponds to the Foldy–Wouthuysen–Tani equation, Eq. (7.5) yields

$$\mathcal{M}\psi' = m\psi', \tag{7.15}$$

$$\mathcal{E}\psi' = \omega(\mathbf{p}, m)\psi', \tag{7.16}$$

and

$$\epsilon(p_0)\gamma_0\psi' = \psi'. \tag{7.17}$$

It follows then from (7.9′) and (7.13)–(7.17) that

$$Q(\tilde{p}'', \tilde{p}')\psi' = [m/\omega(\mathbf{p}, m)]^{\frac{1}{2}}\psi'. \tag{7.18}$$

However, in this example we have

$$Q(\tilde{p}', \tilde{p}) \equiv F(\mathbf{p}, m) \tag{7.19}$$

and

$$Q(\tilde{p}'', \tilde{p}) \equiv C(p), \tag{7.20}$$

and it is known that[19]

$$\psi^{(C)}(x)\,[= C(p)F^{-1}(\mathbf{p}, m)\psi^{(F)}(x)] = [m/\omega(\mathbf{p}, m)]^{\frac{1}{2}}\psi^{(F)}(x). \tag{7.21}$$

Thus from (7.18) we deduce

$$Q(\tilde{p}'', \tilde{p}')\psi^{(F)} = C(p)F^{-1}(\mathbf{p}, m)\psi^{(F)} \tag{7.22}$$

$$= Q(\tilde{p}'', \tilde{p})Q^{-1}(\tilde{p}', \tilde{p})\psi^{(F)}. \tag{7.22′}$$

On inspection, however, $Q(\tilde{p}'', \tilde{p}')$ and $Q(\tilde{p}'', \tilde{p}) \times Q^{-1}(\tilde{p}', \tilde{p})$ are found to be formally distinct, and we conclude that they are related in the manner (7.7), with Λ in the ineffective part of the little group of $\tilde{p}''$.

8. CONCLUSION

The connection between the group $SO(4, 1)$ and the free-particle Dirac equation can be exploited to allow the presentation of a unified treatment of the well-known canonical transformations of the equation. Similarities and relationships between these transformations assume a new and simple significance in such a treatment.

This approach also makes obvious the existence and also the actual form of many other similar canonical transformations, some of which we feel will prove useful[8] in discussing limiting situations other than the nonrelativistic and extreme-relativistic ones (where the Foldy–Wouthuysen–Tani and the Mendlowitz transformations, respectively, are most appropriate).

It is tempting to speculate as to a deeper physical significance of the group $SO(4, 1)$ itself in this context, in view of recent activity centering on this and related groups in connection with dynamical symmetries[4,20] for elementary particles. However, there are relativistic wave equations, linear in the energy–momentum operators, for which there is no simple connection with $SO(4, 1)$.[21] For all such equations describing massive particles, it is, however, a consequence of Lorentz invariance that there will be a Chakrabarti-type transformation corresponding to the transformation of the four momentum to the rest frame: Such a transformation expresses the canonical Wigner amplitudes in terms of the manifestly covariant ones.

Foldy and Wouthuysen[10] generalized their approach to the free-particle Dirac equation to gain considerable insight into the problem of the Dirac particle in interaction with a weak electromagnetic field, and this approach has been pursued consequently by several authors.[22] The Foldy–Wouthuysen method involves a perturbation procedure and yields a "p_0-linear" equation containing an infinite number of terms. In the absence of the interaction, this equation reduces to their form of the free-particle equation. One of us (H. A. C.) has generalized this procedure to develop similar expansions corresponding to various other forms obtainable from the free-particle Dirac equation via canonical $SO(4, 1)$ transformations.

ACKNOWLEDGMENTS

It is a pleasure to record stimulating discussions with Professor H. S. Green and Professor C. A. Hurst, and with C. J. Grigson, G. J. Iverson, M. K. James, and J. Spitzer. One of us (A. J. B.) wishes to acknowledge the financial support provided by a C.S.I.R.O. Senior Postgraduate Studentship.

[19] R. H. Good, Jr., and M. E. Rose, Nuovo Cimento **24**, 864 (1962).

[20] See, for example, A. Böhm, Phys. Rev. **175**, 1767 (1968); C. Fronsdal, Phys. Rev. **156**, 1665 (1967); Y. Nambu, Phys. Rev. **160**, 1171 (1967); A. O. Barut, Phys. Rev. Letters **26B**, 308 (1968).

[21] We are indebted to Professor C. A. Hurst for this information.

[22] See E. Eriksen and M. Kolsrud, Nuovo Cimento Suppl. **18**, 1 (1960), and references therein.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Internal-Labeling Problem*

R. T. SHARP AND C. S. LAM
Physics Department, McGill University, Montreal, Canada

(Received 21 February 1969)

A method is proposed for labeling the bases of a compact group when reduced with respect to an arbitrary subgroup. The scheme is based on the observation that the heaviest state of a multiplet (subgroup irreducible representation) of an IR (group irreducible representation) can be labeled by a product of heaviest states of simpler "elementary" multiplets. Details are worked out for a number of group–subgroup combinations.

1. INTRODUCTION

A recurring problem in the application of group theory to physics is the reduction of the irreducible representations (IR's) of a group into irreducible representations (multiplets) of a subgroup.[1] (For clarity, we use different words for the irreducible representations of the group and its subgroup.) One aspect, the internal-multiplicity problem, is solved by a simple enumeration of the multiplets in each IR. To solve the more general internal-labeling (or state-labeling) problem, it is necessary to specify the multiplets, i.e., define an actual basis for the IR's. We are concerned with the labeling problem in this paper; it must be solved, of course, in order to define and evaluate generator and transformation matrix elements, Clebsch–Gordan coefficients, etc.

Racah[2] showed that the number of internal labels required to specify the basis states of the general IR of a compact group is $\frac{1}{2}(r - l)$, where r is the order of the group (number of generators) and l its rank (number of suitably chosen commuting generators). In general, the nature of these labels and the values they may assume for a particular IR is an unsolved problem. If there is a subgroup which, together with other generators which commute with it and each other, provides the right number of labels, the basis states of the subgroup multiplets may be used as basis states for IR's of the group, solving the internal-labeling problem; such a scheme is called mathematically canonical. (We do not use "mathematically canonical" in the sense "having among the labels a number of commuting generators equal to the rank of the group.") In general, however, no suitable subgroup exists, and even if one does, the states of physical interest may correspond to multiplets of a different subgroup which provides too few labels.

In Sec. 2 we describe an approach which, in principle, solves the internal-labeling problem for any compact group and subgroup. In Sec. 3, a number of specific examples are worked out in detail and we refer to other methods which have been applied to some of these examples by various authors.

2. ELEMENTARY MULTIPLETS AND FACTORS

We follow Cartan's method[3] of constructing the IR $(\lambda_1, \cdots, \lambda_l)$ of a compact group of rank l as the stretched product of the "simple" IR's $(\lambda_1, 0, \cdots, 0), \cdots, (0, \cdots, 0, \lambda_l)$. The stretched product is defined as the IR in the direct product which contains the product of the heaviest states of the factor IR's. The simple IR's $(\lambda_1, 0, \cdots, 0), \cdots, (0, \cdots, 0, \lambda_l)$ are, here, polynomials of degree $\lambda_1, \cdots, \lambda_l$, respectively, in the basis states of the l fundamental IR's $(1, 0, \cdots, 0), \cdots, (0, \cdots, 0, 1)$. For simplicity, polynomials are used for the IR $(0, \cdots, 0, \lambda_i, 0, \cdots, 0)$ rather than the stretched product of λ_i independent copies of the ith-fundamental IR. This has the incidental advantage of eliminating all but completely symmetric IR's and requiring fewer variables. (Since we are not concerned with representation matrices in this paper, no confusion should arise when "IR" or "multiplet" is used for the longer "basis for IR" or "basis for multiplet.") The IR $(\lambda_1, \cdots, \lambda_l)$ thus consists of polynomials of degrees $\lambda_1, \cdots, \lambda_l$ in the respective fundamental IR. States of degree $\lambda_1, \cdots, \lambda_l$ but belonging to IR's lower than $(\lambda_1, \cdots, \lambda_l)$ are called "unwanted" and are to be discarded in general.

The product of the heaviest states of two or more multiplets from the same or different IR's defines, in general, a multiplet of a higher IR, specifically,

* Research supported by the National Research Council of Canada.

[1] For a discussion of the internal-labeling problem with special reference to $SU(N)$ see L. C. Biedenharn, Boulder Lectures at the Summer Institute of Theoretical Physics 5, 258 (1962).

[2] G. Racah, lecture notes on "Group Theory and Spectroscopy," Institute for Advanced Study, Princeton, N.J., 1951; reprinted in Ergeb. Exakt. Naturw. 37, 28 (1965).

[3] We refer to the fact that such a construction is possible as Cartan's fundamental theorem. See E. Cartan, Thèse Paris (1894) reprinted in E. Cartan, *Oeuvres complètes* (Gauthier–Villars, Paris, 1952). Discussions of the result are found in Refs.1 and 2 and in most books on group representations.

the product will be the heaviest state of the multiplet so defined when unwanted states have been projected out of it. Our solution of the labeling problem is based on the observation that a complete set of multiplets of all IR's may be defined in this way through products of powers of certain "elementary factors." The elementary factors evidently include the heaviest states of all the multiplets in the fundamental IR's. In the case of $SU(N) \supset SU(N-1)$, no others are required; in general, elementary factors belonging to higher IR's are needed. They may be found by proceeding systematically through higher IR's. When a multiplet known to be present is not given through the elementary factors already at hand, its heaviest state must be included as a new elementary factor.

The correspondence between multiplets and products of powers of elementary factors is one-to-one when certain relations are taken into account. These relations arise when a linear superposition of products of powers of elementary factors vanishes identically or is equal to an unwanted expression. In each such case, to avoid duplication of multiplets, one term in the superposition must be singled out and regarded as redundant.

For each group–subgroup combination which we have considered, a finite number of elementary factors suffice; the relations can be taken into account in a systematic way by regarding certain combinations of elementary factors as incompatible or redundant by themselves or for the purpose of forming higher multiplets. In all such cases, a solution of the internal-labeling problem is thus obtained in closed form. We are unable to prove that a finite number of elementary factors suffices in general; but even if an infinite number is required, a finite number of them solves the labeling problem up to IR's of any preassigned degree.

A nonredundant product of powers of elementary factors must have unwanted states projected out of it before it is the heaviest state of the multiplet it defines. Equivalently, it is the leading (zero-order) term in the expansion of the state in powers of redundant combinations arising from unwanted expressions. ("Redundant combinations" here include quantities which must be added to the elementary factors themselves to render them free of unwanted states.) The product of powers thus proves useful as a handle for manipulation of the complete state. For example, to expand a state known to be free of unwanted states in basis states labeled by products of powers of elementary factors (as in the determination of generator and transformation matrices, Clebsch–Gordan coefficients, etc.), it is necessary only to pick out the coefficients of the nonredundant products of elementary multiplets.

The labeling procedure described above is applied to particular cases in the next section. The new labels and their relation to more conventional quantum numbers are dealt with under individual cases.

Although we follow Cartan in constructing bases of a group by using its l fundamental IR's, the approach could as well be used with Gel'fand-type bases, i.e., ones which are polynomials in l independent copies of one basic or defining IR. Indeed, particular cases of states of Gel'fand-type, constructed with the help of elementary factors, appear in the literature and will be referred to under the appropriate headings [$SU(3) \supset O(3)$ and $SU(N) \supset SU(N-1)$] in the next section.

Methods analogous to those of this paper can be used for labeling the IR's of a group which are spanned by polynomials in the states of a few specific IR's of the same group.

The external-labeling problem for any group G can be treated as a special case of the internal-labeling problem for the group–subgroup combination $G \times G \supset G$.

3. EXAMPLES

In the following examples, Cartan's $\lambda_1, \cdots, \lambda_l$ are used to label the irreducible representations of groups and subgroups except for $SU(2) \sim O(3)$ and $O(4) \sim SU(2) \times SU(2)$ which, by tradition, are labeled by $j = \frac{1}{2}\lambda$ and $j_1 = \frac{1}{2}\lambda_1$, $j_2 = \frac{1}{2}\lambda_2$, respectively. An abbreviated notation for an elementary factor is $(a; b)$, where a denotes the IR and b the multiplet to which it belongs.

We discuss $SU(3) \supset O(3)$ in some detail since it is perhaps the simplest case which illustrates the general features; the version $SU(3) \supset SU(2)$ which describes the $SU(3)$ symmetry of elementary particles is not considered explicitly, since it is a special case of $SU(N) \supset SU(N-1)$ below.

Case 1: $SU(3) \supset O(3)$ describes the Elliott–Harvey classification of nuclear states[4]; $O(3)$ refers to angular momentum. The subgroup provides one label too few.

The states of the fundamental IR's are shown in Fig. 1. The elementary factors are $(10; 1) = \eta$,

[4] For a discussion of the nuclear $SU(3)$ model and earlier references, see M. Harvey, Advan. Nucl. Phys. **1**, 67 (1968). Elliott solves the labelling problem by projecting good orbital angular momentum states out of certain "intrinsic" states which are simple $SU(3) \supset SU(2)$ states (those with maximum hypercharge) and retaining a nonredundant set. V. Bargmann and M. Moshinsky [Nucl. Phys. **23**, 177 (1961)] and M. Moshinsky and V. Syamala Devi [J. Math. Phys. **10**, 455 (1969)] use elementary factors to obtain $SU(3) \supset O(3)$ states of Gel'fand type equivalent to those defined in the present paper. R. T. Sharp and H. C. von Baeyer [Nucl. Phys. (to be published)] give the transformation matrices connecting Elliott–Harvey and Bargmann–Moshinsky states.

FIG. 1. The fundamental IR's of $SU(3)$; heaviest states of $O(3)$ multiplets are denoted by X.

$(20;0) = \zeta^2 - 2\eta\xi$, $(01;1) = \xi^*$, $(02;0) = \zeta^{*2} - 2\eta^*\xi^*$, and $(11;1) = \eta\zeta^* + \xi^*\zeta$. There is one relation; $(11;1)^2$ is redundant, i.e., $(11;1)$ appears at most linearly.

The elementary factors may be found systematically as follows: First, $(10;1) = \eta$ is needed as the heaviest state of the multiplet in the fundamental IR (10); combining the multiplet $(10;1)$ with itself yields a $j = 2$ multiplet whose heaviest state is η^2 (evidently not an elementary factor) and an $O(3)$ scalar, $(20;0) = \zeta^2 - 2\eta\xi$, which is a new elementary factor. The $j = 1$ multiplet, which might be thought to arise, in fact vanishes because of antisymmetry. The two elementary factors $(10;1)$ and $(20;0)$ are sufficient to generate all IR's of the form $(\lambda_1, 0)$. Similarly, $(01;1) = \xi^*$ and $(02;0) = \zeta^{*2} - 2\eta^*\xi^*$ generate all IR's of the form $(0, \lambda_2)$. The multiplets $(10;1)$, $(01;1)$ can be combined to give three multiplets with $j = 2$, 1, 0, respectively; their heaviest states are, respectively, $\eta\xi^*$, $(11;1) = \eta\zeta^* + \xi^*\zeta$, and $B = \eta\eta^* + \zeta\zeta^* + \xi\xi^*$. Of these, only the second is a new elementary factor; the first is a composite of the elementary factors η and ζ^*, and the third is an unwanted expression belonging to the IR (00). Because of the relation

$$(11;1)^2 = 2B(10;1)(01;1) + (10;1)^2(02;0) + (01;1)^2(20;0),$$

we regard $(11;1)^2$ as redundant.

The general multiplet corresponds to the product $(10;1)^a(01;1)^{a'}(20;0)^b(02;0)^{b'}(11;1)^c$, where $c = 0, 1$ only. The labels are related to the more familiar $\lambda_1\lambda_2 L$ by $\lambda_1 = a + 2b + c$, $\lambda_2 = a' + 2b' + c$, $L = a + a' + c$. The index c is not to be regarded as an independent label; it is 0 or 1 depending on whether $\lambda_1 + \lambda_2 - L$ is even or odd. For the simple IR's $(\lambda_1, 0)$ and $(0, \lambda_2)$, only ab and $a'b'$, respectively, are needed; of the conventional labels, $\lambda_1 L$ and $\lambda_2 L$ then suffice.

To show that the states thus defined form a basis, it is necessary to note that they are linearly independent and then to count them. The counting here is simple enough—it leads to the correct dimension formula:

$$D(\lambda_1, \lambda_2) = \tfrac{1}{2}(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_1 + \lambda_2 + 2).$$

As a model for more complicated groups, e.g., cases

FIG. 2. The fundamental IR's of $O(4)$; heaviest states of $O(3)$ multiplets are denoted by X.

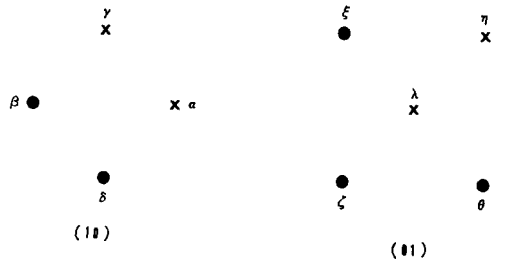

FIG. 3. The fundamental IR's of $O(5)$ with heaviest states of $O(4)$ multiplets denoted by X.

7 and 8 below, it is easier to do the counting for the quantity

$$\Delta^2_{\lambda_1\lambda_2} D = D(\lambda_1 + 2, \lambda_2 + 2) - D(\lambda_1, \lambda_2 + 2) - D(\lambda_1 + 2, \lambda_2) + D(\lambda_1, \lambda_2)$$

instead of for D itself. It may be seen that $\Delta^2_{\lambda_1\lambda_2} D$ is just the number of states with $b = b' = 0$ in the IR $(\lambda_1 + 2, \lambda_2 + 2)$, i.e., the number of states in the multiplets $(10;1)^{\lambda_1+2}(01;1)^{\lambda_2+2}$ and

$$(10;1)^{\lambda_1+1}(01;1)^{\lambda_2+1}(11;1),$$

whose dimensions are, respectively, $2\lambda_1 + 2\lambda_2 + 9$ and $2\lambda_1 + 2\lambda_2 + 7$. This agrees with the result $4(\lambda_1 + \lambda_2 + 4)$, obtained for $\Delta^2_{\lambda_1\lambda_2} D$ by using the known dimension formula. To complete the argument, one can verify directly that the elementary factors lead to the correct dimension formula for those boundary cases in which either λ_1 or λ_2 is equal to 0 or 1.

Case 2: $O(4) \supset O(3)$: special IR's ($j_1 = j_2$) describe the states of the hydrogen atom; $O(3)$ refers to orbital angular momentum.

The states of the fundamental IR's are shown in Fig. 2. The elementary factors are $(\frac{1}{2}0;\frac{1}{2}) = \eta_1$, $(0\frac{1}{2};\frac{1}{2}) = \eta_2$, $(\frac{1}{2}\frac{1}{2};0) = \eta_1\xi_2 - \eta_2\xi_1$. There are no relations or unwanted expressions. The heaviest state of the general multiplet is $\eta_1^a\eta_2^{a'}(\eta_1\xi_2 - \eta_2\xi_1)^b$. The labels are related to the more familiar j_1, j_2, j by $j_1 = \frac{1}{2}(a + b)$, $j_2 = \frac{1}{2}(a' + b')$, $j = \frac{1}{2}(a + a')$.

Case 3: $O(5) \sim O(4)$: the mathematically canonical classification of $O(5)$ states.[5,6]

The states of the fundamental IR's are shown in Fig. 3. The elementary factors are $(10;\frac{1}{2}0) = \alpha$, $(10;0\frac{1}{2}) = \gamma$, $(01;\frac{1}{2}\frac{1}{2}) = \eta$, $(01;00) = \lambda$; a number of other candidates are redundant because of unwanted

[5] Canonical $O(5) \supset O(4)$ states have been derived by K. T. Hecht [Nucl. Phys. **63**, 177 (1965)], by R. T. Sharp and S. C. Pieper [J. Math. Phys. **9**, 663 (1968)], and by N. Kemmer, D. L. Pursey, and S. A. Williams [J. Math. Phys. **9**, 1224 (1968)]. Since $O(4)$ is a mathematically canonical subgroup of $O(5)$, the states of different authors can differ only by phase factors.

[6] K. T. Hecht, Nucl. Phys. **63**, 177 (1965).

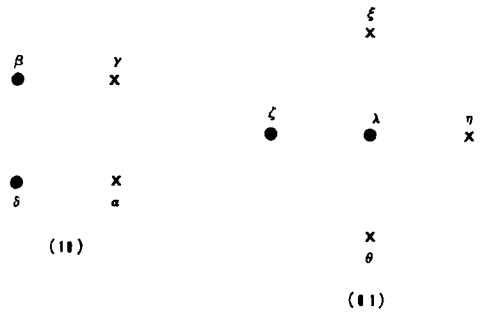

FIG. 4. The fundamental IR's of $O(5)$ with heaviest states of $SU(2)$ multiplets denoted by X.

expressions. The general $O(4)$ multiplet is defined by the product $\alpha^a\gamma^b\eta^c\lambda^d$. The labels are related to the more familiar $\lambda_1\lambda_2 j_1 j_2$ by $\lambda_1 = a + b$, $\lambda_2 = c + d$, $j_1 = \frac{1}{2}(a + c)$, and $j_2 = \frac{1}{2}(b + c)$. For the simple IR's $(\lambda_1, 0)$ and $(0, \lambda_2)$ only ab and cd, respectively, are needed.

Case 4: $O(5) \supset SU(2)$ describes the Hecht–Parikh classification of nuclear states[7] (seniority model); $SU(2)$ refers to isospin and the quantum number V below, plotted vertically in Fig. 4, is related linearly to the number of nucleons. The subgroup, with V, provides one label too few.

The elementary factors [the notation is $(\lambda_1\lambda_2; jV)$] are

$$(10; \tfrac{1}{2}\tfrac{1}{2}) = \gamma, \quad (10; \tfrac{1}{2} - \tfrac{1}{2}) = \alpha,$$
$$(01; 01) = \xi, \quad (01; 10) = \eta,$$
$$(01; 0 - 1) = \theta, \quad (20; 00) = \gamma\delta - \alpha\beta.$$

On account of the unwanted expression

$$(20; 00)(01; 10) + (10; \tfrac{1}{2} - \tfrac{1}{2})^2(01; 01) - (10; \tfrac{1}{2}\tfrac{1}{2})^2(01; 0 - 1),$$

the combination $(20; 00)(01; 10)$ should be regarded as redundant. Thus the general $SU(2)$ multiplet is defined by the product $(10; \frac{1}{2}\frac{1}{2})^a(10; \frac{1}{2} - \frac{1}{2})^b(01; 10)^c (01; 01)^d(01; 0 - 1)^e(20; 00)^f$, with either $c = 0$ or $f = 0$. The more conventional labels are given by $\lambda_1 = a + b + 2f$, $\lambda_2 = c + d + e$, $j = \frac{1}{2}(a + b) + c$, and $V = \frac{1}{2}(a - b) + d - e$. For the simple IR's $(\lambda_1, 0)$ and $(0, \lambda_2)$, the only labels needed are abf and cde, respectively; then the conventional labels $\lambda_1 jV$ and $\lambda_2 jV$ are enough.

[7] For a discussion of the $O(5)$ nuclear model and earlier references, see K. T. Hecht, Nucl. Phys. **A102**, 11 (1967), and R. P. Hemenger, University of Michigan preprint 07591-3-T, 1968. K. T. Hecht, Ref. 6, discusses the transformation between $O(5) \supset O(4)$ and $O(5) \supset SU(2)$ states for some simple IR's. K. Ahmed and R. T. Sharp [J. Math. Phys. (to be published)] have derived general states based on the approach of the present paper, but taking as redundant the combination $(10; \frac{1}{2} - \frac{1}{2})^2 \times (01; 01)$ rather than $(20; 00)(01; 10)$, as proposed in the text.

FIG. 5. The fundamental IR's of $O(5)$ with heaviest states of $SU(2)'$ multiplets denoted by X.

Case 5: $O(5) \supset SU(2)'$: IR's $(0, \lambda_2)$ describe classification of nuclear surfon states, i.e., quantized nuclear surface excitations[8]; $SU(2)'$ refers to orbital angular momentum. The prime is to distinguish the $SU(2)$ subgroup from that of Case 4 above. The subgroup provides two labels too few. Since this case is complicated (more than twenty elementary factors), we only consider the simple IR's $(\lambda_1, 0)$ and $(0, \lambda_2)$; for them the subgroup provides one label too few.

The fundamental IR's are shown in Fig. 5. The elementary factors required for $(\lambda_1, 0)$ IR's are $(10; \frac{3}{2})$, $(20; 1)$, $(30; \frac{3}{2})$, and $(40; 0)$, with $(30; \frac{3}{2})^2$ redundant; we refrain from giving elementary factors and relations in polynomial form. The general $(\lambda_1, 0)$ type $SU(2)'$ multiplet is thus $(10; \frac{3}{2})^a(20; 1)^b$, $(30; \frac{3}{2})^c(40; 0)^d$ with $c = 0$ or 1. The conventional labels are $\lambda_1 = a + 2b + 3c + 4d$, $j = \frac{3}{2}(a + c) + b$. The elementary factors required for $(0, \lambda_2)$ IR's are $(01; 2)$, $(02; 2)$, $(03; 3)$, and $(03; 0)$, with $(03; 3)^2$ redundant. The general $(0, \lambda_2)$ type $SU(2)'$ multiplet is thus defined by $(01; 2)^{a'}(02; 2)^{b'}(03; 3)^{c'}(03; 0)^{d'}$ with $c' = 0$ or 1. The conventional labels are $\lambda_2 = a' + 2b' + 3(c' + d')$ and $j = 2(a' + b') + 3c'$.

Case 6: $O(6) \supset O(5)$: a mathematically canonical scheme for classifying $O(6)$ states.

The states of the fundamental IR's are shown in Fig. 6. The elementary factors are

$$(100; 10) = \alpha, \quad (010; 01) = \eta, \quad (010; 00) = \chi,$$
$$(001; 10) = \beta^*, \quad (101; 01) = \alpha\delta^* - \gamma\beta^*.$$

Thus the general $O(5)$ multiplet is defined by the product

$$(100; 10)^a(010; 01)^b(010; 00)^c(001; 10)^{a'}(101; 01)^d.$$

The labels are related to the IR labels $\lambda_1\lambda_2\lambda_3$ and multiplet labels $\lambda_1'\lambda_2'$ by $\lambda_1 = a + d$, $\lambda_2 = b + c$:

$$\lambda_3 = a' + d, \quad \lambda_1' = a + a', \quad \lambda_2' = b + d.$$

Case 7: $O(6) \simeq SU(4) \supset O(4)$: the Wigner supermultiplet scheme[9]; $O(4) \simeq SU(2) \times SU(2)$ describes

[8] $O(5) \supset SU(2)'$ states of the physically interesting IR's $(0, \lambda_2)$ are discussed by K. T. Hecht, Ref. 6, and by S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).

[9] The Wigner supermultiplet group has often been discussed in the literature. A graphical solution of the internal multiplicity problem is given by A. M. Perelomov and V. S. Popov, Yad. Fiz. **2**, 738 (1965) [Sov. J. Nucl. Phys. **2**, 528 (1966)]; explicit states of Gel'fand type have recently been constructed by M. Resnikoff (report of work prior to publication).

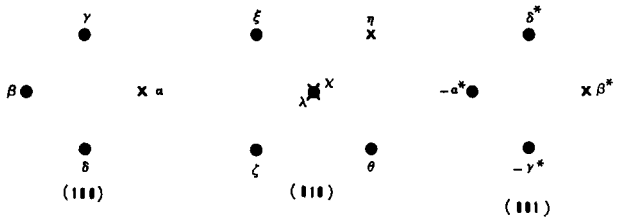

FIG. 6. The fundamental IR's of $O(6)$ with heaviest states of $O(5)$ multiplets denoted by X.

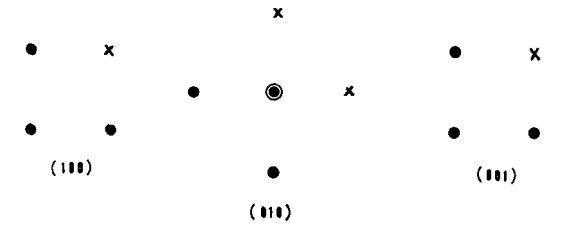

FIG. 7. The fundamental IR's of $O(6)$ with heaviest states of $O(4)$ multiplets denoted by X.

spin and isospin. The subgroup provides two labels too few for the general IR, one too few for IR's in which one of $\lambda_1\lambda_2\lambda_3$ vanishes, just enough for the simple IR's $(\lambda_2 00)$, $(0\lambda_2 0)$, $(00\lambda_3)$.

The states of the fundamental IR's are shown in Fig. 7. The elementary factors, whose algebraic form we omit, are $(100; \frac{1}{2}\frac{1}{2})$, $(200; 00)$, $(010; 10)$, $(010; 01)$, $(020; 00)$, $(001; \frac{1}{2}\frac{1}{2})$, $(002; 00)$, $(101; 10)$, $(101; 01)$, $(110; \frac{1}{2}\frac{1}{2})$, $(011; \frac{1}{2}\frac{1}{2})$, $(11\bar{1}; 10)$, and $(111; 01)$. The redundant combinations are $(111; 10)(101; 01)$; $(111; 01)(101; 10)$; $(101; 10)(101; 01)$; $(110; \frac{1}{2}\frac{1}{2})^2$; $(011; \frac{1}{2}\frac{1}{2})^2$; $(110; \frac{1}{2}\frac{1}{2})(011; \frac{1}{2}\frac{1}{2})$; $(110; \frac{1}{2}\frac{1}{2})(001; \frac{1}{2}\frac{1}{2})$; $(011; \frac{1}{2}\frac{1}{2})(100; \frac{1}{2}\frac{1}{2})$; and $(100; \frac{1}{2}\frac{1}{2})(001; \frac{1}{2}\frac{1}{2})(020; 00)$; and either $(111; 10)$ or $(111; 01)$ with any of $(100; \frac{1}{2}\frac{1}{2})$, $(001; \frac{1}{2}\frac{1}{2})$, $(110; \frac{1}{2}\frac{1}{2})$, $(011; \frac{1}{2}\frac{1}{2})$, $(111; 10)$, and $(111; 01)$.

We omit relating labels which are the indices of elementary factors to the conventional $\lambda_1\lambda_2\lambda_3 j_1 j_2$; the connection is easily written down, since the conventional labels are additive for powers of elementary factors.

Case 8: $G_2 \supset SU(3)$: the exceptional group, at one time considered a candidate for describing particle symmetries[10]; it has no mathematically canonical subgroup. The subgroup $SU(3)$ provides one label too few for the general IR, the right number for the simple IR's $(\lambda_1 0)$, $(0\lambda_2)$.

The fundamental IR's are shown in Fig. 8. The elementary factors are just the heaviest states of the multiplets of the fundamental IR's, i.e., $(10; 10)$, $(10; 01)$, $(10; 00)$, $(01; 10)$, $(01; 01)$, and $(01; 11)$; there is one redundant combination, $(10; 00)(01; 11)$. The conventional labels $\lambda_1\lambda_2\lambda_1'\lambda_2'$ are easily expressed in terms of the exponents of elementary factors, since they are additive for such products of powers.

Case 9: $SU(N) \supset SU(N-1)$; possibly the most studied group–subgroup chain[11]; includes $SU(3) \supset SU(2)$, which is used to classify elementary particles.

[10] Attempts to use G_2 to describe particle symmetries are described by R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1967). They also give the fundamental IR's, the dimension formula, and some simple Clebsch–Gordan series.

$SU(N-1)$ is a canonical subgroup, for, together with the "N-hypercharge" Z_N, it provides enough internal labels for $SU(N)$ states.

There are $N-1$ fundamental IR's. The ith-fundamental IR $(i = 1, \cdots, N-1)$ has one column of i boxes in its Young diagam; it contains two multiplets, which we denote by $(i; 1)$ and $(i; 2)$. For $(i; 1)$, the N-hypercharge Z_N has the value i/N, and the multiplet is the ith-fundamental irreducible representation of $SU(N-1)$, except for $i = N-1$ when it is an $SU(N-1)$ scalar. For $(i; 2)$, the value of Z_N is $i/N-1$, and the multiplet is the $(i-1)$th-fundamental irreducible representation of $SU(N-1)$,

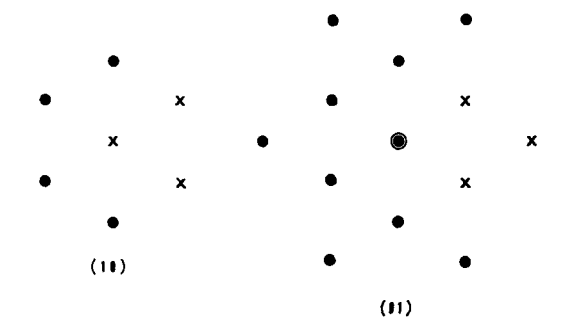

FIG. 8. The fundamental IR's of G_2; heaviest states of $SU(3)$ multiplets are denoted by X.

except for $i = 1$ when it is an $SU(N-1)$ scalar. The elementary factors are just the heaviest states of the $2N-2$ multiplets of the $N-1$ fundamental IR's. There are no redundant combinations. Expressions for the $2N-2$ conventional labels $\lambda_1, \cdots, \lambda_{N-1}$, Z_N, $\lambda_1', \cdots, \lambda_{N-1}'$ are easily found in terms of the exponents of the $2N-2$ elementary factors, since the conventional labels are additive for products of powers of elementary factors.

Case 10: $G \supset U(1)$; here one is, in effect, using no subgroup at all to help label the states.

[11] Heaviest members of $SU(N) \supset SU(N-1)$ multiplets of Gel'fand type are given explicitly by G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963), Eq. 50. The states involves elementary polynomials in boson creation operators (operating on a vacuum ket) similar to the elementary factors of the present paper.

The elementary factors are all the states of all the fundamental IR's. In general, certain combinations are redundant because of unwanted expressions. As an example, we consider the labeling of $SU(3)$ states. The fundamental IR's are shown in Fig. 1. The elementary factors are now $\eta\zeta\xi\eta^*\zeta^*\xi^*$. The unwanted expression $\eta\eta^* + \zeta\zeta^* + \xi\xi^*$ implies a single redundant combination which may be taken to be $\zeta\zeta^*$. Thus the states[12] are defined by $\eta^a\eta^{*a'}\xi^b\xi^{*b'}\zeta^c\zeta^{*c'}$, with $c = 0$ or $c' = 0$; the IR labels are $\lambda_1 = a + b + c$ and $\lambda_2 = a' + b' + c'$.

[12] The states defined in the text are analogous to the "Weyl" states described by Baird and Biedenharn in Ref. 10.

Case 11: $G \supset G$; the trivial case in which the subgroup is the group itself.

Here the terms IR and multiplet are synonymous. The elementary factors are just the heaviest states of the fundamental IR's. That the heaviest state of each multiplet (IR) is a product of powers of the elementary factors is just the content of Cartan's fundamental theorem. There are no redundant combinations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Boson Formalism in Superconductivity

L. LEPLAE AND H. UMEZAWA
Department of Physics, University of Wisconsin, Milwaukee, Wisconsin

(Received 17 January 1969)

After a brief summary of the results obtained previously concerning the boson formalism in superconductivity, the formalism is generalized to include finite temperature and Coulomb interactions. Finally, as an illustration, the formalism is used to derive the Landau–Ginsburg equations and to study the vortices in type II superconductors.

1. INTRODUCTION

The results of previous articles[1,2] suggest that most of the characteristics of the ground state of superconductors are controlled by bosons which can be mainly regarded as bound states of quasi-electrons. The purpose of this paper is to generalize in several directions the formalism developed previously and to extend it.

After briefly summarizing the results already obtained and showing how the boson formalism is particularly well suited to the description of the problems of gauge invariance, we first indicate how to generalize this formalism to all temperatures below T_c, how to include the effect of an external electromagnetic field, and finally how to take into account the Coulomb interaction among electrons.

In this last generalization we find that there appear two types of plasmalike oscillations. One type of oscillation is similar to the one appearing in a normal electron gas, whereas the other type is actually a phase oscillation which recovers the gauge invariance. The characteristics of these two modes are not necessarily the same under every circumstances. The possibility of distinguishing experimentally these two modes will be the object of an article to come.

We also find that the Coulomb effect does not introduce any major change in our previous results. Furthermore, the treatment of the Coulomb effect demonstrates once more the simplicity of the formalism.

After having introduced these generalizations, we proceed to show that in this context the generalized Landau–Ginsburg[3–5] equations can be derived very simply.

Finally, we show that the importance of the bosons becomes most explicit in the formation of vortices. We will see that the radius of the vortex core is a macroscopic manifestation of the uncertainty principle, thus supplying us with another macroscopic example of the quantum mechanical nature of superconductivity.

Let us close this section by a remark on notations. In Sec. 3, a superscript © is used to designate the operators which depend on the Coulomb interaction.

[1] L. Leplae and H. Umezawa, Nuovo Cimento **44**, 410 (1966).
[2] L. Leplae, R. N. Sen, and H. Umezawa, Progr. Theoret. Phys. (Kyoto) Suppl. (1965); L. Leplae, R. N. Sen, and H. Umezawa, Nuovo Cimento **49**, 1 (1967).
[3] L. P. Gor'kov, Zh. Eksp. Teor. Fiz. **36**, 1918 (1959) [Sov. Phys.—JETP **9**, 1364 (1959)].
[4] N. R. Werthamer, Phys. Rev. **132**, 663 (1963).
[5] L. Tewardt, Phys. Rev. **132**, 595 (1963).

The elementary factors are all the states of all the fundamental IR's. In general, certain combinations are redundant because of unwanted expressions. As an example, we consider the labeling of $SU(3)$ states. The fundamental IR's are shown in Fig. 1. The elementary factors are now $\eta\zeta\xi\eta^*\zeta^*\xi^*$. The unwanted expression $\eta\eta^* + \zeta\zeta^* + \xi\xi^*$ implies a single redundant combination which may be taken to be $\zeta\zeta^*$. Thus the states[12] are defined by $\eta^a\eta^{*a'}\xi^b\xi^{*b'}\zeta^c\zeta^{*c'}$, with $c = 0$ or $c' = 0$; the IR labels are $\lambda_1 = a + b + c$ and $\lambda_2 = a' + b' + c'$.

[12] The states defined in the text are analogous to the "Weyl" states described by Baird and Biedenharn in Ref. 10.

Case 11: $G \supset G$; the trivial case in which the subgroup is the group itself.

Here the terms IR and multiplet are synonymous. The elementary factors are just the heaviest states of the fundamental IR's. That the heaviest state of each multiplet (IR) is a product of powers of the elementary factors is just the content of Cartan's fundamental theorem. There are no redundant combinations.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Boson Formalism in Superconductivity

L. LEPLAE AND H. UMEZAWA
Department of Physics, University of Wisconsin, Milwaukee, Wisconsin

(Received 17 January 1969)

After a brief summary of the results obtained previously concerning the boson formalism in superconductivity, the formalism is generalized to include finite temperature and Coulomb interactions. Finally, as an illustration, the formalism is used to derive the Landau–Ginsburg equations and to study the vortices in type II superconductors.

1. INTRODUCTION

The results of previous articles[1,2] suggest that most of the characteristics of the ground state of superconductors are controlled by bosons which can be mainly regarded as bound states of quasi-electrons. The purpose of this paper is to generalize in several directions the formalism developed previously and to extend it.

After briefly summarizing the results already obtained and showing how the boson formalism is particularly well suited to the description of the problems of gauge invariance, we first indicate how to generalize this formalism to all temperatures below T_c, how to include the effect of an external electromagnetic field, and finally how to take into account the Coulomb interaction among electrons.

In this last generalization we find that there appear two types of plasmalike oscillations. One type of oscillation is similar to the one appearing in a normal electron gas, whereas the other type is actually a phase oscillation which recovers the gauge invariance. The characteristics of these two modes are not necessarily the same under every circumstances. The possibility of distinguishing experimentally these two modes will be the object of an article to come.

We also find that the Coulomb effect does not introduce any major change in our previous results. Furthermore, the treatment of the Coulomb effect demonstrates once more the simplicity of the formalism.

After having introduced these generalizations, we proceed to show that in this context the generalized Landau–Ginsburg[3–5] equations can be derived very simply.

Finally, we show that the importance of the bosons becomes most explicit in the formation of vortices. We will see that the radius of the vortex core is a macroscopic manifestation of the uncertainty principle, thus supplying us with another macroscopic example of the quantum mechanical nature of superconductivity.

Let us close this section by a remark on notations. In Sec. 3, a superscript © is used to designate the operators which depend on the Coulomb interaction.

[1] L. Leplae and H. Umezawa, Nuovo Cimento **44**, 410 (1966).
[2] L. Leplae, R. N. Sen, and H. Umezawa, Progr. Theoret. Phys. (Kyoto) Suppl. (1965); L. Leplae, R. N. Sen, and H. Umezawa, Nuovo Cimento **49**, 1 (1967).
[3] L. P. Gor'kov, Zh. Eksp. Teor. Fiz. **36**, 1918 (1959) [Sov. Phys.—JETP **9**, 1364 (1959)].
[4] N. R. Werthamer, Phys. Rev. **132**, 663 (1963).
[5] L. Tewardt, Phys. Rev. **132**, 595 (1963).

In the sections following Sec. 3, this is not done because all the arguments are true whether the Coulomb interaction exists or not.

2. SUMMARY OF PREVIOUS RESULTS

To begin with, let us summarize briefly the results of our previous articles.

The Hamiltonian used here is the familiar one

$$H = \sum \epsilon_k (a_{k\uparrow}^+ a_{k\uparrow} + a_{k\downarrow}^+ a_{k\downarrow}) + (-\lambda) \sum a_{k\uparrow}^+ a_{-k+q\downarrow}^+ a_{-k'+q\downarrow} a_{k'\uparrow}. \quad (1)$$

We looked for a set of creation and annihilation operators of free fermions ($\alpha_{k\downarrow,\uparrow}^+(t)$, $\alpha_{k\downarrow,\uparrow}(t)$) and free bosons ($B_k^+(t)$, $B_k(t)$) which are determined by the conditions that their time dependence has the form

$$\alpha_{k\downarrow,\uparrow}(t) = \alpha_{k\downarrow,\uparrow} e^{-iE_k t}, \qquad B_k(t) = B_k e^{-i\omega_k t} \quad (2)$$

and that the electron operators $a_{k\downarrow,\uparrow}(t)$ are expressed as a linear combination of normal products of these free operators

$$a_{k\uparrow}(t) = \bar{\alpha}_{k\uparrow}(t) \cos\theta_k - \bar{\alpha}_{-k\downarrow}^+(t) \sin\theta_k, \quad (3a)$$

etc., where

$$\bar{\alpha}_{k\uparrow}(t) = \alpha_{k\uparrow}(t) + \sum_{p,l} f_k(p, \mathbf{l}) \alpha_{-k+l\downarrow}^+(t) \alpha_{p+\frac{1}{2}l\uparrow}(t) \alpha_{-p+\frac{1}{2}l\uparrow}(t) + \sum_{p,l} h_k(p, \mathbf{l}) \alpha_{-k+l\downarrow}^+(t) \alpha_{p-\frac{1}{2}l\downarrow}^+(t) \alpha_{p-\frac{1}{2}l\uparrow}(t) + \sum_l g_k^{(1)}(\mathbf{l}) \alpha_{-k+l\downarrow}^+(t) B_l(t) + \sum_l g_k^{(2)}(\mathbf{l}) \alpha_{-k+l\downarrow}^+(t) B_{-l}^+(t) + \cdots, \quad (3b)$$

etc. Here, the intermediate operators $\bar{\alpha}(t)$ were introduced to simplify the formula. Note that the quasifermion operators (α^+, α) and boson operators (B^+, B) commute with each other. The quantities θ, f, h, $g^{(1)}$, $g^{(2)}$, E_k, and ω_l are so determined that $a(t)$ satisfies the canonical equations derived from the Hamiltonian (1). *It should be noted that the existence of the Bose field B_k is not assumed, but is required by the fact that $a(t)$ has to satisfy the above-mentioned equations as is extensively shown in Ref.* 1. There it is also shown that B is a bound state.

It can be shown that the Hamiltonian given in (1) takes the form

$$H = \sum E_k (\alpha_{k\uparrow}^+ \alpha_{k\uparrow} + \alpha_{k\downarrow}^+ \alpha_{k\downarrow}) + \sum \omega_l B_l^+ B_l, \quad (4)$$

when expressed in terms of the free operators.

The calculation of θ, f, h, $g^{(1)}$, $g^{(2)}$, E_k, and ω_l are presented in Ref. 1. In particular, ω_l is found to be proportional to l:

$$\omega_l = v_0 |\mathbf{l}|. \quad (5)$$

Using the expressions in Eqs. (3), we express the charge and current densities

$$\rho(\mathbf{x}, t) = \psi_\uparrow^+ \psi_\uparrow + \psi_\downarrow^+ \psi_\downarrow, \quad (6)$$

$$\mathbf{j}(\mathbf{x}, t) = (2m)^{-1} [\psi_\uparrow^+ \nabla \psi_\uparrow - \nabla \psi_\uparrow^+ \cdot \psi_\uparrow + \psi_\downarrow^+ \nabla \psi_\downarrow - \nabla \psi_\downarrow^+ \cdot \psi_\downarrow] \quad (7)$$

explicitly in terms of free operators ($\alpha_{k\downarrow,\uparrow}$, B_k), where

$$\psi_{\uparrow,\downarrow} = \sum_k a_{k\uparrow,\downarrow}(t) e^{i\mathbf{k}\cdot\mathbf{x}}.$$

The results are the following[1]:

$$\rho(\mathbf{x}, t) = \rho^{(1)}(\mathbf{x}, t) + \rho^{(2)}(\mathbf{x}, t), \quad (8)$$

$$\mathbf{j}(\mathbf{x}, t) = \mathbf{j}^{(1)}(\mathbf{x}, t) + \mathbf{j}^{(2)}(\mathbf{x}, t), \quad (9)$$

$$\rho^{(2)}(\mathbf{x}, t) = -\eta \frac{\partial}{\partial t} B(\mathbf{x}, t), \quad (10)$$

$$\mathbf{j}^{(2)}(\mathbf{x}, t) = v_0^2 \eta \nabla B(\mathbf{x}, t). \quad (11)$$

Here the boson operator $B(\mathbf{x}, t)$ is defined as

$$B(\mathbf{x}, t) = \sum_{|\mathbf{l}| < l_0} (2\omega_l)^{-\frac{1}{2}} (B_l e^{i(\mathbf{l}\cdot\mathbf{x} - \omega_l t)} + B_l^+ e^{-i(\mathbf{l}\cdot\mathbf{x} - \omega_l t)}). \quad (12)$$

The cut off momentum l_0 is given by

$$l_0 = 2\Delta / v_0, \quad (13)$$

because the bosons decay into quasifermion pairs when $v_0 l > 2\Delta$.

The operator $B(\mathbf{x}, t)$ satisfies the equation

$$\left(\frac{\partial^2}{\partial t^2} - v_0^2 \nabla^2\right) B(\mathbf{x}, t) = 0. \quad (14)$$

It can be seen from (10) and (11) that $\mathbf{j}^{(2)}$ and $\rho^{(2)}$ satisfy the conservation law

$$\nabla \cdot \mathbf{j}^{(2)} + \frac{\partial}{\partial t} \rho^{(2)} = 0.$$

The constant η in (10) and (11) is given by[6]

$$\eta = -2^{\frac{3}{2}} \Delta [R(0)]^{\frac{1}{2}}, \quad (15)$$

where Δ is the energy gap [i.e., $E_k = (\epsilon_k^2 + \Delta^2)^{\frac{1}{2}}$] and R is defined by[7,8]

$$R(\mathbf{l}) = -\frac{i}{2} \int \frac{d^3k}{(2\pi)^3} \times \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1}{(\omega^2 - E_{k+\frac{1}{2}l}^2 + i\epsilon)(\omega^2 - E_{k-\frac{1}{2}l}^2 + i\epsilon)} = \frac{1}{4} \int \frac{d^3k}{(2\pi)^3} \frac{(E_{k+\frac{1}{2}l} + E_{k-\frac{1}{2}l})}{(E_{k+\frac{1}{2}l} E_{k-\frac{1}{2}l})} \frac{1}{(E_{k+\frac{1}{2}l} + E_{k-\frac{1}{2}l})^2}. \quad (16a)$$

[6] When $\hbar\omega_c / E_F \ll 1$ and $\Delta^2 / \hbar^2 \omega_c^2 \ll 1$, one can show easily that $\eta v_0 \simeq -(n/m)^{\frac{1}{2}}$, where n is the electron density.

[7] $R(\mathbf{l})$ defined here corresponds to $R(\mathbf{l}, 0)$ in Ref. 1.

[8] There are two misprints in Ref. 1: Eq. (4.3) in Ref. 1 should be replaced by (15) of this article.

The boson velocity v_0 is defined by the relation

$$v_0^2 l^2 R(\mathbf{l}) = -\frac{i}{2}\int\frac{d^3k}{(2\pi)^3} \times\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{(\epsilon_{k+\frac{1}{2}l}-\epsilon_{k-\frac{1}{2}l})^2}{(\omega^2-E_{k+\frac{1}{2}l}^2+i\epsilon)(\omega^2-E_{k-\frac{1}{2}l}^2+i\epsilon)} = \frac{1}{4}\int\frac{d^3k}{(2\pi)^3}\frac{(E_{k+\frac{1}{2}l}+E_{k-\frac{1}{2}l})}{E_{k+\frac{1}{2}l}E_{k-\frac{1}{2}l}}\frac{(\epsilon_{k+\frac{1}{2}l}-\epsilon_{k-\frac{1}{2}l})^2}{(E_{k+\frac{1}{2}l}+E_{k-\frac{1}{2}l})^2}. \tag{16b}$$

This gives $v_0 \simeq (3)^{-\frac{1}{2}}V_F$ when the approximation $\mathbf{k}\simeq\mathbf{k}_F$ is made in the above integration. (This approximate form of v_0 is the one given by Bogoliubov[9] for the collective modes.) The quantities $\rho^{(1)}$ and $\mathbf{j}^{(1)}$ in (10) and (11) mainly depend on the quasi-electron operators (α, α^+) and satisfy the conservation law

$$\nabla\cdot\mathbf{j}^{(1)}+\frac{\partial}{\partial t}\rho^{(1)}=0.$$

It was proved[1] that

$$\int d^3x\rho^{(1)}(\mathbf{x},t)=0, \tag{17}$$

which shows that the Fourier component $\rho^{(1)}(\mathbf{l}, E)$ vanishes at the low-momentum limit ($\mathbf{l}\simeq\mathbf{0}$). To see more closely the situation, the reader is referred to the relations (3.4) and (3.7) in Ref. 1, which show that $\rho^{(1)}(\mathbf{l}, E)$ is small when $v_0|\mathbf{l}|\ll 2\Delta$. Therefore, we find that

$$\rho(\mathbf{l},E)\simeq\rho^{(2)}(\mathbf{l},E),\quad\text{for}\quad v_0|\mathbf{l}|\ll 2\Delta. \tag{18}$$

Let us now discuss the gauge transformation. The transformation

$$\psi(\mathbf{x},t)\to\exp[if(\mathbf{x},t)]\psi(\mathbf{x},t)$$

is generated by

$$N_f=\int d^3x f(\mathbf{x},t)\rho(\mathbf{x},t)$$

which can be rewritten as

$$N_f=\int d^3x f(\mathbf{x},t)\rho^{(2)}(\mathbf{x},t) = -\eta\int d^3x f(\mathbf{x},t)\frac{\partial}{\partial t}B(\mathbf{x},t), \tag{19}$$

when $f(\mathbf{x}, t)$ is a slowly varying function of $\mathbf{x}$ so that only low momenta are important. In this case, it can be seen from (19) *that the gauge transformation is induced solely by the boson.*

[9] N. N. Bogoliubov, *A New Method in the Theory of Superconductivity* (Consultants Bureau, New York, 1959).

To express the gauge transformation explicitly in terms of free operators, we first compute the commutator

$$\left[B(\mathbf{x},t),\frac{\partial}{\partial t}B(\mathbf{y},t)\right]=ic(\mathbf{x}-\mathbf{y}). \tag{20}$$

Since (20) is close to the canonical commutation relation, we introduce the following notation:

$$\pi(\mathbf{x},t)=\frac{\partial}{\partial t}B(\mathbf{x},t). \tag{21}$$

Then the approximation (18) gives

$$\rho(\mathbf{x},t)=-\eta\pi(\mathbf{x},t). \tag{22}$$

Making use of (12), we find that

$$c(\mathbf{x}-\mathbf{y})=\int_0^{2\pi}d\varphi\int_0^{\pi}\sin\theta\,d\theta\int_0^{l_0}l^2\,dl e^{il|\mathbf{x}-\mathbf{y}|\cos\theta} \tag{23}$$

which has the following normalization:

$$\int d^3y c(\mathbf{x}-\mathbf{y})=1. \tag{24}$$

We are now ready to compute the commutator between the generator and the boson operator,

$$[N_f,\eta^{-1}B(\mathbf{x},t)]=i\int d^3y c(\mathbf{x}-\mathbf{y})f(\mathbf{y},t), \tag{25}$$

where (19) was used. From this, the gauge transformation is found to be

$$\varphi_{\uparrow,\downarrow}(\mathbf{x},t)\to\exp[-iN_f]\varphi_{\uparrow,\downarrow}(\mathbf{x},t)\exp[iN_f]=\varphi_{\uparrow,\downarrow}(\mathbf{x},t),$$
$$B(\mathbf{x},t)\to\exp[-iN_f]B(\mathbf{x},t)\exp[iN_f] = B(\mathbf{x},t)+\eta\int d^3y c(\mathbf{x}-y)f(\mathbf{y},t). \tag{26}$$

Here the quasifermion field $\varphi(\mathbf{x}, t)$ is defined as

$$\varphi_{\uparrow,\downarrow}(\mathbf{x},t)=\int d^3k\alpha_{k\uparrow,\downarrow}e^{i(\mathbf{k}\cdot\mathbf{x}-E_k t)}.$$

In the special case of $f(\mathbf{x}, t) = 1$, Eq. (25) becomes

$$[N,\eta^{-1}B(\mathbf{x},t)]=i. \tag{27}$$

This shows that $\eta^{-1}B$ is the canonical conjugate of the number operator of electrons. Owing to this reason, B was called the phason field. When $f(\mathbf{x}, t)$ is a constant, say θ, the gauge transformation (26) is

$$\varphi\to\varphi,\quad B\to B+\eta\theta. \tag{28}$$

This shows that ψ takes the form

$$\psi_{\uparrow,\downarrow}=e^{iB/\eta}F(\varphi,\nabla B,\pi) \tag{29}$$

when expressed in terms of the free fields B and φ. Obviously, the phason equation (14) is invariant under the gauge transformation (28) as it should be.

This last expression shows explicitly that the gauge transformation is controlled mainly by the phasons when $f(\mathbf{x}, t)$ is a slowly varying function of $\mathbf{x}$. On the other hand it can be shown that, when f varies strongly so that high momenta dominate, the gauge transformation is essentially controlled by the quasi-fermions.

We close this section by noting that the gauge transformation (26) modifies the ground state current by the amount

$$\langle \delta j \rangle = v_0^2 \eta^2 \int d^3y c(\mathbf{x} - \mathbf{y}) \nabla f(\mathbf{y}, t), \tag{30}$$

where use was made of (11).

3. COULOMB EFFECTS

In the previous section we completely disregarded the Coulomb interaction. As the following considerations show, it is a simple matter to take into account the Coulomb effects in this formalism.

Let us first recall that $\varphi(\mathbf{x}, t)$ and $B(\mathbf{x}, t)$, introduced in the previous section, are the free operators and, therefore, that the Hamiltonian (1) can be expressed in the form of a free Hamiltonian of these operators [cf. (4)]:

$$H = H_0(B) + H_0(\varphi), \tag{31}$$

where

$$H_0(\varphi) = \sum_k E_k(\alpha_{k\uparrow}^+ \alpha_{k\uparrow} + \alpha_{k\downarrow}^+ \alpha_{k\downarrow}) \tag{32}$$

and

$$\begin{aligned} H_0(B) &= \sum_{l<l_0} \omega_l B_l^+ B_l \\ &= \tfrac{1}{2} \int d^3x [\pi^2 + v_0^2 \nabla B \cdot \nabla B] + c \text{ number}. \end{aligned} \tag{33}$$

Here π is defined by (21), i.e.,

$$\pi = \frac{\partial}{\partial t} B. \tag{34}$$

When the Coulomb interaction is introduced, the new electron field $\psi^{\mathbb{C}}$ satisfies the canonical equation

$$\frac{\partial}{\partial t} \psi^{\mathbb{C}}(\mathbf{x}, t) = -i[\psi^{\mathbb{C}}(\mathbf{x}, t), \bar{H}^{\mathbb{C}}]. \tag{35}$$

Here the new Hamiltonian $\bar{H}^{\mathbb{C}}$ has the form

$$\bar{H}^{\mathbb{C}} = H^{\mathbb{C}} + H_{(c)}^{\mathbb{C}} \tag{36}$$

in which the Coulomb interaction $H_{(c)}^{\mathbb{C}}$ is given by

$$H_{(c)}^{\mathbb{C}} = \frac{e^2}{2} \int d^3x \int d^3y \frac{\rho^{\mathbb{C}}(\mathbf{x}, t) \rho^{\mathbb{C}}(\mathbf{y}, t)}{|\mathbf{x} - \mathbf{y}|} \tag{37}$$

and $H^{\mathbb{C}}$ is defined by replacing ψ by $\psi^{\mathbb{C}}$ in the old Hamiltonian (1):

$$H^{\mathbb{C}} = H_{\psi \to \psi^{\mathbb{C}}}. \tag{38}$$

The "neutral" electron field ψ is related to $\psi^{\mathbb{C}}$ by a unitary transformation $S(t)$:

$$\psi^{\mathbb{C}}(\mathbf{x}, t) = S^{-1}(t) \psi(\mathbf{x}, t) S(t). \tag{39}$$

Let us define

$$\varphi^{\mathbb{C}}(\mathbf{x}, t) = S^{-1}(t) \varphi(\mathbf{x}, t) S(t), \tag{40}$$

$$B^{\mathbb{C}}(\mathbf{x}, t) = S^{-1}(t) B(\mathbf{x}, t) S(t), \tag{41}$$

$$\begin{aligned} \pi^{\mathbb{C}}(\mathbf{x}, t) &= S^{-1}(t) \pi(\mathbf{x}, t) S(t) \\ &= S^{-1}(t) \frac{\partial}{\partial t} B(\mathbf{x}, t) S(t), \end{aligned} \tag{42}$$

where φ and B are the operators introduced in the last section.

We see from (39) that

$$\begin{aligned} \rho^{\mathbb{C}}(x, t) &= S^{-1}(t) \rho(\mathbf{x}, t) S(t), \\ \mathbf{j}^{\mathbb{C}}(x, t) &= S^{-1}(t) \mathbf{j}(\mathbf{x}, t) S(t), \end{aligned} \tag{43}$$

and that

$$\bar{H}^{\mathbb{C}} = S^{-1}(t)(H + H_{(c)}) S(t), \tag{44}$$

in which

$$H_{(c)} = \frac{e^2}{2} \int d^3x \int d^3y \frac{\rho(\mathbf{x}, t) \rho(\mathbf{y}, t)}{|\mathbf{x} - \mathbf{y}|}. \tag{45}$$

Making use of (31), we can rewrite (44) as

$$\bar{H}^{\mathbb{C}} = S^{-1}(t)[H_0(\varphi) + H_0(B) + H_{(c)}] S(t). \tag{46}$$

Since we are interested in the long-range effects of the Coulomb potential, we are not concerned with the high-momentum contributions of the Coulomb potential. To be specific, we keep only the momentum components of the Coulomb potential corresponding to $l < l_0$. This choice of cutoff will be justified later. We can now write ρ in (37) in terms of the boson operator as

$$\rho^{\mathbb{C}}(\mathbf{x}, t) = -\eta \pi^{\mathbb{C}}(\mathbf{x}, t), \tag{47}$$

so that we obtain

$$H_{(c)}^{\mathbb{C}} = \frac{\mu^2}{2(4\pi)} \int d^3x \int d^3y \frac{\pi^{\mathbb{C}}(\mathbf{x}, t) \pi^{\mathbb{C}}(\mathbf{y}, t)}{|\mathbf{x} - \mathbf{y}|}, \tag{48}$$

where

$$\mu = (4\pi)^{\frac{1}{2}} e \eta. \tag{49}$$

It is worthwhile to note that, as $H_{(c)}^{\mathbb{C}}$ in (48) contains only the boson operators, the quasifermion operators are not modified; i.e.,

$$\varphi^{\mathbb{C}} = \varphi, \quad S^{-1}(t) H_0(\varphi) S(t) = H_0(\varphi). \tag{50}$$

Because of this, only the bosons need to be studied.

The boson Hamiltonian is

$$H^{\mathbb{C}(B)} = H_0^{\mathbb{C}}(B) + H_{(c)}^{\mathbb{C}}, \tag{51}$$

with

$$H_0^{\mathbb{C}}(B) = S^{-1}(t)H_0(B)S(t)$$
$$= \tfrac{1}{2}\int d^3x[(\pi^{\mathbb{C}})^2 + v_0^2\nabla B^{\mathbb{C}}\cdot\nabla B^{\mathbb{C}}] + c\text{ number}. \tag{52}$$

To derive the new boson equation, we note the following relations:

$$[B^{\mathbb{C}}(\mathbf{x}, t), \pi^{\mathbb{C}}(\mathbf{y}, t)] = ic(\mathbf{x}-\mathbf{y}), \tag{53}$$

$$\int d^3y c(\mathbf{x}-\mathbf{y})B^{\mathbb{C}}(\mathbf{y}, t) = B^{\mathbb{C}}(\mathbf{x}, t),$$

$$\int d^3y c(\mathbf{x}-\mathbf{y})\pi^{\mathbb{C}}(\mathbf{y}, t) = \pi^{\mathbb{C}}(\mathbf{x}, t). \tag{54}$$

The above commutation relation is due to (20) and the relations in (54) can be derived by means of (23).

The canonical equation

$$\frac{\partial}{\partial t}B^{\mathbb{C}}(\mathbf{x}, t) = -i[B^{\mathbb{C}}(\mathbf{x}, t), H^{\mathbb{C}(B)}]$$

leads to

$$\frac{\partial}{\partial t}B^{\mathbb{C}}(\mathbf{x}, t) = \pi^{\mathbb{C}}(\mathbf{x}, t) + \frac{\mu^2}{4\pi}\int d^3y\,\frac{\pi^{\mathbb{C}}(\mathbf{y}, t)}{|\mathbf{x}-\mathbf{y}|}. \tag{55}$$

Then the canonical equation for $\partial B/\partial t$ gives

$$\left(\frac{\partial^2}{\partial t^2} - v_0^2\nabla^2\right)B^{\mathbb{C}}(\mathbf{x}, t)$$
$$- \frac{\mu^2 v_0^2}{4\pi}\int d^3y\,\frac{1}{|\mathbf{x}-\mathbf{y}|}\nabla^2 B^{\mathbb{C}}(\mathbf{y}, t) = 0. \tag{56}$$

The last term on the left-hand side shows that the bosons at two points $\mathbf{x}$ and $\mathbf{y}$ are correlated with each other even when $\mathbf{x}$ is very far from $\mathbf{y}$.

It should be noted that partial integration applied to the integral in (56) does not give $(-\mu^2 v_0^2 B^{\mathbb{C}})$ because of the long-range nature of the Coulomb potential. To estimate this integral, we must remember that the integration domain is the metal of *finite* volume $V = L^3$. This means the d^3y in (56) must be replaced by $g(\mathbf{y})\,d^3y$ where $g(\mathbf{y})$ is a function confined inside the metal. We can approximately express $g(\mathbf{y})$ by $\exp(-\epsilon\,|\mathbf{x}-\mathbf{y}|)$ when V is large. Here ϵ is given by

$$\epsilon \simeq 1/L. \tag{57}$$

The last term in (56) then gives

$$[\mu^2 v_0^2/(\nabla^2 - \epsilon^2)]\nabla^2 B^{\mathbb{C}}.$$

Therefore, (56) becomes

$$\left(\frac{\partial^2}{\partial t^2} + v_0^2\mu^2 - v_0^2\nabla^2 - \frac{\mu^2 v_0^2\epsilon^2}{\epsilon^2 - \nabla^2}\right)B^{\mathbb{C}}(\mathbf{x}, t) = 0. \tag{58}$$

Thus, the energy spectrum of the boson of momentum l is given by

$$\tilde{\omega}_l = [v_0^2\mu^2 + v_0^2 l^2 - \mu^2 v_0^2\epsilon^2/(\epsilon^2 + l^2)]^{\frac{1}{2}}$$
$$= (v_0^2\mu^2 + v_0^2 l^2)^{\frac{1}{2}}, \quad \text{for} \quad l \gg \epsilon,$$
$$= v_0 l, \qquad\qquad \text{for} \quad l \ll \epsilon. \tag{59}$$

One should note that *the excited energy spectrum practically starts at the plasma frequency* $v_0\mu$, because $l \gtrsim \epsilon$ for excited levels. The bosons of extremely low momentum ($l \ll \epsilon$) can, however, influence the ground state and, as is shown later, play an important role in the gauge transformation. The transformation $S(t)$ can be computed in many ways. An easy method[10] is to diagonalize the Hamiltonian $H^{(B)}$ by means of the Bogoliubov transformation. To do this, we express the Hamiltonian $H_0(B)$ and $H_{(c)}(t)$ in the momentum representation by using (12) and a similar expression for $\pi = \partial B/\partial t$:

$$H_0(B) + H_{(c)}(t)$$
$$= \left[\sum_k \omega_k B_k^+ B_k - \frac{\mu^2\omega_k}{k^2+\epsilon^2}(B_k e^{-i\omega_k t} - B_{-k}^+ e^{i\omega_k t})\right.$$
$$\left.\times\,(B_{-k}e^{-i\omega_k t} - B_k^+ e^{i\omega_k t})\right].$$

Here ϵ is introduced to take into account the fact that the size of the system is not really infinite [see Eq. (57)]. It can be shown that $H^{(B)}$ in (27) takes the diagonal form

$$H^{\mathbb{C}(B)} = \sum_k \tilde{\omega}_k B_k^+ B_k + c\text{ number}, \tag{60}$$

where we choose $S(t)$ to satisfy

$$S^{-1}(t)B_l S(t)e^{-i\omega_l t}$$
$$= \tfrac{1}{2}(l v_0\tilde{\omega}_l)^{-\frac{1}{2}}[(\tilde{\omega}_l + v_0 l)B_l e^{-i\tilde{\omega}_l t} + (\tilde{\omega}_l - v_0 l)B_{-l}^+ e^{i\tilde{\omega}_l t}]. \tag{61}$$

Here, the energy $\tilde{\omega}_l$ is given in (59).

Then, using (12), we can express $B^{\mathbb{C}}$ and $\pi^{\mathbb{C}}$ in the momentum representation:

$$B^{\mathbb{C}}(\mathbf{x}, t) = \sum_{l<l_0} v_0 l^{-1}(\tfrac{1}{2}\tilde{\omega}_l)^{\frac{1}{2}}$$
$$\times\,(B_l e^{i\mathbf{l}\cdot\mathbf{x} - i\tilde{\omega}_l t} + B_l^+ e^{-i\mathbf{l}\cdot\mathbf{x} + i\tilde{\omega}_l t}), \tag{62}$$

$$\pi^{\mathbb{C}}(\mathbf{x}, t) = -i\sum_{l<l_0} v_0 l/(2\tilde{\omega}_l)^{-\frac{1}{2}}$$
$$\times\,(B_l e^{i\mathbf{l}\cdot\mathbf{x} - i\tilde{\omega}_l t} - B_l^+ e^{-i\mathbf{l}\cdot\mathbf{x} + i\tilde{\omega}_l t}). \tag{63}$$

To see if the boson $B^{\mathbb{C}}$ still acts as a phason, we shall examine the gauge transformation. Following an argument similar to the one presented in Sec. 2, it can be seen from (47) and (53) that *the gauge transformation is induced by the boson operator* in the

[10] Another way, where the momentum representation is not needed, is to use the equation $i[dS(t)/dt] = H_{(c)}(t)S(t)$.

following manner:

$$B^{©}(\mathbf{x}, t) \to B^{©}(\mathbf{x}, t) + \eta \int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}), \quad (64)$$

even when the Coulomb potential exists. We thus see that $B^{©}$ is the phason and that the electron operators keep the form (29), i.e.,

$$\psi_{\uparrow,\downarrow} = \exp(iB^{©}/\eta) F(\varphi, \nabla B^{©}, \pi^{©}). \quad (65)$$

Let us further note that (27) remains true even when B is replaced by $B^{©}$, showing that $(1/\eta)B^{©}$ is the canonical conjugate of total number of electrons.

It is worth noting that the phason equations (56) and (58) are invariant under the gauge transformation with a uniform phase $f(\mathbf{y}) = \theta$ (i.e., $B^{©} \to B^{©} + \eta\theta$), as they should be. This transformation is induced by the phasons with $l \ll \epsilon$, showing the important role played by these phasons.

Besides the phasons described above, there may exist other excitations of the plasma type with $l > l_0$ corresponding to the Coulomb interactions among quasi-electrons (or "normal" electrons). Nevertheless, these plasmons will not play any role in the phase properties of the superconductor and, in this sense, are not of immediate interest in this article. This justifies the cutoff $l = l_0$ introduced in the Coulomb potential.

To see how the Coulomb effect influences the density, we rewrite (55) in the form

$$\pi^{©}(\mathbf{x}, t) = -\frac{1}{4\pi} \int d^3y \frac{e^{-\mu|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} \nabla^2 \frac{\partial}{\partial t} B^{©}(\mathbf{y}, t). \quad (66)$$

Performing a partial integration, we find

$$\pi^{©}(\mathbf{x}, t) = \frac{\partial}{\partial t} B^{©}(\mathbf{x}, t) - \frac{\mu^2}{4\pi} \int d^3y \frac{e^{-\mu|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} \frac{\partial}{\partial t} B^{©}(\mathbf{y}, t), \quad (67)$$

which gives

$$\rho^{©}(\mathbf{x}, t) = -\eta \frac{\partial}{\partial t} B^{©}(\mathbf{x}, t) + \frac{\eta\mu^2}{4\pi} \int d^3y \frac{e^{-\mu|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} \frac{\partial}{\partial t} B^{©}(\mathbf{y}, t). \quad (68)$$

The second term in (68) shows that the charge density at a point $\mathbf{x}$ is influenced by the phasons in the domain of radius $1/\mu$ [$= 1/(4\pi)^{\frac{1}{2}}e\eta$].

The current can be derived from (43) and (11):

$$\mathbf{j}^{©}(x, t) = v_0^2 \eta \nabla B^{©}(\mathbf{x}, t). \quad (69)$$

To prove the conservation law, we note that the phason equation (58) implies

$$\left(\frac{\partial^2}{\partial t^2} - v_0^2\nabla^2 + \mu^2 v_0^2\right) \nabla^2 B^{©}(\mathbf{x}, t) = 0. \quad (70)$$

This relation together with (66) gives

$$\frac{\partial}{\partial t} \pi^{©}(\mathbf{x}, t) = v_0^2 \nabla^2 B^{©}(\mathbf{x}, t). \quad (71)$$

The conservation law

$$\nabla \cdot \mathbf{j}^{©} + \frac{\partial}{\partial t} \rho^{©} = 0$$

follows immediately from (71) because of (47) and (69).

4. INVARIANT TRANSFORMATION[11]

Let us now consider the special case of gauge transformations which leave the system invariant. We thus assume that $f(\mathbf{x}, t)$ is independent of t, denoting it by $f(\mathbf{x})$, and that

$$[N_f, H] = 0. \quad (72)$$

The latter condition leads to

$$\int d^3x f(\mathbf{x}) \frac{\partial}{\partial t} \rho(\mathbf{x}, t) = 0,$$

which in turn gives

$$\int d^3x f(\mathbf{x}) \nabla \cdot \mathbf{j}(\mathbf{x}, t) = 0.$$

This is satisfied when $\nabla f = 0$. The invariance condition, however, is still weakened when $f(\mathbf{x})$ is a slowly varying function ($l \ll l_0$). In that case,

$$\int d^3x f(\mathbf{x}) \frac{\partial}{\partial t} \rho(\mathbf{x}, t) = -\int d^3x f(\mathbf{x}) \nabla \cdot \mathbf{j}(\mathbf{x}) = -v_0^2 \eta \int d^3x f(\mathbf{x}) \nabla^2 B(\mathbf{x}, t),$$

owing to Eq. (11) [or (69)]. The requirement that $\dot{N}_f = 0$ implies that

$$\left[B(\mathbf{x}), \int d^3y \nabla^2 B(\mathbf{y}, t) f(\mathbf{y})\right] = \nabla^2 \int d^3y i c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) = \nabla^2 f(\mathbf{x}) = 0,$$

using the fact that $f(\mathbf{x})$ is varying slowly with $\mathbf{x}$. Thus, the gauge transformation is an invariant transformation when

$$\nabla^2 f(\mathbf{x}) = 0. \quad (73)$$

It is obvious that the gauge transformation (26) leaves the phason equation (14) [or (56)] invariant under the condition (73), as it should.

A simple example of (73) is given by

$$f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x},$$

[11] In this and following sections, no © superscripts are used because all the arguments are true whether the Coulomb interaction exists or not.

where **a** is a constant vector. This corresponds to the following persistent current:

$$\langle \delta j \rangle = v_0^2 \eta^2 \mathbf{a},$$

according to (30).

The condition (73) is no longer necessary when there is an external electromagnetic vector potential **A**, since the effect of the gauge transformation of ψ can be cancelled out by the gauge transformation of **A** (i.e., $e\mathbf{A} \to e\mathbf{A} + \vec{\partial} f$), and this leaves the system invariant. In that case, (26) shows that **A** always appears in the following combination with ∇B:

$$\nabla B - \eta \int d^3y c(\mathbf{x} - \mathbf{y}) \mathbf{A}(\mathbf{y}). \tag{74}$$

By the same argument, we can show that the scalar potential ϕ always appears in the combination

$$\frac{\partial}{\partial t} B - \int d^3y c(\mathbf{x} - \mathbf{y}) \phi(\mathbf{y}). \tag{75}$$

5. FINITE TEMPERATURE

The arguments in the previous section can be easily extended to finite temperatures by means of the well-known quantum-statistical technique,[12,13] i.e., by making the following replacements in expressions of Eqs. (16):

$$\omega \to i\omega_n,$$
$$(2\pi)^{-1} \int_{-\infty}^{\infty} d\omega \to i k_B T \sum_n, \tag{76}$$

where

$$\omega_n = 2n\pi k_B T, \quad \text{for bosons,}$$
$$= (2n+1) k_B T, \quad \text{for fermions.}$$

Similar replacements should be performed everywhere in the computations of the previous article.[1] In particular, the expansion coefficients in Eqs. (3) also depend on the temperature.

By (76) we obtain

$$R^T(\mathbf{l}) = \frac{k_B T}{2} \int \frac{d^3k}{(2\pi)^3} \sum_n \frac{1}{(\omega_n^2 + E_{k+\frac{1}{2}l}^2)(\omega_n^2 + E_{k-\frac{1}{2}l}^2)} \tag{77a}$$

and

$$v_0^2 l^2 R^T(\mathbf{l}) = \frac{k_B T}{2} \int \frac{d^3k}{(2\pi)^3} \sum_n \frac{(\epsilon_{k+\frac{1}{2}l} - \epsilon_{k-\frac{1}{2}l})^2}{(\omega_n^2 + E_{k+\frac{1}{2}l}^2)(\omega_n^2 + E_{k-\frac{1}{2}l}^2)}$$
$$= \frac{k_B T}{2} l^2 \int \frac{d^3k}{(2\pi)^3} \frac{k^2}{3m^2} \sum_n \frac{1}{(\omega_n^2 + E_{k+\frac{1}{2}l}^2)(\omega_n^2 + E_{k-\frac{1}{2}l}^2)}. \tag{77b}$$

[12] H. Ezawa, Y. Tomozawa, and H. Umezawa, Nuovo Cimento 5, 810 (1957).

[13] A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshiskii, Zh. Eksp. Teor. Fiz. 36, 900 (1959) [Sov. Phys.—JETP 9, 636 (1959)].

The constant η depends on T through the relation (15), i.e., $\eta(T) = -2^{\frac{3}{2}} \Delta(T) [R^T(0)]^{\frac{1}{2}}$.

We thus find that

$$v_0^2 \eta_{(T)}^2 = 4\Delta_{(T)}^2 k_B T \int \frac{d^3k}{(2\pi)^3} \frac{k^2}{3m^2} \times \sum_n \frac{1}{(\omega_n^2 + E_{k+\frac{1}{2}l}^2)(\omega_n^2 + E_{k-\frac{1}{2}l})}. \tag{78}$$

It can easily be shown that $\eta^2(T) v_0^2 \simeq n_s(T)/m$, where $n_s(T)$ is the density of superelectrons. At finite temperature, Eq. (59) should be replaced by

$$\tilde{\omega}_l(T) = [v_0^2 \mu(T)^2 + v_0^2 l^2 - v_0^2 \mu^2(T) \epsilon^2 / (l^2 + \epsilon^2)]^{\frac{1}{2}},$$

where

$$v_0^2 \mu^2(T) = 4\pi e^2 \eta^2(T) v_0^2 \simeq 4\pi e^2 n_s(T)/m.$$

Besides these phase oscillations there can also exist, at finite temperature, plasma oscillations of the quasi-electrons.

6. THE GENERALIZED LANDAU–GINSBURG EQUATIONS

In this section we derive the second Landau–Ginsburg equation and compare the expression obtained with the one given by Werthamer.[4] As our purpose is essentially to show the simplicity of the derivation, it is not necessary to give a full derivation of the first Landau–Ginsburg equation. (The methods used in the derivation of both equations are quite similar.)

In order to simplify the notation, we replace the function $c(\mathbf{x} - \mathbf{x}')$ [see (23)] by $\delta(\mathbf{x} - \mathbf{x}')$ in this section. This simplification does not induce any significant change in the results. Thus, (26) shows us that a gauge transformation

$$\psi \to e^{if(\mathbf{x},t)} \psi$$

is induced by the transformation of $B(\mathbf{x}, t)$,

$$B(\mathbf{x}, t) \to B(\mathbf{x}, t) + \eta f(\mathbf{x}, t), \tag{79}$$

when $f(\mathbf{x}, t)$ is a slowly varying function of **x**.

Making the replacement (74) in the relation (30), we obtain the ground-state current[14]

$$\langle \mathbf{j} \rangle = v_0^2 \eta^2 (\nabla f - e\mathbf{A}). \tag{80}$$

[14] Combining (30) and (74), we find that

$$\langle \mathbf{j}(\mathbf{x}) \rangle = v_0^2 \eta^2 \int d^3y c(\mathbf{x} - \mathbf{y}) [\nabla f(\mathbf{y}) - e\mathbf{A}(\mathbf{y})].$$

Equation (80) in the text is obtained from this by approximating the c function by the δ function. The second term in the above expression gives a Meissmer current of the form

$$-e^2 v_0^2 \eta^2 \int d^3y c(\mathbf{x} - \mathbf{y}) \mathbf{A}(\mathbf{y}).$$

Owing to the relation (24), this can be put in the form

$$-e^2 v_0^2 \eta^2 (1 + F(\nabla^2)\nabla^2) \mathbf{A}(\mathbf{x})$$

with a certain function F which should be computed from the knowledge of the c function. Note that the c function has the range of order of $\xi = 1/l_0$.

To show that this is one of the Landau–Ginsburg equations, we recall that

$$\Delta = \lambda \langle \psi_\downarrow \psi_\uparrow \rangle, \tag{81}$$

where λ is the coupling constant introduced in Eq. (1). After gauge transformation, (81) becomes

$$\Delta(\mathbf{x}, t) = \langle e^{2iB/\eta + 2if} F^2(\varphi, \nabla B - \eta e\mathbf{A} + \eta \nabla f, \pi) \rangle, \tag{82}$$

where use was made of (29) and (74).

We write (82) as

$$\Delta(\mathbf{x}, t) = e^{2if(\mathbf{x},t)} |\Delta(\mathbf{x}, t)|. \tag{83}$$

It is now easy to see that (80) can be put in the form

$$\langle \mathbf{j} \rangle = \frac{v_0^2 \eta^2}{i4\,|\Delta|^2} [\Delta^*(\mathbf{x}, t) \nabla \Delta(\mathbf{x}, t) - \Delta(\mathbf{x}, t) \nabla \Delta^*(\mathbf{x}, t)]. \tag{84}$$

Inserting the value of $v_0^2 \eta^2$ given in (78), Eq. (84) becomes identical to the generalized Landau–Ginsburg equation obtained by Werthamer [see Eq. (16) of Ref. 4].

It may be seen from (82) that $|\Delta|$ is practically constant when Δ is varying so slowly that $\xi_0 \nabla f$ (where ξ_0 is the coherence length) is negligible and when the temperature is so low that the variation of F in (82) is mainly due to the space dependence of f.

As we mentioned in the beginning of this section, we do not give a full derivation of the first Landau–Ginsburg equation, but only outline its derivation. This time we are concerned with $\psi_\downarrow \psi_\uparrow$: first one has to express $\psi_\downarrow \psi_\uparrow$ in terms of the free operators (φ, B), then perform the gauge transformation (79), and finally compute the expectation value of the expression thus obtained.

7. VORTICES

As was shown in Sec. 4, the condition for the gauge transformation (26),

$$B(\mathbf{x}, t) \to B(\mathbf{x}, t) + \eta \int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}), \tag{85}$$

to leave the system invariant is expressed by the condition (73), i.e.,

$$\nabla^2 f(\mathbf{x}) = 0. \tag{86}$$

We notice that this condition leads to

$$\nabla^2 \int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) = 0, \tag{87}$$

provided the system is simply connected.

The condition (86) admits a variety of non-simply-connected solutions. The simplest solution of a cylindrically symmetric nature is given by

$$f(\mathbf{x}) = \tfrac{1}{2} \nu \varphi. \tag{88}$$

φ is the cylindrical angle of the vector $\mathbf{x}$ and ν is required to be an integer in order to make the order parameter $\Delta(x)$ single-valued.

Let us note that the function $c(\mathbf{x} - \mathbf{y})$ is an oscillating function of $|\mathbf{x} - \mathbf{y}|$ which is practically confined in a small domain of radius of the order of $1/l_0$ around the point $\mathbf{x}$. [l_0 is the cutoff momentum given by (13).] Since $f(x)$, given by (88), is multiply connected around the cylindrical axes, (87) does not hold anymore. Let us estimate how much the left-hand side of (87) deviates from zero. To do this, we define a new function

$$\begin{aligned} D(\mathbf{x} - \mathbf{y}) &= c(\mathbf{x} - \mathbf{y}), \quad \text{for} \quad |\mathbf{x} - \mathbf{y}| < r(\mathbf{x}), \\ &= 0, \qquad\qquad \text{otherwise}, \end{aligned} \tag{89}$$

where $r(\mathbf{x})$ is the distance of the point $\mathbf{x}$ from the axes. According to (89), the cylindrical axes are excluded from the domain of nonvanishing $D(\mathbf{x} - \mathbf{y})$. Therefore, the partial integration simply leads to

$$\nabla^2 \int d^3y D(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) = \int d^3y D(\mathbf{x} - \mathbf{y}) \nabla^2 f(\mathbf{y}) = 0. \tag{90}$$

When $r(\mathbf{x}) \gg 1/l_0$, then

$$\int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) = \int d^3y D(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \tag{91}$$

and, therefore,

$$\nabla^2 \int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \simeq 0. \tag{92}$$

On the other hand, the above quantity is considerably different from zero for $r(\mathbf{x}) \leqslant 1/l_0$, because in such a case (91) is not true anymore. This indicates that the persistent current

$$\langle \mathbf{j} \rangle = v_0^2 \eta^2 \nabla \int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \tag{93}$$

is conserved in the region far from the axes [i.e., $r(\mathbf{x}) \gg 1/l_0$], but is not in the region close to the axes. Therefore, some quasifermions need to be present in order to keep the current conservation.[15] These quasifermions are mainly concentrated in a cylinder of radius of the order

$$r_0 \approx 1/l_0 = v_0/2\Delta \approx \xi_0. \tag{94}$$

[15] To keep the conservation law we must give up the condition $\nabla^2 f = 0$ in the domain $r(\mathbf{x}) \leqslant 1/l_0$. Then, according to (74), $\mathbf{A}$ must exist in this domain so that we can construct the gauge-invariant combination $\mathbf{A} - \nabla f$ (see also Footnote 14). This $\mathbf{A}$ exhibits the response of the quasifermions effects. This argument shows also that creation of vortices requires that $\lambda \gtrsim 1/l_0$ where λ is the penetration depth.

We are thus led to the following picture.[16,17] Excited quasifermions (or "normal" electrons) form a core of radius ξ_0. Far from the core there exists a persistent current (93), which is given by

$$\langle \mathbf{j}(x) \rangle \simeq \tfrac{1}{2} v v_0^2 \eta^2 \nabla \varphi = \tfrac{1}{2} v v_0^2 \eta^2 [r(\mathbf{x})]^{-1} \mathbf{e}\varphi, \qquad (95)$$

where $\mathbf{e}\varphi$ is the unit vector in the circular direction.[18] In the intermediate region there is a mixture of persistent current and excited quasifermions.

To complete the picture, let us note that

$$e \oint_c \langle \mathbf{j}(\mathbf{x}) \rangle \cdot d\mathbf{s} = -e v_0^2 \eta^2 \pi v, \qquad (96)$$

where the integration is taken over a circle of radius $R \gg 1/l_0$. The left-hand side of (96) can be rewritten as

$$e \int d\mathbf{S} \cdot \nabla \times \langle \mathbf{j} \rangle = -\eta^2 v_0^2 e^2 \int d\mathbf{S} \cdot \nabla \times \mathbf{A},$$

where Eq. (80) has been used. Therefore, there appears a magnetic field concentrated mainly in the core of radius ξ_0.

The total magnetic flux is[19]

$$\Phi = \pi v / e. \qquad (97)$$

Let us close this section by noting that the result (94) can be understood by a simple intuitive argument based on the uncertainty principle. We have seen that the phase

$$\int d^3y c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}),$$

which corresponds to a vortex, is generated by the phason $B(\mathbf{x}, t)$ [cf. (85)]. Since there is a cutoff l_0 in phason momentum, the uncertainty principle

$$\Delta \mathbf{x} \cdot \Delta \mathbf{l} \gtrsim 1$$

leads us to conclude that the phasons cannot stay in a region smaller than

$$\Delta x \approx 1/l_0 \simeq \xi_0.$$

When they are confined in such a region they acquire, by the same uncertainty principle, enough energy to decay into quasifermions which remain after equilibrium is established to form the core of the vortex.

[16] G. Rickayzen, *Theory of Superconductivity* (Interscience Publishers, Inc., New York, 1965).

[17] P. G. De Gennes, *Superconductivity of Metals and Alloys* (W. A. Benjamin, Inc., New York, 1966).

[18] *Note Added in Proof:* In the case of charged superconductors the expression (93) should be replaced, according to Footnote (14), by

$$\langle \mathbf{j} \rangle = e v_0^2 \eta^2 \int d^3y c(\mathbf{x} - \mathbf{y}) [\nabla f(\mathbf{y}) - e\mathbf{A}(\mathbf{y})],$$

where $f(\mathbf{y})$ is given by (88) and $\mathbf{A}(\mathbf{y})$ represents the self-consistent vector potential. A detailed calculation of the vortex current using this formula has been done recently by our group. In that case the current takes the well-known form

$$\langle \mathbf{j} \rangle = \frac{\Phi}{8\pi^2 \lambda^3} K_1\left(\frac{r}{\lambda}\right) \mathbf{e}\varphi, \quad \text{for} \quad r \gg \xi_0.$$

An expression for the current inside the core has also been obtained. (See L. Leplae, N. Mancini, and H. Umezawa, report of work prior to publication.)

[19] In this article, $\hbar = 1$ and $c = 1$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Short-Range Interactions and Analyticity in Momentum Transfer*

J. D. FINLEY III†
University of California, Berkeley, California

(Received 13 February 1969)

We investigate possible characterizations of "short-range" interactions by means of conditions on the total transition probability $\|T\psi\|^2$. We desire to require that the total transition probability decrease exponentially as the displacement of the wavepacket ψ increases. It is shown that this can only be done for certain types of sequences of wavepackets, and a criterion is developed for selecting these "proper" wavepackets. We then show for all such interactions that what is essentially the square of the scattering amplitude is analytic, in the cosine of the scattering angle, in an ellipse which always includes the physical region. We then compare these results with the Schrödinger potential theory and thereby relate the type of the exponential decrease of a potential to the type of the exponential decrease of the corresponding transition probability. Finally, our results are compared with similar results previously obtained by others working on this problem.

1. INTRODUCTION

This study is concerned with a characterization of the short-range property of strong interactions by means of statements about observable quantities, and with the resulting implications for the scattering matrix. In particular, we impose conditions on the transition probabilities related to elastic cross sections for a fixed incoming wavepacket, and show that there is a relationship between these conditions and the (traditional) notions of "short range" which apply within potential theory.

It is a basic assumption of hadron physics (the dynamics of the strongly interacting particles) that the interactions between particles are, in some sense, of short range. In a very intuitive sense, this means that two particles interact only when they are "near" each other, so that, asymptotically, a system can be described by free-particle states. In the Schrödinger potential theory, "short range" usually means that the interaction potential is of exponential decrease, or faster, as the distance between packet and scattering center increases. This behavior of the potential implies certain characteristic properties of other quantities in the theory, such as the phase shifts, the scattering matrix, and the transition probabilities. For example, the fact that the potential is of "exponentially" short range implies that, for any fixed energy, the scattering amplitude will be analytic in the $\cos\theta$ plane in an ellipse which contains the physical region, where θ is the scattering angle.

If we consider a more realistic, fundamental physical theory in which the forces are not derivable from a potential, we may wonder how to characterize a short-range interaction. In analytic S-matrix theory,[1] the existence of regions of analyticity for the scattering amplitude is taken as a fundamental characterization of short range.

We intend here to characterize short-range interactions by a property of the total transition probability: Consider a scattering situation in which ψ represents the initial asymptotic state-vector. Then, $S\psi$ is the final asymptotic state-vector in the outcome of the scattering event, where S is the scattering matrix. If $iT = S - 1$, the quantity $\|T\psi\|^2 = \|(S-1)\psi\|^2$, the square of the norm of the vector $T\psi$, is a measure of the amount of scattering. In the following we call this measure the total transition probability. (We call it this, following custom, although it is not strictly a probability. If the vector $T\psi$ is orthogonal to the original state ψ, it will be a strict probability.)

If the packet ψ stays away at all times from the scattering center, then we expect $\|T\psi\|^2$ to be "small." In particular, if we construct a sequence of packets which stay more and more away from the scattering center, then we expect the corresponding sequence of transition probabilities to decrease. We may regard the sequence of wavepackets as a probe to measure the rate of decrease of the interaction with distance. Because of the evidence from Lagrangian field theory (via perturbation theory), it is natural to expect exponentially decreasing forces even though these forces are not derivable from a potential. Therefore, at least for sequences which are "efficient probes," we expect the transition probabilities associated with short-range interactions to decrease exponentially, as the sequence of wavepackets stays more and more away from the scattering center.

* Research supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under Grant No. AF-AFOSR-68-1471.

† Present address: University of New Mexico, Albuquerque, New Mexico.

[1] See, for instance, G. F. Chew, *The Analytic S-Matrix* (W. A. Benjamin, Inc., New York, 1966), or D. Olive, Phys. Rev. **135B**, 745 (1964).

We would like to specify the relation between "short range" and the analyticity of the scattering amplitude a bit more fully. In Appendix A, it is shown that if a potential is of exponential decrease[2] of order $b > 1$, then the corresponding scattering amplitude [which depends upon the magnitude of the incoming (relative) momentum p and the cosine of the scattering angle z] will be an entire analytic function in the variable z, for every fixed nonnegative p; while if $b < 1$, there will be no such region of analyticity, but rather the boundary of the physical region will include a singularity. If, however, $b = 1$, then the corresponding scattering amplitude will be a function analytic in z (for every fixed nonnegative p) in some ellipse with foci ± 1, the size of which depends upon the type. Therefore, in our attempts to characterize short-range interactions which are not derivable from a potential by an exponential decrease of a sequence of transition probabilities, we show that this implies an ellipse of analyticity in z for the scattering amplitude. [We find it convenient to use the notation $f(s) = \bar{O}(e^{-as^r})$ to indicate that $f(s)$ is of exponential decrease of order r and type a, in an analogy with the standard order symbol.]

Recently there has been some interest in the relation between the decrease of the transition probability and the region of analyticity of the scattering amplitude. Omnes[3] assumed an exponential decrease for the transition probability for a selected sequence of Gaussian wavepackets, and he derived an ellipse of analyticity in z for fixed nonnegative p for the square of that scattering amplitude. Similar studies were also carried out by Kugler and Roskies,[4] who derived an ellipse of analyticity, as well as MacDowell, Roskies, and Schroer.[5] By making similar assumptions about the decrease of the transition probability (using wavepackets of compact support in momentum space), Stapp and Chandler[6] showed that the scattering amplitude is C^∞ in all arguments in the physical region, excluding Landau singularities.

In Sec. 2 we discuss whether or not certain specific families of wavepackets are suitable for use as probes. In Sec. 3 we derive an ellipse of analyticity for what is essentially the square of the scattering amplitude, based on the assumption that $\|T\psi^s\|^2 = \bar{O}(e^{-as})$ for some specific families of wavepackets ψ^s, where s is a parameter which measures how far away that wavepacket is from the scattering center, and $a > 0$. In Sec. 4 we compare the results obtained in Sec. 3 with the Schrödinger potential theory. In particular, we attempt to relate the constant a [in the assumption that $\|T\psi^s\|^2 = \bar{O}(e^{-as})$] to the range of a potential $V(r) = O(e^{-\mu r})$. We also compare the various results which we have obtained with the results obtained by others working on this problem.

2. A MEASURE OF THE EXTENT TO WHICH A PARTICLE APPROACHES THE SCATTERING CENTER

Although the problem with which we are truly concerned is one in which there are many different kinds of particles, each with different spins and statistics, the essentials do not depend upon all these details. We will, therefore, restrict our discussion to dealing merely with one type of scalar boson. Furthermore, we only consider the case of two incident particles, which is equivalent to the case of a single particle (tracing out the relative motion) incident upon a fixed target. We intend that a quite general interaction (of short-range) may occur while this particle is in the neighborhood of the target, so that there may be an arbitrary number of particles after the collision. (The quantity $\|T\psi\|^2$ is the total transition probability *into all channels*.)

We take ψ^s to be a one-parameter family of initial asymptotic states ordered so that s is a measure of their displacement from the scattering center. We fix the normalization[7] of the S matrix in the momentum-space representation by

$$(T_e\psi^s)(\mathbf{q}) = \int d\mathbf{p}\delta(p - q)T_e'(\mathbf{p}, \mathbf{q})\psi^s(\mathbf{p}), \quad (2.1)$$

where T_e is the elastic portion of the T matrix. We designate the corresponding family of transition probabilities by

$$P(s) = \|T\psi^s\|^2. \quad (2.2)$$

Then,

$$P(s) = \int dp' \int d\mathbf{p}\psi^{*s}(\mathbf{p}')\psi^s(\mathbf{p})\delta(p - p')A(p, z), \quad (2.3)$$

where $pp'z = \mathbf{p} \cdot \mathbf{p}'$, and $A(p, z)$ is the absorptive part of the S matrix. We expand $A(p, z)$ in the standard

[2] We say that a function $f(r)$ is of exponential decrease of order b if for every $\epsilon > 0$, $\lim \exp(r^{b-\epsilon})f(r) = 0$ as $r \to \infty$, while $\exp(r^{b+\epsilon})f(r)$ is unbounded as $r \to \infty$. If there exists $a > 0$ such that

$$\lim_{r\to\infty} \exp(ar^b)f(r) = 0,$$

then the type [of $f(r)$] is the least upper bound of all such a. Otherwise, the type is 0. [If for every $a > 0$, $\lim \exp(ar^b)f(r) = 0$ as $r \to \infty$, then the type is infinite.] Obviously, a function which decreases like some inverse power of its argument is of order 0. If no order is mentioned, then the order will be assumed to be 1.

[3] R. Omnes, Phys. Rev. **146**, 1123 (1966).

[4] M. Kugler and R. Roskies, Phys. Rev. **155**, 1685 (1967).

[5] S. MacDowell, R. Roskies, and B. Schroer, Phys. Rev. **166**, 1691 (1968).

[6] H. Stapp and C. Chandler, J. Math. Phys. **10**, 826 (1969); C. Chandler, Phys. Rev. **174**, 1749 (1968).

[7] In the following, we use units such that $c = 1 = \hbar$.

partial-wave expansion,

$$A(p, z) = (\pi p^2)^{-1} \sum_{l=0}^{\infty} (2l + 1) P_l(z) A_l(p). \quad (2.4)$$

From the unitarity of the S matrix, we have that

$$0 \leq A_l(p) \leq 1. \quad (2.5)$$

From Eqs. (2.3) and (2.4),

$$P(s) = 4 \sum_{l=0}^{\infty} \int_0^{\infty} dp p^2 A_l(p) f_l^s(p), \quad (2.6)$$

where

$$f_l(p) \equiv \frac{2l + 1}{4\pi} \int d\Omega' \int d\Omega \psi^*(\mathbf{p}') \psi(\mathbf{p}) P_l(z)|_{p=p'}. \quad (2.7)$$

We note that, for every $b \geq a \geq 0$,

$$0 \leq \int_a^b dp p^2 f_l^s(p) \leq 1,$$

because of the normalization of the wavefunction $\psi(\mathbf{p})$. Finally, we define our conventions so that $\psi(\mathbf{x}, t)$, the solution to the Klein–Gordon (or Schrödinger) equation, is related to its momentum-space representation $\psi(\mathbf{p})$ by

$$\psi(\mathbf{x}, t) = (2\pi)^{-\frac{3}{2}} \int d\mathbf{p} e^{i\mathbf{p}\cdot\mathbf{x} - i\omega(p)t} \psi(\mathbf{p}), \quad (2.8)$$

where $\omega(p)$ is the energy corresponding to the momentum $\mathbf{p}$. [In the nonrelativistic (Schrödinger) case, $\omega(p) = p^2/2m$, while in the relativistic (Klein–Gordon) case, $\omega(p) = (p^2 + m^2)^{\frac{1}{2}}$, where m is the reduced mass of the particle system under consideration.]

We need a method of selecting "efficient" wavepackets: those which always stay far away from the scattering center. Obviously, some measure is needed of how much of the wavepacket is in the neighborhood of the scattering center for all times—a "measure of closeness." We will use

$$G[R, \psi] = \int_{-\infty}^{\infty} dt \int_{|\mathbf{x}| \leq R} d\mathbf{x} \, |\psi(\mathbf{x}, t)|^2 \quad (2.9)$$

as our measure of closeness. Then, $G[R, \psi]$ is the integral over all times of the probability, at the time t, that the wavepacket is within a sphere of radius R about the scattering center. We will take the "smallness" of $G[R, \psi]$ to mean that the wavepacket ψ is, at all times, "away" from the scattering center. There are obviously some quantities other than G which might have been considered as a means of selecting "efficient" wavepackets. However, this particular measure is rather reasonable and convenient for actual computation with specific wavepackets. Also, for the particular wavepackets considered in this paper, any other reasonable choice would give the same results.

A relation between G and $\|T\psi\|^2$ can be seen by considering a theorem proved by Cook.[8] If, in a potential theory, V is the potential and H the total Hamiltonian, both of these being represented as operators on the Hilbert space of states, and $\psi(t) = e^{-it(H-V)}\psi$, Cook shows that

$$\|T\psi\|^2 = \left\| \int_{-\infty}^{\infty} dt e^{itH} V \psi(t) \right\|^2 \leq \left(\int_{-\infty}^{\infty} dt \, \|V\psi(t)\| \right)^2.$$

The size of $\|T\psi\|^2$ depends on both the choice of a sequence of wavepackets and on the rate of decrease of the interaction. In particular, if the interaction decreases very slowly with distance, then we would not expect $\|T\psi\|^2$ to be very "small," regardless of the choice of sequence of wavepackets. In order, therefore, to see the dependence of $\|T\psi\|^2$ on the wavepacket, we choose

$$V(r) = C, \quad r < R,$$
$$= 0, \quad r > R.$$

For this potential,

$$\int_{-\infty}^{\infty} dt \, \|V\psi(t)\|^2 = C^2 G[R, \psi].$$

Although this is not quite the same as the expression appearing in Cook's theorem, we can see that if G is not "small," then $\|T\psi\|^2$ is not likely to be "small" either.

Using Eqs. (2.7) and (2.8), we have

$$G[R, \psi^s] = (2\pi)^{-2} \int d\mathbf{p}' \int d\mathbf{p} \psi^{*s}(\mathbf{p}') \psi^s(\mathbf{p}) \delta(\omega - \omega') \times \int_{|\mathbf{x}| \leq R} d\mathbf{x} e^{i\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')}$$
$$= 4 \sum_{l=0}^{\infty} \int_0^{\infty} dp p^4 \frac{dp}{d\omega} f_l^s(p) \int_0^R dx x^2 j_l^2(xp), \quad (2.10)$$

where $\omega(p)$ is the energy corresponding to momentum p, and $j_l(xp)$ is the spherical Bessel function of order l. [We note that $\omega(p) \geq m$, which implies that the relativistic version of G is always larger than the nonrelativistic version.] It is convenient to take the last expression in Eq. (2.10) as the *definition* of $G[R, \psi^s]$. We then find (see Appendix B) that

$$G[R, \psi] < 2\pi R(1 + 4m^2R^2)^{\frac{1}{2}} \|\psi\|^2;$$

hence $G[R, \psi]$ is certainly well defined for all square-integrable wavefunctions ψ. These questions will be

[8] J. M. Cook, J. Math. & Phys. **36**, 82 (1957).

discussed further in Appendix B. We see that $G[R, \psi]$ is a monotonically increasing function of R. Also, we note that it is a sum over l of terms each of which is positive, so that it can decrease no faster in s than any of its terms. We denote these terms by $G_l[R, \psi^s]$.

Omnes[3] considers Gaussian wavepackets of the form

$$\psi^s(\mathbf{p}) = (b^2/\pi)^{\frac{3}{4}} \exp[-b^2(\mathbf{p} - \mathbf{k})^2/2] \exp(-i\mathbf{s}\cdot\mathbf{p}), \tag{2.11}$$

where $\mathbf{s}\cdot\mathbf{k} = 0$. He shows that, in potential theory, $P(s)$ decreases exponentially with s [for a potential $V(r)$ which is everywhere finite, and $|V(r)|\, e^{\mu r} < C$, $C = \text{const}$] provided that the width b increases with s so that

$$b^2k = s. \tag{2.12}$$

If, instead, b is independent of s, the spreading of the wavepacket causes it to have nonnegligible overlap with the scattering center at very large times, even though it had been displaced a very great distance away from the scattering center at time $t = 0$, as can be seen from Eq. (2.8) of Omnes.[3]

The measure $G[R, \psi]$ should give similar results. By some simple inequalities, we find that

$$G[R, \psi^s] \geq G_0[R, \psi^s] > C_1 m b^3 R^3 d^{-4} \times \exp(-b^2k^2) + O(R^5/d^6),$$

for sufficiently large values of d/R, where C_1 is a numerical constant, and $d^2 = s^2 - b^4k^2$. However, by insisting that the width increase with s according to Eq. (2.12), we find that

$$G[R, \psi^s] \leq C_2 m R^3 b^{-1} \exp[-sk(1 - R/s)^2],$$

for $s > R$, which agrees well with Omnes' results. It is therefore necessary to use this family of wavepackets, a different one for each impact parameter, in order to keep "small" the probability that the wavepacket will be near the scattering center at infinite time.

Another interesting family of wavepackets is

$$\psi_2^s(\mathbf{x}) = Bp^{-\frac{1}{2}} \exp(b^2\mathbf{p}\cdot\mathbf{k} - i\mathbf{p}\cdot\mathbf{s}), \quad p < k,$$
$$= 0, \quad p > k, \tag{2.13}$$

where $B = (2\pi)^{-\frac{1}{2}}(x/k)(\sinh x)^{-1}$, $x = (bk)^2$, $\mathbf{k}\cdot\mathbf{s} = 0$, and $\psi_2^s(\mathbf{p})$ has been displaced by an amount $\mathbf{s}$ from $\psi_2^0(\mathbf{p})$. We note that in $\mathbf{x}$ space this wavepacket is extremely broad—$\langle\mathbf{x}^2\rangle_{\text{av}} = \infty$—although $\langle\mathbf{p}\rangle_{\text{av}}$ and $\langle\mathbf{p}^2\rangle_{\text{av}}$ are well defined. We then find that, for large values of d/R, $d^2 = s^2 - b^4k^2$,

$$G[R, \psi_2^s] \geq G_0[R, \psi_2^s] \geq (\tfrac{4}{3}) m R^3 b^4 k^2 d^{-3} \sinh^{-2} x.$$

[See Eqs. (2.7), (2.10), and (2.13).] Again we choose $b^2k = s$ and then find that $G[R, \psi_2^s] = O(e^{-2s(k-\delta)})$, $\delta > 0$, as s (and b) tend to infinity, for fixed k.

3. SOME RESULTS FOR SPECIFIC WAVEPACKETS

We now develop the basic results for the families of Gaussian wavepackets given by Eq. (2.11), with the extra condition $b^2k = s$. Particular cases of these results have previously been given by Omnes,[3] and Kugler and Roskies.[4] Our basic assumption at this point is the exponential decrease of $P(s)$. Using Eqs. (2.6) and (2.11), we have[9]

$$P(s) = 16\pi(s/\pi k)^{\frac{3}{2}} e^{-sk} \sum_{l=0}^{\infty} [(2l+1)!]^{-1} \times \int_0^{\infty} dp\, p^2 A_l(p)(2sp)^{2l} e^{-sp^2/k}. \tag{3.1}$$

In studying the consequences of this assumption, we work with what is essentially the Laplace transform of $P(s)$, namely[10]

$$\tilde{P}(w) = \int_0^{\infty} ds e^{-ws}(\pi k^3/16s)^{\frac{1}{2}} P(s) = \sum_{l=0}^{\infty} \int_0^{\infty} dp A_l(p) h_l(w, p), \tag{3.2}$$

where

$$h_l(w, p) = [2pk/(p^2 + k^2 + kw)]^{2l+2}. \tag{3.3}$$

The assumption that $P(s) = \bar{O}(e^{-as})$ implies[11] that $\tilde{P}(w)$ is an analytic function of w in the half-plane $\text{Re}\,(w) > -a$. On physical grounds, we assume further that there is at least one l, say l_0, such that for all $p_0 > 0$

$$\int_0^{p_0} dp A_{l_0}(p) > 0. \tag{3.4}$$

We now investigate the behavior of $\tilde{P}(w)$ as a function of the complex variable w, for a *fixed* $k > 0$. Under the conditions on $A_l(p)$ [given by Eqs. (2.5) and (3.4)], there exists a nonnegative number a such that the series in Eq. (3.2), defining $\tilde{P}(w)$, converges to an analytic function of w whenever $\text{Re}\,(w) > -a$, and such that $\tilde{P}(w)$ is singular at $w = -a$. We relegate this proof to Appendix C since the details are totally mathematical. It is shown there that if we define

$$K_\epsilon^2(p) \equiv \limsup_{l\to\infty} \left(\int_{p-\epsilon}^{p+\epsilon} dx A_l(x) \right)^{1/l} \tag{3.5}$$

[9] The calculation of $f_l^s(p)$ for the Gaussian wavepackets may be facilitated by the use of the standard expansion of a plane wave in terms of spherical waves. See, for instance, Ref. 5, Appendix C.

[10] This particular method for study of the exponential decrease of a function was pointed out by Omnes in Ref. 3.

[11] D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, 1946).

and

$$K(p) \equiv \lim_{\epsilon \to 0} [K_\epsilon(p)], \tag{3.6}$$

then

$$0 \leq a = -\sup_{0 \leq p \leq \infty} [2pK(p) - (p^2 + k^2)/k] \leq k. \tag{3.7}$$

Therefore $2pK(p) - (p^2 + k^2)/k \leq -a$, for every $p \geq 0$ and $k \geq a$. However, since $K(p)$ is certainly independent of k, we see that

$$K(p) \leq \inf_{k \geq a} [(p^2 + k^2)/2pk - a/2p].$$

By calculating this infimum, we find the useful condition

$$\begin{aligned} K(p) \leq \gamma(p) &= 1 - a/2p, \quad p \geq a, \\ &= p/2a, \qquad p \leq a, \end{aligned} \tag{3.8}$$

which is the necessary condition that the total transition probability be such that

$$P(s) = \bar{O}(e^{-as}), \tag{3.9}$$

for all $k \geq a > 0$, where a is a number intended to characterize the exponential range of the forces.

However, the converse is also true. In order to see this, we use Eqs. (3.1), (3.2), and the standard inversion formula for the Laplace transform to obtain

$$\begin{aligned} &(\pi k^3/16s)^{\frac{1}{2}} P(s) \\ &\quad = 4se^{-sk} \int_0^\infty dp p^2 A_0(p) e^{-sp^2/k} \\ &\qquad + (2\pi i)^{-1} \int_{c-i\infty}^{c+i\infty} dw e^{sw} \sum_{l=1}^{\infty} \int_0^\infty dp A_l(p) h_l(w, p), \end{aligned}$$

which is valid for any $c > -a$. In Appendix C it is then shown that

$$\begin{aligned} P(s) \leq 4e^{-sk} &+ (16s/\pi k^3)^{\frac{1}{2}} \\ &\times [\tfrac{1}{2}(17k + c)\tilde{P}(c) + 8k^2/3]e^{cs}, \end{aligned}$$

for every $c > -a \geq -k$. We therefore conclude that the condition (3.8) on $K(p)$ is in fact a necessary and sufficient condition for Eq. (3.9) to hold.

Consider now the integral of $A(p, z)$, given by Eq. (2.4), over a small interval $[p - \epsilon, p + \epsilon]$ of the nonnegative real axis, i.e.,

$$\int_{p-\epsilon}^{p+\epsilon} dx A(x, z) = \pi^{-1} \sum_{l=0}^{\infty} (2l + 1) P_l(z) \int_{p-\epsilon}^{p+\epsilon} dx x^{-2} A_l(x). \tag{3.10}$$

By standard theorems on series of Legendre polynomials,[12] the series converges to an analytic function of z when z is in the interior of an ellipse with foci at $z = \pm 1$, and semimajor axis

$$z_1(p, \epsilon) = \tfrac{1}{2}\{[K_\epsilon(p)]^{-2} + [K_\epsilon(p)]^2\}.$$

Let

$$z_1(p) \equiv \tfrac{1}{2}[K^{-2}(p) + K^2(p)] \geq \tfrac{1}{2}[\gamma^{-2}(p) + \gamma^2(p)],$$

where the last inequality holds since $K(p) \leq 1$. Therefore,[13] for sufficiently small $\epsilon > 0$,

$$z_1(p, \epsilon) \geq [\gamma^{-2}(p) + \gamma^2(p)].$$

So Eq. (3.8) gives a lower bound on the semimajor axis of the ellipse of analyticity, which depends on the momentum p, and the "range of the interaction," as specified by the quantity a.

It is customary to consider, rather than the quantity $A(p, z)$, the quantity

$$\mathcal{A}(p, t) = A[p, z(p, t)], \tag{3.11}$$

where

$$z = z(p, t) = 1 + t/2p^2, \quad \text{or} \quad t = 2p^2(z - 1). \tag{3.12}$$

The variable t is the momentum transfer and has the real interval $[-4p^2, 0]$ as its physical region.[14] Let us set $t_1(p, \epsilon) = t[p, z_1(p, \epsilon)]$ and $\bar{t}_1(p) = \{p[\gamma^{-1}(p) - \gamma(p)]\}^2$. It then follows from Eqs. (3.8) and (3.10) that, for sufficiently small $\epsilon > 0$,

$$\begin{aligned} t_1(p, \epsilon) \geq \bar{t}_1(p) &= a^2[(4p - a)/(4p - 2a)]^2, \quad p \geq a, \\ &= 4a^2[1 - (p/2a)2]^2, \qquad p \leq a. \end{aligned} \tag{3.13}$$

One sees that $\bar{t}_1(p)$ is a monotonically decreasing function of p such that $\bar{t}_1(0) = 4a^2$ and

$$\lim_{p \to \infty} \bar{t}_1(p) = a^2.$$

[The curve $\bar{t}_1(p)$ is plotted, for $a = \frac{3}{4}(\sqrt{3})\mu$, in Fig. 1.]

Summarizing, we find that the total transition probability $P(s)$ decreases exponentially like $\bar{O}(e^{-as})$ if and only if, for every p and ϵ such that $p \geq 0$, $\epsilon > 0$, and $p - \epsilon \geq 0$, $\int_{p-\epsilon}^{p+\epsilon} dx \mathcal{A}(x, t)$ is an analytic function of t, analytic in an ellipse with foci at $-4p^2$

[12] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (University Press, Cambridge, 1962), p. 323.

[13] In fact we have this condition only when $K(p) < \gamma(p)$. If the equality $K(P) = \gamma(p)$ holds, we do not have the above. Rather, since $\gamma(p)$ is monotonically increasing, for sufficiently small ϵ there exists $\eta > 0$, as small as one likes, so that $z_1(p, \epsilon) \geq \frac{1}{2}[\gamma^{-2}(p + \eta) + \gamma^2(p + \eta)]$. Since η is as small as desirable, the equation is true with an error as small as one wants if $K = \gamma$, and with no error in all other cases. If one desires to determine the analyticity region of $\int_{p-\epsilon}^{p+\epsilon} dx A(x, z)$ for larger ϵ, he may decompose it into a finite sum of similar quantities with sufficiently small ranges of integration so that their ellipses of analyticity are easily shown to be bounded by $\gamma(p)$ for some p. Then $\int_{p-\epsilon}^{p+\epsilon} dx A(x, z)$ will certainly be analytic in the smallest such ellipse.

[14] It is customary to refer to the set of *physical* values which the variable t (or z) can assume as the "physical region," although the set is actually only a line segment.

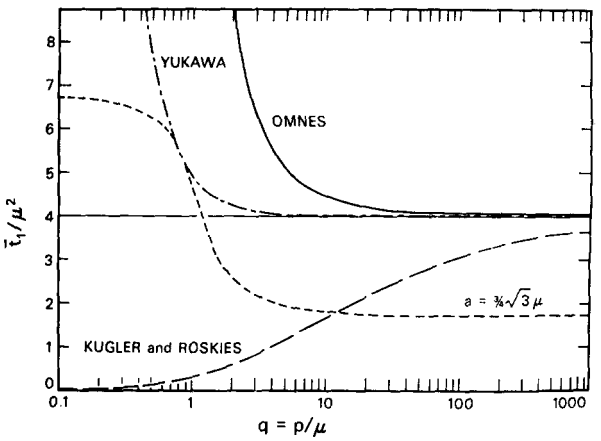

FIG. 1. A comparison of our results with similar results derived by others: the rightmost points of the ellipses of analyticity of $\mathcal{A}(p, t)$.

and 0, the size of which depends upon p and ϵ. This ellipse has a semimajor axis which is larger than a^2, for all $p \geq 0$, and so is always considerably larger than the physical region.

The quantity which has been shown to be analytic is $\int_{p-\epsilon}^{p+\epsilon} dx\mathcal{A}(x, t)$ and not $\mathcal{A}(p, t)$ itself. This is all the information contained in our assumption about the transition probability. We may then look at $\mathcal{A}(p, t)$ as a distribution. For instance, let $\mathfrak{D}$ be the space of all C^∞-testing functions of compact support defined over $[0, \infty)$ in the variable p. Then, instead of the requirement that, for every $p \geq 0$ and for every $\epsilon > 0$, $\int_{p-\epsilon}^{p+\epsilon} dx\mathcal{A}(x, t)$ be an analytic function of t in the given ellipse, we could instead require that, for every $f(p)$ in $\mathfrak{D}$, $\int_0^\infty dpf(p)\mathcal{A}(p, t)$ be an analytic function of t in some ellipse determined by the support of $f(p)$.[15] This fact has already been pointed out by Kugler and Roskies,[4] and MacDowell, Roskies, and Schroer,[5] but is worth repeating here.

We would also like to show that the above arguments are not solely dependent upon the nature of the Gaussian wavepackets for their validity. In particular, let us consider the family of wavepackets $\psi_2^s(\mathbf{p})$ given by Eq. (2.13). For that wavepacket[9] with the change of width characterized by $b^2k = s$,

$$f_l^s(p) = \frac{2}{p}\frac{(2sp)^{2l}}{(2l+1)!}\left(\frac{s}{\sinh sk}\right)^2, \quad p < k,$$
$$= 0, \quad p > k. \tag{3.14}$$

From Eq. (2.6), using the Laplace transform, we have

$$\tilde{P}'(w) = \int_0^\infty dse^{-ws}P(s) = 4\sum_{l=0}^{\infty}\int_0^k dpA_l(p)g_l(w, p), \tag{3.15}$$

where

$$g_l(w, p) = \frac{(2p)^{2l+1}}{(2l+1)!}\int_0^\infty dse^{-ws}s^{2l+2}(\sinh sk)^{-2}$$
$$= 8(l+1)(2p)^{2l+1}\sum_{m=1}^{\infty} m(w + 2mk)^{-2l-3}. \tag{3.16}$$

The study of the function $\tilde{P}'(w)$ in this case can be carried out in complete analogy with the case of the Gaussian wavepackets. The half-plane of analyticity is again determined by the convergence of the series defining $\tilde{P}'(w)$ for real w. From Eq. (3.16), we find that

$$\lim_{l\to\infty}[g_l(w, p)]^{1/l} = [2p/(w + 2k)]^2.$$

In analogy with Eq. (3.8), we find the condition

$$K(p) \leq 1 - a/2p, \quad p \geq a. \tag{3.17}$$

[We obtain no information about the size of $K(p)$ for values of $p < a$.] The condition (3.17), which we note is exactly the same as condition (3.8) (in the range of validity) derived on the basis of Gaussian wavepackets, would then again give us an ellipse of analyticity for $\int_{p-\epsilon}^{p+\epsilon} dx\mathcal{A}(x, t)$.

4. A COMPARISON WITH POTENTIAL THEORY

Suppose that the interaction which we have been studying is such that for large distances it is $O(e^{-\mu r})$. An obvious question which then arises is the relation between the number a, in the assumption $P(s) = \bar{O}(e^{-as})$, and the quantity μ. It has also been explicitly assumed in Sec. 3 that a is independent of k, the mean momentum of the wavepacket. It is not clear that this assumption is actually possible. In fact, we have already found that in the case of the Gaussian packets, we had to restrict $k \geq a$ in order to maintain this independence.

In order to study this problem more fully, let us consider it in potential theory. We assume the Yukawa potential,

$$V(r) = Ce^{-\mu r}/r. \tag{4.1}$$

From some results of Carter, it follows that the first singularity in t of $\mathcal{A}(p, t)$ will be the same as in the first Born approximation to the T matrix[16] (see Appendix A for the details),

$$T'^B(p, z) = -(2\pi)^{-2}(m/p)\int d\mathbf{r}e^{i\mathbf{r}\cdot(\mathbf{p}-\mathbf{q})}\big|_{p=q}. \tag{4.2}$$

Then, for our potential we have

$${}_yT'^B(p, z) = -(mC/2\pi p^3)(1 + \mu^2/2p^2 - z)^{-1} \tag{4.3}$$

and[16]

$${}_yA_l^B(p) = [(mC/p)Q_l(1 + \mu^2/2p^2)]^2, \tag{4.4}$$

[15] The actual ellipse can be determined by the method explained in Ref. 13.

[16] See, for instance, M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964).

where $Q_l(v)$ is the Legendre function of the second kind. [We use ${}_yT'$ to indicate T' for the specific potential $V(r)$ given by Eq. (4.1).]

We may ask if the analyticity region of the above T matrix is consistent, for some choice of a, with the analyticity region derived for a general $A(p, z)$ in Sec. 3. [We note that it is permissible here to use ${}_yA(p, z)$ rather than an integral over a small region, since ${}_yA(p, z)$ is itself an analytic function.] From Sec. 3 we recall that ${}_yA(p, z)$ is a function of z, analytic in an ellipse with foci at $z = \pm 1$, and semimajor axis ${}_yz_1 = \frac{1}{2}({}_yK^2 + {}_yK^{-2})$, where[17]

$$ {}_yK(p) \equiv \limsup_{l\to\infty} [{}_yA_l(p)]^{\frac{1}{2}l} = \limsup_{l\to\infty} [{}_yA_l^B(p)]^{\frac{1}{2}l} $$
$$ = \lim_{l\to\infty} [Q_l(z_0)]^{1/l} = z_0 - (z_0^2 - 1)^{\frac{1}{2}}, \qquad (4.5) $$

with

$$ z_0 = 1 + \mu^2/2p^2. \qquad (4.6) $$

We therefore have that [since all the ${}_yA_l(p) \geq 0$] the nearest singularity to the physical region is at

$$ z = {}_yz_1 = 2z_0^2 - 1 = 1 + 2\mu^2/p^2 + \mu^4/2p^4. $$

The corresponding value of t is

$$ t_1(p) = 2p^2({}_yz_1 - 1) = \mu^2(4 + \mu^2/p^2). \qquad (4.7) $$

Therefore, the largest ellipse with foci at $t = -4p^2$ and 0, in which ${}_y\mathcal{A}(p, t)$ is analytic, has rightmost point $t_1(p)$, given by Eq. (4.7). We may compare this ellipse with the ellipse of analyticity derived in Sec. 3. The idea behind this is that, by a proper choice of the constant a, we may obtain an analyticity domain which is most like that of the Yukawa interaction.

If we choose $a = 2\mu$, which was the choice made by Omnes,[3] we may then compare the two curves obtained—namely, t_1 from Eq. (4.7) for the Yukawa potential (labeled Yukawa in Fig. 1), and t_1 from Eq. (3.13) for the choice $a = 2\mu$, $p > \mu$ (labeled Omnes in Fig. 1). We find that they agree only at infinite momentum, while for all finite $p > \mu$, t_1 (Omnes) > t_1 (Yukawa).[18] This would mean that our assumption, which we desire to make for arbitrary short-range interactions, would not be valid for the Yukawa potential. From this we see that the obvious first choice ($a = 2\mu$) is not suitable. In a similar manner, we might determine that there are, in fact, suitable constant choices for a. An example is to be seen in Fig. 1 by the curve for $t_1(p)$ given by Eq. (3.13) with the choice $a = \frac{3}{4}(\sqrt{3})\mu$, which for no value of p lies above the curve t_1 (Yukawa). We will show below that this choice of a is the maximum constant choice consistent with the restriction $k \geq a$. We will derive this along with more complete statements about the decrease of the transition probability.

We now compute the transition probability for this particular combination of an incoming Gaussian wavepacket, whose shape is changed according to the formula $b^2k = s$ and the Yukawa potential given by Eq. (4.1). From the analysis in Sec. 3 [see, in particular, Eq. (3.7) and the following material], we have that ${}_y\tilde{P}(w)$ is analytic for every value of w such that, for all $p \geq 0$,

$$ |w + k + p^2/k| > 2p_yK(p), $$

where ${}_yK(p)$ is given by Eq. (4.5). Therefore, if we designate $w_0(k)$ as the (real) location of the rightmost singularity of ${}_y\tilde{P}(w)$, and set

$$ q = p/\mu, \quad u = k/\mu, \quad \text{and} \quad w_0(k) = -\mu v_0(u), \qquad (4.8) $$

we have that

$$ v_0(u) = \inf_{0\leq q\leq\infty} v(q, u), \qquad (4.9) $$

with

$$ v(q, u) = (q - u)^2/u + [(1 + 4q^2)^{\frac{1}{2}} - 1]/q. \qquad (4.10) $$

We find that $v_0(u)$ is a monotonically increasing continuous function of u such that

$$ v_0(0) = 0, \quad \lim_{u\to\infty} [v_0(u)] = 2. $$

Moreover, for $u \leq \frac{3}{4}\sqrt{3}$, $v_0(u) = u$. The graph of $v_0 = v_0(u)$ is given in Fig. 2 (labeled Gaussian). (The calculation was done numerically.) Since ${}_yP(s) = \tilde{O}(e^{-\mu s v_0(k/\mu)})$ for large s, it follows that for extremely large values of k/μ the decrease of ${}_yP(s)$ is given approximately by $\tilde{O}(e^{-2\mu s})$. Since $v_0(u)$ is monotonically increasing, it follows that any choice of $a < 2\mu$ would be suitable, provided that the incoming wavepackets are restricted to large enough values of their mean momentum k. So long as $a \leq \frac{3}{4}(\sqrt{3})\mu$, it is sufficient to restrict $k \geq a$.

Instead of using the Gaussian wavepackets, we could have used the family of wavepackets $\psi_2^s(\mathbf{p})$ given by Eq. (2.13). From Eq. (3.17) we find that for the Yukawa potential and these wavepackets,

$$ {}_yP_2(s) = \tilde{O}(e^{-2ks[1-{}_yK(k)]}). $$

Of interest then is the coefficient of decrease [the type of ${}_yP_2(s)$] which, measured in units of μ, is $2(k/\mu) \times [1 - {}_yK(k)]$. With $u = k/\mu$, we find that this is a monotonically increasing function of u which vanishes at $u = 0$ and has limit value 2 as u goes to infinity. In Fig. 2 this function (labeled ϕ_2) is compared with the function $v_0(u)$ for the Gaussian wavepackets

[17] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Cambridge, England, 1931), p. 58.

[18] This was previously pointed out by Kugler and Roskies in Ref. 4.

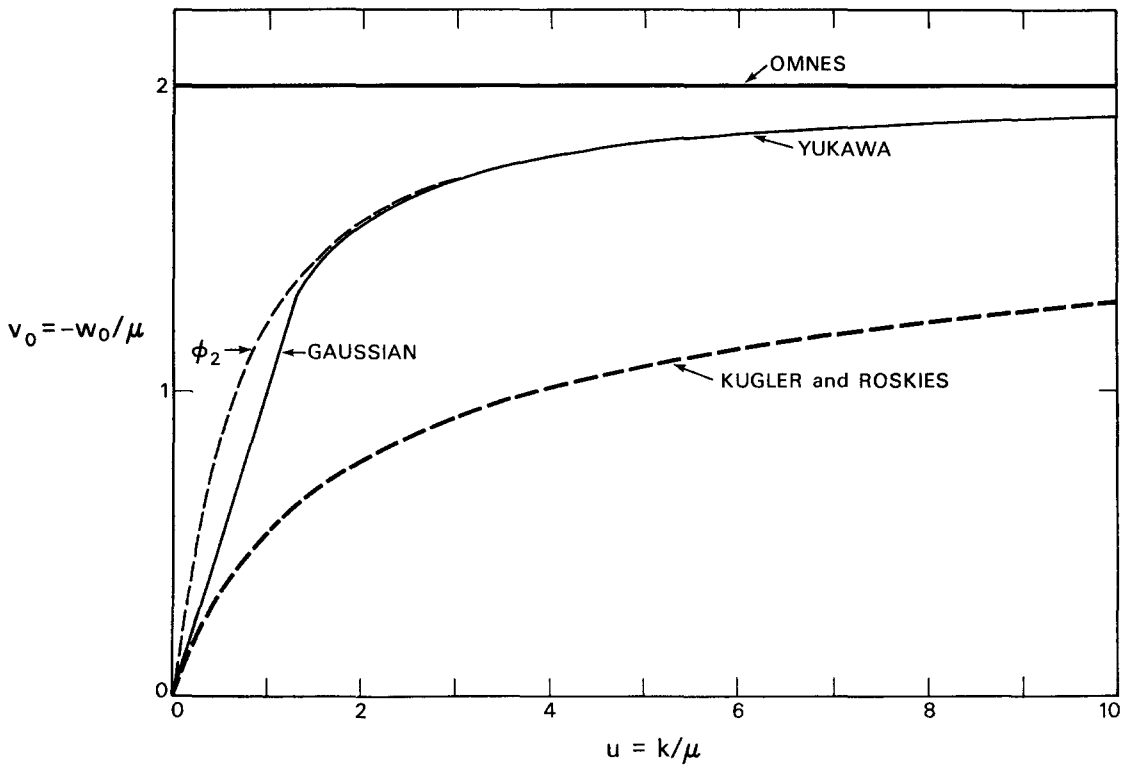

FIG. 2. A comparison of the types of the exponential decrease of various transition probabilities.

(labeled Gaussian). We see that they are extremely similar although for small u, the function obtained from the wavepackets $\psi_2^s(\mathbf{p})$ rises more steeply. The point to be gathered from this example is that the particular type of the decrease of $P(s)$ depends upon the wavepacket for its detail, but the basic features would appear to be somewhat the same for all. Also on Fig. 2 are plotted the functions $v_0(u)$—the type of the decrease in units of μ—which were assumed by Omnes,[3] and by Kugler and Roskies.[4] We see, in particular, that the decrease assumed by Kugler and Roskies is consistent with the actual mode of decrease for the transition probability obtained from the Yukawa potential, although it is noticeably smaller for all finite values of k/μ.

It should be obvious that a choice of a as a constant, independent of k, somehow misses some of the essential points involved. Moreover, any such choice of a does not pick out the maximum ellipse of analyticity for $\mathcal{A}(p, t)$. Another approach would be to take some k-dependent choice of a. However, there do not seem to be any very natural ones. It is in fact worth noting that an assumption that $P(s) = \bar{O}(e^{sw_0(k)})$ does not imply the analyticity of $\int_{p-\epsilon}^{p+\epsilon} dx \mathcal{A}(x, t)$ in the entire region characteristic of the Yukawa interaction. Specifically, let us assume (in the discussion of Sec. 3) that $a = -w_0(k)$ for Gaussian wavepackets of mean momentum k. We then apply our formalism and determine the required region of analyticity. From Eq. (3.7) we have

$$2K(p) \leq p/k + k/p + w_0(k)/p, \quad \text{for all } p. \tag{4.11}$$

Since $K(p)$ is independent of k, we may write

$$2K(p) \leq \inf_{0\leq u\leq\infty} \left(\frac{q}{u} + \frac{u}{q} - \frac{1}{q} \cdot \inf_{0\leq y\leq\infty} \left[\frac{y^2}{u} + u - 2y\, {}_yK(\mu y) \right] \right),$$

where we have used Eqs. (4.8) and (4.10). This implies

$$2[K(p) - {}_yK(p)] \leq \frac{1}{q} \inf_{0\leq u\leq\infty} [v(q, u) - \inf_{0\leq y\leq\infty} v(y, u)]. \tag{4.12}$$

Since for every value of u, $v(q, u) \geq \inf v(y, u)$, $0 < y \leq \infty$, we see that the right-hand side of Eq. (4.12) is nonnegative for every value of q. It can in fact be zero if and only if the infimum in y is taken on at a value $y_0(u) = q$, for a fixed value of q. From the properties of $v(q, u)$, we easily see that $y_0(u)$ is monotonically decreasing, taking on the value $\frac{1}{2}\sqrt{3}$ at $u = \frac{3}{4}\sqrt{3}$, and $y_0(u) = 0$ for $u < \frac{3}{4}\sqrt{3}$. Therefore the open interval $(0, \frac{1}{2}\sqrt{3})$ is not included in the range of $y_0(u)$.

Then, for those values of q, we have that $y_0(u) = 0$ and, therefore,

$$2K(p) \leq \frac{1}{q} \inf_{0 \leq u \leq \infty} \left(\frac{q^2}{u}\right), \quad u \leq \tfrac{3}{4}\sqrt{3}, \quad 0 < q < \tfrac{1}{2}\sqrt{3},$$

while for $q \geq \frac{1}{2}\sqrt{3}$ we find that $2[K(p) - {}_yK(p)] \leq 0$. Therefore we conclude that

$$\begin{aligned} K(p) &\leq {}_yK(p), \qquad p \geq \tfrac{1}{2}(\sqrt{3})\mu, \\ &\leq \left(\frac{2\sqrt{3}}{9}\right)\frac{p}{\mu}, \quad p \leq \tfrac{1}{2}(\sqrt{3})\mu. \end{aligned} \tag{4.13}$$

Inserting this into the equations for the minimum semimajor axis of the ellipse of analyticity of $\mathcal{A}(p, t)$ in the variable t, one obtains

$$\begin{aligned} t_1(p, \epsilon) &\geq \mu^2(4 + \mu^2/p^2), \qquad p \geq \tfrac{1}{2}(\sqrt{3})\mu, \\ &\geq \tfrac{27}{4}\mu^2(1 - 4p^2/27\mu^2)^2, \quad p \leq \tfrac{1}{2}(\sqrt{3})\mu, \end{aligned} \tag{4.14}$$

for sufficiently small $\epsilon > 0$.

We find that the entire region of analyticity for the amplitude corresponding to the Yukawa potential is recovered only for $p \geq \frac{1}{2}(\sqrt{3})\mu$. For smaller values of p, a smaller region of analyticity is obtained—the region which would be obtained (for those values of p) if we assume a constant decrease of the transition probability of value $a = \frac{3}{4}(\sqrt{3})\mu$ [see Eq. (3.13)]. Obviously, then, by this method of attack we will never be able to derive the entire region of analyticity characteristic of the Yukawa interaction.

In a forthcoming paper we indicate another possible approach to the entire question of what kind of decrease assumptions one should make concerning the total transition probability. We show that under those assumptions it is possible to make a choice of a parameter which recovers the entire ellipse of analyticity characteristic of the Yukawa potential, although that choice is still not particularly natural. For quite simple choices, however, regions approximating the Yukawa region may be obtained.

5. CONCLUSION

It was our desire to construct a possible characterization of "short-range" interactions by means of conditions on the total transition probability. We first showed that there exist certain sequences of wavepackets which stay more and more away from the scattering center such that the integral over all times of the probability, at the time t, that the wavepacket is within a sphere of radius R about the scattering center, decreases exponentially. We then assumed that, for such wavepackets, the corresponding sequence of total transition probabilities should be of exponential decrease. This assumption enabled us to show that $\mathcal{A}(p, t)$ [considered as a distribution in p] is analytic in t in a certain ellipse which contains the physical region.

In Fig. 1, a comparison of the results which we have obtained is plotted. The curves are the locations of the rightmost point $[t_1(p)/\mu^2]$ of the ellipse of analyticity of $\mathcal{A}(p, t)$ [considered as a function of t for fixed p] as a function of p/μ. We compare $t_1(p)$ of the type derived using the Gaussian wavepackets [for the choice $a = \frac{3}{4}(\sqrt{3})\mu$] and the similar function $\tilde{t}_1(p)$ for the Yukawa interaction, with the similar results obtained by Omnes, and by Kugler and Roskies.

The above results suffer from the fact that they depend on the type of wavepackets used and on the existence of sufficiently "nice" sequences of wavepackets. In a subsequent paper we indicate some steps which may be taken to remove these difficulties.

ACKNOWLEDGMENT

The author is deeply indebted to Professor E. H. Wichmann for encouragement and guidance during the course of this study.

APPENDIX A: A REGION OF ANALYTICITY FOR SCATTERING AMPLITUDES DERIVABLE FROM A POTENTIAL

We will show the relation between the order of exponential decrease of a given potential $V(r)$ and a region of analyticity of the corresponding elastic scattering amplitude. Let $V(r)$ be a function of r such that the (3-dimensional) Fourier transform exists,

$$\begin{aligned} \tilde{V}(s) &\equiv (2\pi)^{-\frac{3}{2}} \int d\mathbf{r} e^{-i\mathbf{s}\cdot\mathbf{r}} V(r) \\ &= (2\pi)^{-\frac{1}{2}} 2/s \int_0^\infty dr r \sin(rs) V(r). \end{aligned} \tag{A1}$$

We note that the Born approximation to the elastic scattering amplitude is given by (with respect to our normalizations)

$$T'^B(p, z) = -(2\pi)^{-\frac{1}{2}}(m/p)\tilde{V}(\mathbf{p} - \mathbf{q})|_{p=q}. \tag{A2}$$

For fixed nonnegative values of p, we will consider $T'(p, z)$ as a function of z. It follows from some results of Carter[19] that the maximum ellipse of analyticity with foci at $z = \pm 1$ is the same as the corresponding ellipse for the Born approximation term,

[19] See W. Brenig and R. Haag, Fortschr. Physik 17, 183 (1959), where reference is made to Carter's thesis (Princeton University, 1952). Carter shows that when $\int_0^\infty dr r V(r) < \infty$, for large l,

$$T_l(p) = -(\pi/2)\int_0^\infty ds s J^2_{l+\frac{1}{2}}(ps)V(s) + o(1)\cdot\int_0^\infty ds s J^2_{l+\frac{1}{2}}(ps)|V(s)|,$$

where

$$-(\pi/2)\int_0^\infty ds s J^2_{l+\frac{1}{2}}(ps)V(s)$$

is just $T_l^B(p)$.

$T'^{B}(p, z)$. We therefore examine the expression (A2) to determine the location of the singularity of $T'(p, z)$ which is nearest to the physical region.

From Eq. (A1), since $\tilde{V}(s)$ is an even function of the magnitude of s only, we have that

$$\tilde{V}(s) = (2\pi)^{-\frac{1}{2}} i/s \int_{-\infty}^{+\infty} dr r e^{-irs} V(|r|). \tag{A3}$$

We may then use standard knowledge about 1-dimensional Fourier transforms. Let $V(r)$ be of order b. If $b > 1$, then $\tilde{V}(s)$ is an entire function of s, while if $b < 1$, then $\tilde{V}(s)$ is not extensible to an analytic function of s. If $b = 1$, $\tilde{V}(s)$ is an analytic function of s within a strip around the real axis of width 2μ, where $V(r)$ is of type $\mu > 0$. [If $V(r)$ is of order 1 but type 0, then $\tilde{V}(s)$ again is not extensible to an analytic function of s.]

To consider $T'^{B}(p, z)$, we take $\mathbf{s} = \mathbf{p} - \mathbf{q}$ with $p = q$, in Eq. (A2)—this implies $s^2 = 2p^2(1 - z)$. A singularity in s therefore implies *a* singularity in z, for fixed real p. If $b > 1$ we see that $T'(p, z)$ is an entire function of z [since $\tilde{V}(s)$ is even], while if $b < 1$, we will not have an analytic function of z. In the case $b = 1$, for type $\mu > 0$, we have a region of analyticity described by $|\mathrm{Im}\,[2p^2(1 - z)]^{\frac{1}{2}}| < \mu$. For real p, this is the interior of a parabola in the z plane, with vertex at the point $z = 1 + \mu^2/2p^2$, focus at $z = 1$, and symmetric with respect to the real axis. The largest ellipse which will fit in this region has semimajor axis $1 + \mu^2/2p^2$. If there exists r_0 such that, for all $r > r_0$, $V(r)$ is of fixed sign, then the singularity in s will be on the imaginary axis[11] and, therefore, the singularity in z will be on the real axis; and the above ellipse will be the largest ellipse of analyticity, with foci at $z = \pm 1$. From the theorems on Legendre series already discussed, we then have (for the case $b = 1$, $\mu > 0$)

$$\limsup_{l\to\infty} |T_l(p)|^{1/l} \leq {}_{\nu}K(p) = 1 + \mu^2/p^2 - [(1 + \mu^2/2p^2)^2 - 1]^{\frac{1}{2}}.$$

APPENDIX B: THE EXISTENCE OF $G[R, \psi]$

It is convenient to define our measure of closeness $G[R, \psi]$ in terms of the last expression in Eq. (2.10), namely

$$G[R, \psi] = 4\sum_{l=0}^{\infty} \int_0 dp p^4 \frac{dp}{d\omega} f_l(p) \int_0^R dx x^2 j_l^2(xp), \tag{B1}$$

where $f_l(p)$ is given by Eq. (2.7). Therefore,

$$\|\psi\|^2 = \sum_{l=0}^{\infty} \int_0^{\infty} dp p^2 f_l(p) \tag{B2}$$

for every vector in our Hilbert space of wavefunctions.

We will now show that $G[R, \psi]$ is bounded by $C\,\|\psi\|^2$ for every fixed value of $R \geq 0$, where C is a constant depending on R and m. In order to do this, we will need some elementary bounds on the behavior of the spherical Bessel functions. We consider the function

$$g_l(u, \beta) = \frac{(u^2 + \beta^2)}{u^2} \int_0^u dx x^2 j_l^2(x). \tag{B3}$$

Since $(x^2 + u^2)^{-1}$ is a monotonically decreasing function of x, it follows that

$$g_l(u, \beta) < 2(u^2 + \beta^2)^{\frac{1}{2}} \int_0^{\infty} dx \frac{x^2}{x^2 + u^2} j_l^2(x).$$

We use the integral representation[20] for the spherical Bessel function,

$$j_l^2(x) = \tfrac{1}{2} \int_{-1}^{+1} dt P_l(t) \frac{\sin(x[2(1 - t)]^{\frac{1}{2}})}{x[2(1 - t)]^{\frac{1}{2}}} = (2x)^{-1} \int_0^2 dy \sin(xy) P_l(1 - \tfrac{1}{2}y^2).$$

We then have that

$$\begin{aligned} g_l(u, \beta) &< (u^2 + \beta^2)^{\frac{1}{2}} \int_0^{\infty} dx x (x^2 + u^2)^{-1} \\ &\quad \times \int_0^2 dy P_l(1 - \tfrac{1}{2}y^2) \sin(xy) \\ &= \tfrac{1}{2}\pi(u^2 + \beta^2)^{\frac{1}{2}} \int_0^2 dy P_l(1 - \tfrac{1}{2}y^2) e^{-uy} \\ &\leq \tfrac{1}{2}\pi u^{-1}(u^2 + \beta^2)^{\frac{1}{2}}(1 - e^{-2u}) \equiv \hat{g}(u, \beta). \end{aligned}$$

Since $1 - e^{-2u}$ is a monotonically increasing function while $(1 - e^{-2u})/u$ is a monotonically decreasing function, for every $\rho > 0$,

$$\hat{g}(u, \beta) \leq 2\pi(\rho^2 + \beta^2)^{\frac{1}{2}}/2, \quad 0 \leq u \leq \rho,$$
$$\hat{g}(u, \beta) \leq \tfrac{1}{2}\pi(\rho^2 + \beta^2)^{\frac{1}{2}}/\rho, \quad \rho \leq u.$$

The best bound is obtained by taking $\rho = \frac{1}{2}$, which implies

$$g_l(u, \beta) < \hat{g}(u, \beta) \leq \tfrac{1}{2}\pi(1 + 4\beta^2)^{\frac{1}{2}}. \tag{B4}$$

From Eqs. (B1) and (B3),

$$\begin{aligned} G[R, \psi] &\leq 4R \sum_{l=0}^{\infty} \int_0^{\infty} dp p^2 f_l(p) g_l(Rp, Rm) \\ &< 2\pi R[1 + (2mR)^2]^{\frac{1}{2}} \|\psi\|^2. \end{aligned} \tag{B5}$$

[20] In *Handbook of Mathematical Functions* [M. Abramowitz and I. A. Stegun, eds. (National Bureau of Standards, Washington, D.C., 1965), p. 440] it is shown that

$$\frac{\sin R}{R} = \sum_{l=0}^{\infty} (2l + 1) j_l(r) j_l(\rho) P_l(x),$$

where $R^2 = r^2 + \rho^2 - 2r\rho x$, from which our representation follows immediately from the orthogonality of the Legendre functions $P_l(x)$.

It follows from Eq. (B5) that $G[R, \psi]$ determines a bounded operator G_R defined on the entire Hilbert space by

$$(\psi, G_R\psi) = G[R, \psi].$$

If we restrict ourselves to a suitable dense set of the Hilbert space such as the set of all continuous functions which decrease faster than any inverse power of $|\mathbf{p}|$ as $|\mathbf{p}| \to \infty$, then we have [from Eq. (2.7)] that

$$\begin{aligned}
G[R, \psi] &= \frac{1}{\pi}\int_0^\infty dpp^4\frac{dp}{d\omega}\int d\Omega'\int d\Omega\psi^*(\mathbf{p}')\psi(\mathbf{p}) \\
&\quad\times\int_0^R dxx^2\sum_{l=0}^\infty(2l+1)P_l(z)j_l^2(xp) \\
&= 4\int d\mathbf{p}'\int d\mathbf{p}\frac{dp}{d\omega}\delta(p-p')\psi^*(\mathbf{p}')\psi(\mathbf{p}) \\
&\quad\times\int_0^R dxx^2\sum_{l=0}^\infty\sum_{m=-l}^{+l}Y_{lm}^*(\hat{p}')Y_{lm}(\hat{p})j_l^2(xp) \\
&= 4\int d\mathbf{p}'\int d\mathbf{p}\frac{dp}{d\omega}\delta(p-p')\psi^*(\mathbf{p}')\psi(\mathbf{p}) \\
&\quad\times\int_0^R dxx^2\sum_{l'=0}^\infty i^{-l'}\sum_{m'=-l'}^{+l'}\sum_{l=0}^\infty i^{+l}\sum_{m=-l}^{+l}\delta_l^{l'}\delta_m^{m'} \\
&\quad\times Y_{l'm'}^*(\hat{p}')Y_{lm}(\hat{p})j_{l'}(xp')j_l(xp) \\
&= 4\int d\mathbf{p}'\int d\mathbf{p}\delta(\omega-\omega')\psi^*(\mathbf{p}')\psi(\mathbf{p})\int_{|\mathbf{x}|\le R}d\mathbf{x} \\
&\quad\times\sum_{l'=0}^\infty i^{-l'}\sum_{m'=-l'}^{+l'}Y_{l'm'}(\hat{x})Y_{l'm'}^*(\hat{p}')j_{l'}(xp') \\
&\quad\times\sum_{l=0}^\infty i^{+l}\sum_{m=-l}^{+l}Y_{lm}^*(\hat{x})Y_{lm}(\hat{p})j_l(xp) \\
&= (2\pi)^{-2}\int_{|\mathbf{x}|\le R}d\mathbf{x}\int d\mathbf{p}'\int d\mathbf{p}\delta(\omega-\omega') \\
&\quad\times\psi^*(\mathbf{p}')\psi(\mathbf{p})\exp[+i\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')] \\
&= (2\pi)^{-3}\int_{-\infty}^{+\infty}dt\int_{|\mathbf{x}|\le R}d\mathbf{x}\int d\mathbf{p}'\int d\mathbf{p}\psi^*(\mathbf{p}')\psi(\mathbf{p}) \\
&\quad\times\exp[+i\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')-it(\omega-\omega')] \\
&= \int_{-\infty}^{+\infty}dt\int_{|\mathbf{x}|\le R}d\mathbf{x}\,|\psi(\mathbf{x}, t)|^2. \qquad \text{(B6)}
\end{aligned}$$

We therefore see that our definition of $G[R, \psi]$ agrees with the other notion given by Eq. (2.9) [which is physically more appealing], at least on a dense set in the Hilbert space. Although we do not give the details here, we believe the equality (B6) actually holds for the entire space.

APPENDIX C: RIGOROUS PROOF OF THE ASSERTIONS ABOUT ANALYTICITY IN SEC. 3

Remembering Eqs. (3.2) and (3.3), let us write $w = u + iv$, $u > -k$, v real, and define the integrals

$$\tilde{P}_l(w; p_1, p_2) = \int_{p_1}^{p_2}dpA_l(p)h_l(w, p). \qquad \text{(C1)}$$

Whenever $u > -k$, these integrals exist for all l and for all $p_1, p_2 \in [0, \infty]$ and define functions of w analytic for $u > -k$. We also define, for $u > -k$,

$$D(u; p_1, p_2) = \limsup_{l\to\infty}[\tilde{P}_l(u; p_1, p_2)]^{1/l}. \qquad \text{(C2)}$$

Since, for $u > -k$,

$$h_l(u_1, p) \ge |h_l(u_2 + iv, p)|, \quad \text{for} \quad u_2 \ge u_1,$$

it follows that

$$\tilde{P}_l(u_1; p_1, p_2) \ge |\tilde{P}_l(u_2 + iv; p_1, p_2)| \qquad \text{(C3)}$$

and, therefore,

$$D(u_1; p_1, p_2) \ge D(u_2; p_1, p_2), \quad \text{for all} \quad u_2 \ge u_1. \qquad \text{(C4)}$$

Also, since

$$h_l(u, p) \le h_0(u, p)[k/(k+u)]^l, \qquad \text{(C5)}$$

it follows that

$$D(u; p_1, p_2) \le D(u; 0, \infty) \le k(k+u). \qquad \text{(C6)}$$

From the definition of $D(u_1; 0, \infty)$, it follows that the series defining $\tilde{P}(u_1)$ converges for every $u_1 \ge -k$ such that $D(u_1; 0, \infty) < 1$. Because of the inequality (C6) such a u_1 certainly exists. Then, from Eq. (C3), we see that the series converges to a function $\tilde{P}(w)$ analytic in the half-plane Re $(w) > u_1$.

In order to specify more fully the properties of $D(u; 0, \infty)$, let $u_2 > u_1 > -k$ and $p_0 > 2k$. We then have that

$$h_l(u_2, p) \le h_0(u_2, p)(2k/p_0)^{2l}, \quad p \ge p_0,$$

and

$$h_l(u_2, p) \le h_l(u_1, p)\left[\frac{p_0^2 + k(k+u_1)}{p_0^2 + k(k+u_2)}\right]^{2l+2}, \quad p \le p_0.$$

It follows that

$$\tilde{P}_l(u_2; 0, \infty) \le \tilde{P}_l(u_1; 0, p_0)\left[\frac{p_0^2 + k(k+u_1)}{p_0^2 + k(k+u_2)}\right]^{2l+2} + \left(\frac{2k}{p_0}\right)^{2l}\int_0^\infty dph_0(u_2, p),$$

from which

$$D(u_2; 0, \infty) \le \sup\left\{\left(\frac{2k}{p_0}\right)^2; \left[\frac{p_0^2 + k(k+u_1)}{p_0^2 + k(k+u_2)}\right]^2 D(u_1; 0, \infty)\right\}. \qquad \text{(C7)}$$

Since p_0 can be selected arbitrarily large, we conclude that $D(u_1; 0, \infty) > D(u_2; 0, \infty)$ unless $D(u_1; 0, \infty) = 0$, in which case the two are equal. As a result, we may

now uniquely define the number a as follows:

(1) If $D(u; 0, \infty) \geq 1$ for some $u > -k$, then a is the unique solution of the equation $D(-a; 0, \infty) = 1$;

(2) If $D(u; 0, \infty) < 1$ for all $u > -k$, then $a = k$.

It follows, therefore, that for $u > -a$, the series in Eq. (3.2) converges to a function $\tilde{P}(w)$ analytic in $\mathrm{Re}\,(w) > -a$.

We must now show that $\tilde{P}(w)$ is singular at $w = -a$. We first consider the case $a < k$. If $\tilde{P}(w)$ were analytic at $w = -a$, then $\tilde{P}(w)$ would be analytic in the disk $|w - 1| < 1 + a + 2\delta$ for some sufficiently small δ, which we select such that $k > a + \delta > a$. We would then have that

$$\tilde{P}(-a - \delta) = \sum_{n=0}^{\infty} (-1)^n \frac{(1 + a + \delta)^n}{n!} \tilde{P}^{(n)}(1).$$

Since

$$(-1)^n h^{(n)}(1, p) \geq 0,$$

we can invert the order of summation and obtain

$$\begin{aligned}\tilde{P}(-a - \delta) &= \sum_{l=0}^{\infty} \sum_{n=0}^{\infty} \frac{(1 + a + \delta)^n}{n!} \\ &\quad \times \int_0^{\infty} dp A_l(p)(-1)^n h_l^{(n)}(1, p) \\ &= \sum_{l=0}^{\infty} P_l(-a - \delta; 0, \infty) < \infty.\end{aligned}$$

However, since

$$\begin{aligned}\limsup_{l \to \infty} [\tilde{P}_l(-a - \delta; 0, \infty)]^{1/l} &= D(-a - \delta; 0, \infty) \\ &> D(-a; 0, \infty) = 1,\end{aligned}$$

this is an absurdity. To settle the case $a = k$, we note that

$$(-1)^n \tilde{P}^{(n)}(0) \geq (-1)^n P_l^{(n)}(0; 0, \infty) \geq 0$$

for any l. Let us select $l = l_0$ so that Eq. (3.4) holds. Then

$$\tilde{P}_{l_0}(w; 0, \infty) = \int_0^{\infty} dp A_{l_0}(p) \left(\frac{2pk}{p^2 + k(k + w)} \right)^{2l_0 + 2}$$

is analytic for $\mathrm{Re}\,(w) > -k$, but obviously *not* analytic at $w = -k$. It follows that

$$\limsup_{n \to \infty} [(1/n!)\, |\tilde{P}_{l_0}^{(n)}(0; 0, \infty)|]^{1/n} = 1/k$$

and, therefore,

$$\limsup_{n \to \infty} [(1/n!)\, |\tilde{P}^{(n)}(0)|]^{1/n} > 1/k.$$

Since $\tilde{P}(w)$ is analytic for $\mathrm{Re}\,(w) > -k$, we can conclude that the point $w = -k$ must be a singularity.

We should now like to have a simple condition which determines the number a. Since

$$\tilde{P}_l(u; 0, p + \epsilon) = \tilde{P}_l(u; 0, p) + \tilde{P}_l(u; p, p + \epsilon),$$

it follows that

$$D(u; 0, p + \epsilon) = \sup\, [D(u; 0, p); D(u; p, p + \epsilon)], \tag{C8}$$

from which, in particular, $D(u; 0, p)$ is a non-decreasing function of p. For any $p > 0$, we have

$$\begin{aligned}\tilde{P}_l(u; p, p + \epsilon) &\leq \tilde{P}_l(u; p, \infty) \\ &\leq (2k/p)^{2l} \int_0^{\infty} dp' h_0(u, p'),\end{aligned}$$

from which $D(u; p, p + \epsilon) \leq (2k/p)^2$, for any $\epsilon > 0$. Only the case when $D(u; 0, \infty) > 0$ need be considered. It then follows that $D(u; 0, p) = D(u; 0, \infty)$ for all $p > 2k/[D(u; 0, \infty)]^{\frac{1}{2}}$. Let p_0 be the greatest lower bound on all numbers p for which $D(u; 0, p) = D(u; 0, \infty)$. We should note that [for $D(u; 0, \infty) \neq 0$] $p_0 > 0$, since

$$\begin{aligned}D(u; 0, p) &\leq [2/(k + u)]^2 \limsup_{l \to \infty} \left(\int_0^p dq q^{2l+2} \right)^{1/l} \\ &= [2p/(k + u)]^2.\end{aligned}$$

It then follows from Eq. (C8) that

$$D(u; p_0 - \epsilon, p_0 + \epsilon) = D(u; 0, \infty).$$

Hence,

$$\lim_{\epsilon \to 0^+} D(u; p_0 - \epsilon, p_0 + \epsilon) = D(u; 0, \infty).$$

However, for all p and ϵ such that $p - \epsilon > 0$, $D(u; p - \epsilon, p + \epsilon) \leq D(u; 0, \infty)$, from the definition of D. Therefore,

$$D(u; 0, \infty) = \sup_{0 < p < \infty} \left[\lim_{\epsilon \to 0^+} D(u; p - \epsilon, p + \epsilon) \right]. \tag{C9}$$

We then note that

$$\begin{aligned}&\lim_{\epsilon \to 0^+} D(u; p - \epsilon, p + \epsilon) \\ &\quad = \lim_{\epsilon \to 0^+} \limsup_{l \to \infty} \left(\int_{p-\epsilon}^{p+\epsilon} dx A_l(x) h_l(u, x) \right)^{1/l} \\ &\quad = h_0(u, p) \lim_{\epsilon \to 0^+} \limsup_{l \to \infty} \left(\int_{p-\epsilon}^{p+\epsilon} dx A_l(x) \right)^{1/l} \\ &\quad = h_0(u, p) K^2(p),\end{aligned}$$

from which

$$D(u; 0, \infty) = \sup_{0<p<\infty} [2pk/(p^2 + k^2 + ku)]^2 K^2(p), \tag{C10}$$

where $K(p)$ is defined by Eq. (3.6). If we now take into account our previous results on the relation between $D(u; 0, \infty)$ and the half-plane Re $(w) > -a$, in which $\tilde{P}(w)$ is analytic, we obtain the very simple result [stated in Eq. (3.7)],

$$a = -\sup_{0<p<\infty} [2pK(p) - (p^2 + k^2)/k].$$

To show the converse result, we start with the equation obtained from Eqs. (3.1), (3.2), and the standard Laplace transform inversion formula:

$$\begin{aligned}(\pi k^3/16s)^{\frac{1}{2}} P(s) &= 4se^{-sk}\int_0^\infty dpp^2 A_0(p)e^{-sp^2/k} \\ &\quad + (2\pi i)^{-1}\int_{c-i\infty}^{c+i\infty} dwe^{sw}\sum_{l=1}^{\infty}\int_0^\infty dpA_l(p)h_l(w, p),\end{aligned}$$

valid for any $c > -a$. Since

$$\int_0^\infty dpp^2 A_0(p)e^{-sp^2/k} \leq (\pi k^3/16s^3)^{\frac{1}{2}},$$

it suffices to show that the second term is $O(e^{cs})$, $c > -a \geq -k$. We have that

$$\begin{aligned}&\left|\int_{c-i\infty}^{c+i\infty} dwe^{sw}\sum_{l=1}^{\infty}\int_0^\infty dpA_l(p)h_l(w, p)\right| \\ &\quad \leq e^{cs}\sum_{l=1}^{\infty}\int_0^\infty dpA_l(p)h_{l-1}(c, p) \\ &\qquad \times \int_{-\infty}^{+\infty} dx(2pk)^2 |p^2 + k(k + c + ix)|^{-2} \\ &\quad = \pi k^{-1}e^{cs}\sum_{l=1}^{\infty}\int_0^\infty dp(p^2 + k^2 + kc)A_l(p)h_l(c, p) \\ &\quad \leq \pi(17k + c)e^{cs}\sum_{l=1}^{\infty}\int_0^{4k} dpA_l(p)h_l(c, p) \\ &\qquad + \pi k^{-1}e^{cs}\sum_{l=1}^{\infty}\int_{4k}^\infty dph_{l-2}(c, p)\frac{(2pk)^4}{(p^2 + k^2 + kc)^3} \\ &\quad \leq \pi(17k + c)e^{cs}[\tilde{P}(c) - \tilde{P}_0(c)] \\ &\qquad + 2\pi(2k)^3 e^{cs}\int_{4k}^\infty dpp^{-2}\sum_{l=0}^{\infty}(\tfrac{1}{2})^{2l} \\ &\quad \leq \pi(17k + c)e^{cs}\tilde{P}(c) + (16k^2/3)e^{cs}.\end{aligned}$$

Therefore, dividing by $(\pi k^3/16s)^{\frac{1}{2}}$, we have that

$$P(s) \leq 4e^{-sk} + (16s/\pi k^3)^{\frac{1}{2}} \times [\tfrac{1}{2}(17k + c)\tilde{P}(c) + 8k^2/3]e^{cs}, \tag{C11}$$

for every $c > -a \geq -k$; i.e., $P(s) = \bar{O}(e^{-as})$, for all $k \geq a$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Possible Characterization of Short-Range Interactions*

J. D. FINLEY III†
University of California, Berkeley, California

(Received 13 February 1969)

In the preceding paper [J. Math. Phys. **10**, 2047 (1969)], we investigated possible characterizations of "short-range" interactions by means of conditions on the total transition probability $\|T\psi\|^2$ which were valid only for certain types of sequences of wavepackets. In this paper we generalize the assumption on the transition probability so that it will apply to an arbitrary wavepacket. We do this by utilizing a measure $G[R, \psi]$ of how much of a wavepacket ψ is within a sphere of radius R about the scattering center at all times. We then characterize a "short-range" interaction by the existence of a positive constant A and positive functions $B(p)$ and $C(p)$ such that for all $R \geq 0$,

$$\|T\psi\|^2 \leq AG[R, \psi] + \langle C(p)e^{-RB(p)}\rangle_\psi .$$

We then show that for such interactions the square of the scattering amplitude is analytic, in the cosine of the scattering angle, in an ellipse the size of which depends upon $B(p)$. We compare our results with the similar results obtained by the method in the preceding paper and show that these new results much more closely approximate the behavior of the Yukawa interaction.

1. INTRODUCTION

In a previous paper (I) we showed that, if one uses certain special sequences of wavepackets which have the property that they stay more and more away from the scattering center, it is possible to require that the (corresponding) sequence of total transition probabilities decreases exponentially with distance from the scattering center. It was then shown that the existence of certain analyticity properties of the absorptive part of the S matrix is equivalent to this assumption, and that these analyticity properties are similar to some of those the S matrix has in the ordinary Schrödinger theory of "short-range" potentials (those which decrease exponentially or faster with distance). It was therefore suggested that this type of decrease of the total transition probability might be used to characterize "short-range" interactions in a theory which does not lend itself to potentials. However, the fact that the equivalence of the decrease of the transition probability and the analyticity properties of the S matrix could be shown only for certain selected sequences of wavepackets makes this a rather inefficient characterization.

2. DEFINITION OF THE MEASURE OF CLOSENESS

In this paper we show that there exists a generalization of the above notions applicable to arbitrary wavepackets. In order to do this, we use the notion of a measure of closeness introduced in I. It was seen that there was a need for a measure which would determine how close a wavepacket is to the scattering center. Since, in fact, any wavepacket will have some overlap with the scattering center at most finite times. the probability at some time t that the packet is within a sphere of radius R, centered at the scatterer, is a realistic quantity which could be used. Moreover, since the scattering process relates an event at a time in the far past with an event in the far future, all times are pertinent to the scattering event. Therefore we introduce the quantity

$$G[R, \psi] = \int_{-\infty}^{\infty} dt \int_{|\mathbf{x}| \leq R} d\mathbf{x}\, |\psi(\mathbf{x}, t)|^2 \qquad (2.1)$$

as a measure to determine that portion of a wavepacket that is ever "near" the scattering center.

As in I, we consider a scattering situation in which ψ represents the initial asymptotic state and $S\psi$ the final asymptotic state, where S is the scattering matrix. If $iT = S - 1$, then $\|T\psi\|^2$, the square of the norm of the vector $T\psi$, is a measure of the amount of scattering, which we call the total transition probability. In the momentum-space representation we fix the normalization of the S matrix by

$$(T_e\psi)(\mathbf{q}) = \int d\mathbf{p}\delta(p - q)T_e(\mathbf{p}, \mathbf{q})\psi(\mathbf{p}) \qquad (2.2)$$

and

$$\|T\psi\|^2 = \int d\mathbf{p}' \int d\mathbf{p}\psi^*(\mathbf{p}')\psi(\mathbf{p})\delta(p - p')A(p, z), \quad (2.3)$$

where T_e is the elastic portion of the T matrix, $pp'z = \mathbf{p} \cdot \mathbf{p}'$, and $A(p, z)$ is the absorptive part of the S matrix. Expanding $A(p, z)$ in the standard

* Research supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under Grant No. AF-AFOSR-68-1471.

† Present address: University of New Mexico, Albuquerque, New Mexico.

partial-wave expansion, we have

$$A(p, z) = (\pi p^2)^{-1} \sum_{l=0}^{\infty} (2l + 1) P_l(z) A_l(p), \quad (2.4)$$

with $0 \leq A_l(p) \leq 1$. If we also set

$$f_l(p) = \frac{2l+1}{4\pi} \int d\Omega' \int d\Omega \psi^*(\mathbf{p}') \psi(\mathbf{p}) P_l(z)\big|_{p=p'}, \quad (2.5)$$

we have

$$\|T\psi\|^2 = 4 \sum_{l=0}^{\infty} \int_0^{\infty} dp p^2 A_l(p) f_l(p). \quad (2.6)$$

Lastly, we define

$$\mathcal{A}(p, t) \equiv A(p, z(p, t)), \quad \text{where } z = 1 + t/2p^2. \quad (2.7)$$

In an appendix to I, we showed that if we define the measure of closeness as

$$G[R, \psi] \equiv 4 \sum_{l=0}^{\infty} \int_0^{\infty} dp p^4 \frac{dp}{d\omega} f_l^s(p) \int_0^R dx x^2 j_l^2(xp), \quad (2.8)$$

where $j_l(u)$ is the spherical Bessel function of order l, then $G[R, \psi] \leq 2\pi R(1 + 4m^2R^2)^{\frac{1}{2}} \|\psi\|^2$, and, at least on a dense subset of all square-integrable wavepackets, this definition is equivalent to the one given by Eq. (2.1).

3. SHORT-RANGE INTERACTIONS

We surmise, "naively," that a transition probability should consist of a portion near the center and another part describing the interaction of the packet with the "tail" of the interaction. For the "short-range" interactions which occur in hadron physics, this latter part may well be expected to decrease exponentially with distance. Since the notion of "short-range" is not completely defined in the absence of potentials, we can utilize the existence of such a relation as a characterization of a "short-range" interaction. We would like to explore the possibilities of an upper bound on the transition probability, which would hold for all wavepackets. Therefore we assume that for every interaction of "short-range," there exist positive functions $B(p)$ and $C(p)$ and a positive constant A such that, for all wavepackets,

$$\|T\psi\|^2 \leq AG[R, \psi] + \langle C(p) e^{-RB(p)} \rangle_\psi, \quad R \geq 0, \quad (3.1)$$

where the notation $\langle f(\mathbf{p}) \rangle_\psi$ denotes the expectation value of the quantity $f(\mathbf{p})$ in the state determined by the wavefunction $\psi(\mathbf{p})$. Our purpose then is to determine what constraining relations must exist, if any, between the quantities A, $B(p)$, and $C(p)$, and a given interaction of "short-range" in order that Eq. (3.1) may be satisfied for all wavepackets and for all $R \geq 0$. We see immediately that $R = 0$ implies that

$$0 \leq \|T\psi\|^2 \leq \langle C(p) \rangle_\psi.$$

It will be more convenient to study Eq. (3.1) in terms of the partial-wave decomposition. From Eqs. (2.6) and (2.8), we have

$$4 \sum_{l=0}^{\infty} \int_0^{\infty} dp p^2 f_l(p) \left[Ap^2 \frac{dp}{d\omega} \int_0^R dx x^2 j_l^2(xp) + \tfrac{1}{4} C(p) e^{-RB(p)} - A_l(p) \right] \geq 0, \quad (3.2)$$

for all wavefunctions $\psi(\mathbf{p})$ and for all $R \geq 0$. Since $f_l(p) \geq 0$, it will suffice to study the quantity in the bracket, which should be nonnegative for almost all $p \geq 0$ and for all l. Since $A_l(p)$ is independent of R, we write

$$A_l(p) \leq \inf_R \left[Ap^2 \frac{dp}{d\omega} \int_0^R dx x^2 j_l^2(xp) + \tfrac{1}{4} C(p) e^{-RB(p)} \right]. \quad (3.3)$$

Setting $\alpha = Rp/(l + \frac{1}{2})$, we define

$$H_l(\alpha, p) = \frac{A}{p} \frac{dp}{d\omega} (l + \tfrac{1}{2})^3 \int_0^{\alpha} dx x^2 j_l^2[(l + \tfrac{1}{2})x] + \tfrac{1}{4} C(p) e^{-\alpha(l+\frac{1}{2})B(p)/p}, \quad (3.4)$$

from which

$$A_l(p) \leq \inf_{0 \leq \alpha \leq \infty} H_l(\alpha, p), \quad (3.5)$$

for almost all p and all l.

To determine the approximate location of the minimum of $H_l(\alpha, p)$, we first look at only the first term. We have that,[1] for $\alpha \geq 1$,

$$\begin{aligned}(l + \tfrac{1}{2})^3 \int_0^{\alpha} dx x^2 j_l^2[(l + \tfrac{1}{2})x] &\geq (l + \tfrac{1}{2})^3 \int_0^1 dx x^2 j_l^2[(l + \tfrac{1}{2})x] \\ &= \frac{\pi}{4} (l + \tfrac{1}{2})^2 [J'_{l+\frac{1}{2}}(l + \tfrac{1}{2})]^2 \\ &\to \frac{1}{8\pi} 6^{\frac{1}{3}} \Gamma^2(\tfrac{2}{3}) (l + \tfrac{1}{2})^{\frac{2}{3}}, \quad \text{as} \quad l \to \infty.\end{aligned}$$

However, the other term in $H_l(\alpha, p)$ is less than $\frac{1}{4} C(p) \exp[-(l + \frac{1}{2})B(p)/p]$, for $\alpha \geq 1$, which is certainly much smaller for large l. Therefore, if α_l is the value of α for which the minimum of $H_l(\alpha, p)$ occurs, we see that, for very large l, $\alpha_l < 1$. Differentiating $H_l(\alpha, p)$ with respect to α, we find that α_l is the smallest positive solution of the equation

$$4A \frac{dp}{d\omega} (l + \tfrac{1}{2})^2 \alpha^2 j_l^2[(l + \tfrac{1}{2})\alpha] = BC(p) \exp\left(\frac{-\alpha(l + \frac{1}{2})B(p)}{p} \right), \quad (3.6)$$

[1] The spherical Bessel function is such that $j_l(x) = (\pi/2x)^{\frac{1}{2}} J_{l+\frac{1}{2}}(x)$. G. N. Watson, *A Treatise on the Theory of Bessel Functions* (University Press, Cambridge, 1962), 2nd ed., p. 232. Watson shows that $J'_\nu(\nu) = 3^{\frac{1}{3}} \Gamma(\frac{2}{3}) 2^{-\frac{1}{3}} \pi^{-1} \nu^{-\frac{2}{3}} + o(\nu^{-\frac{2}{3}})$.

since $xj_l(x)$ is monotonically increasing for $x \leq l + \frac{1}{2}$. We then desire to determine the properties of $\alpha_l = \alpha_l(p)$ as l becomes arbitrarily large. To do this, we first note some standard properties of Bessel functions. Watson[2] shows that, for $0 \leq z < 1$,

$$J_\nu(\nu z) = (\nu z)^\nu \exp[\nu(1-z^2)^{\frac{1}{2}} - V_\nu] \times \{\exp(\nu)\Gamma(\nu+1)(1-z^2)^{\frac{1}{4}}[1+(1-z^2)^{\frac{1}{2}}]^\nu\}^{-1},$$

where $V_\nu = O(z^2\nu^{-1}(1-z^2)^{-\frac{3}{2}})$. It follows that

$$j_l^2[(l+\tfrac{1}{2})z] = [4(l+\tfrac{1}{2})^2 z(1-z^2)^{\frac{1}{2}}]^{-1}\{z/[1+(1-z^2)^{\frac{1}{2}}]\}^{2l+1} \times \exp[(2l+1)(1-z^2)^{\frac{1}{2}}] \times [1+O(z^2 l^{-1}(1-z^2)^{-\frac{3}{2}})], \quad (3.7)$$

from whence

$$\lim_{l\to\infty}\left[\frac{4A}{BC}\frac{dp}{d\omega}(l+\tfrac{1}{2})^2 z^2 j_l^2[(l+\tfrac{1}{2})z]\right]^{1/(2l+1)} = \frac{z\exp(1-z^2)^{\frac{1}{2}}}{1+(1-z^2)^{\frac{1}{2}}} \equiv h(z),\ 0 \leq z < 1.$$

Therefore, since $\alpha_l < 1$ for large l, it follows from Eq. (3.6) that $v \equiv \lim_{l\to\infty} \alpha_l$ is the unique solution of the equation

$$\exp[-vB(p)/2p] = h(v), \quad 0 \leq v \leq 1; \quad (3.8)$$

i.e., for $0 \leq v \leq 1$, $v = v[B(p)/2p]$ is uniquely determined by the above equation, and $v = \lim \alpha_l$ as $l \to \infty$. Hence, v is *independent* of A, and of $C(p)$, when $A > 0$, $C(p) > 0$. We note that $h(v)$ is a monotonically increasing, convex function of v such that $h(v) \geq v$. In addition, $h(0) = 0$, and $h(1) = 1$.

We may now consider $H_l(\alpha_l, p)$ and determine the restrictions which it places on the partial-wave amplitudes $A_l(p)$ by way of Eq. (3.3). We first note that

$$H_l(\alpha_l, p) > \tfrac{1}{4}C(p)\exp[-(l+\tfrac{1}{2})\alpha_l B(p)/p] > 0.$$

Since $0 \leq A_l(p) \leq 1$, we can always pick $C(p)$ so that the inequality (3.3) is satisfied for any finite range of values of l. We need therefore only to determine the form of $H_l(\alpha_l, p)$ as l becomes arbitrarily large. We use

$$\int_0^x dy f(y) e^{2lg(y)} = \frac{f(x)e^{2lg(x)}}{2lg'(x)}\left[1+O\left(\frac{1}{l}\right)\right],$$

for sufficiently smooth functions f and g, where g is monotonically increasing. Via Eq. (3.7) we see that

$$\int_0^{\alpha_l} dx x^2 j_l^2[(l+\tfrac{1}{2})x] = (1-\alpha_l^2)^{-1}\alpha_l^2(2l+1)^{-3}[h(\alpha_l)]^{2l+1} \times [1+O(l^{-1}\alpha_l^2(1-\alpha_l^2)^{-\frac{3}{2}})].$$

[2] G. N. Watson, Ref. 1, p. 227.

Therefore, using Eq. (3.6), we have that

$$H_l(\alpha_l, p) = \tfrac{1}{4}C(p)\exp[-(l+\tfrac{1}{2})\alpha_l B(p)/p] \times [1+\alpha_l(1-\alpha_l^2)^{-\frac{1}{2}}B(p)/2p] \times [1+O(l^{-1}\alpha_l^2(1-\alpha_l^2)^{-\frac{3}{2}})]. \quad (3.9)$$

We see that the quantities $H_l(\alpha_l, p)$ decrease approximately like an lth power of a number less than unity, for sufficiently large l. Therefore, in order for the inequality (3.3) to hold, the $A_l(p)$ must be bounded by an lth power of a number less than unity. [We recall that this is indeed the case for the $A_l(p)$ derivable from the Yukawa potential; in fact, it is shown in I that this is true for all $A_l(p)$ derived from a potential $V(r) = O(e^{-\mu r})$, $\mu > 0$.] More precisely, if the $A_l(p)$ are continuous,

$$K(p) \equiv \lim_{l\to\infty}[A_l(p)]^{1/2l} \leq \lim_{l\to\infty}[H_l(\alpha_l, p)]^{1/2l} = h(v); \quad (3.10)$$

i.e., the requirement on the interaction made by the inequality (3.1) is $K(p) \leq h\{v[B(p)/2p]\}$, for all p,[3] under the assumption that, for a given interaction, there exist quantities A, $B(p)$, and $C(p)$ as specified above.

We now proceed to show that such functions do actually exist, for a given interaction of "exponential range." Let $A_l(p)$ be specific given functions for a specific interaction ($0 \leq A_l(p) \leq 1$) such that the corresponding $K(p)$ exists and is less than unity, for finite nonnegative values of p. We may then define a function $\bar{B}(p)$ [using Eqs. (3.8) and (3.10)] by the equation

$$K(p) = h\{v[\bar{B}(p)/2p]\}. \quad (3.11)$$

Now let $B_\epsilon(p) \equiv (1-\epsilon)\bar{B}(p)$. Since $h\{v[B(p)/2p]\}$ is a monotonically decreasing function of $B(p)/2p$ (see the discussion of this function in the Appendix), it follows that $K(p) < h\{v[B_\epsilon(p)/2p]\}$ and that, for sufficiently large values of l, $A_l(p) < H_l(\alpha_l, p)$, where this α_l is the one corresponding to $B_\epsilon(p)$. Since we are still free to choose the function $C(p)$, we choose it so that the inequality (3.3) is satisfied for "small" values of l, which is obviously always possible, since for any given p there is only a finite range of values of l for which this needs to be done. We note that any reasonable choice of the constant A will do, such as m.

From the above we see that the function $\bar{B}(p)/2p$ is just the least upper bound of all the functions $B_\epsilon(p)/2p$ which satisfy the inequality (3.1) for a given interaction and some proper choice of $C(p)$ and A. Let us set

[3] If the $A_l(p)$ are *not* continuous, then $A_l(p)$ is to be replaced by the function

$$\tilde{A}_l(p) = \lim_{\epsilon\to 0}(2\epsilon)^{-1}\int_{p-\epsilon}^{p+\epsilon} dx A_l(x).$$

$\beta = \beta(K)$ such that $\beta[K(p)] = \bar{B}(p)/2p$. We may then study $\beta = \beta(K)$ as given by Eqs. (3.8) and (3.11), namely

$$e^{-v\beta} = h(v) = K,$$

which implies

$$\beta = v^{-1} \log K^{-1}. \tag{3.12}$$

We note that (see the Appendix) the main features of $\beta = \beta(K)$ are that it is a monotonically decreasing function of K such that $\lim \beta(K) = \infty$ as $K \to 0$, while $\beta(1) = 0$. If we are given a specific interaction, we will then know $K = K(p)$, which would allow us to determine $\beta[K(p)]$—we could then determine $\bar{B} = \bar{B}(p)$. A particular $K(p)$ which is of considerable interest is the one corresponding to a Yukawa potential. In I we show that for every scattering amplitude derivable from a potential function $V(r) = O(e^{-\mu r})$, $\mu > 0$, where $V(r)$ is of constant sign for all $r > r_0 \geq 0$, the corresponding function $K(p) = {}_yK(p)$,

$${}_yK(p) = z_0 - (z_0^2 - 1)^{\frac{1}{2}}, \quad \text{with} \quad z_0 = 1 + \mu^2/2p^2, \tag{3.13}$$

which is the function derived from the Yukawa potential (see I). There is also good reason to believe, from Lagrangian field theory, that every interaction does indeed approach that derivable from a Yukawa potential at very small values of the momentum. Therefore, using the equations in the Appendix and setting ${}_yB(p) = 2p\beta[{}_yK(p)]$, we find that ${}_yB(p)$ is a monotonically decreasing function of p with the following asymptotic behavior:

$${}_yB(p)/\mu = 2e(\mu/p) \log (\mu/p)[1 + O(p^2/\mu^2)],$$

for p near 0, and

$${}_yB(p)/\mu = 2 + (3\mu/p)^{\frac{2}{3}} + O(\mu/p), \quad \text{for} \quad p \to \infty.$$

The function ${}_yB(p)/\mu$ is plotted as a function of p/μ in Fig. 1.

It is now our intent to use the upper bound on $\|T\psi\|^2$, given by Eq. (3.1), for a realistic theory of the interactions of elementary particles, in which there are no potentials from which to derive the interactions. We can use the inequality (3.1) to characterize an interaction of "exponential range," or "short range." We would like to have a universal function $B(p)$ such that, for proper choice of A and $C(p)$, any interaction of "exponential range" would satisfy Eq. (3.1). Since we know that for any interaction such that $K(p) \leq 1$, there does indeed exist such a function $B(p)$, dependent on the choice of the interaction, we merely need to know if these functions B have an upper bound. Since we believe that the interaction should be similar to the Yukawa interaction for small values of the momentum, it would follow that in the range of small p we must require $B(p) \leq {}_yB(p)$.

Since we suppose that our interaction decreases like $O(e^{-\mu r})$, a possible choice is $B(p) = 2\mu < {}_yB(p)$. This choice is quite simple and natural and agrees quite well with ${}_yB(p)$ for very large momenta. However, since ${}_yB(p) \to \infty$ as $p \to 0$, this choice disagrees quite

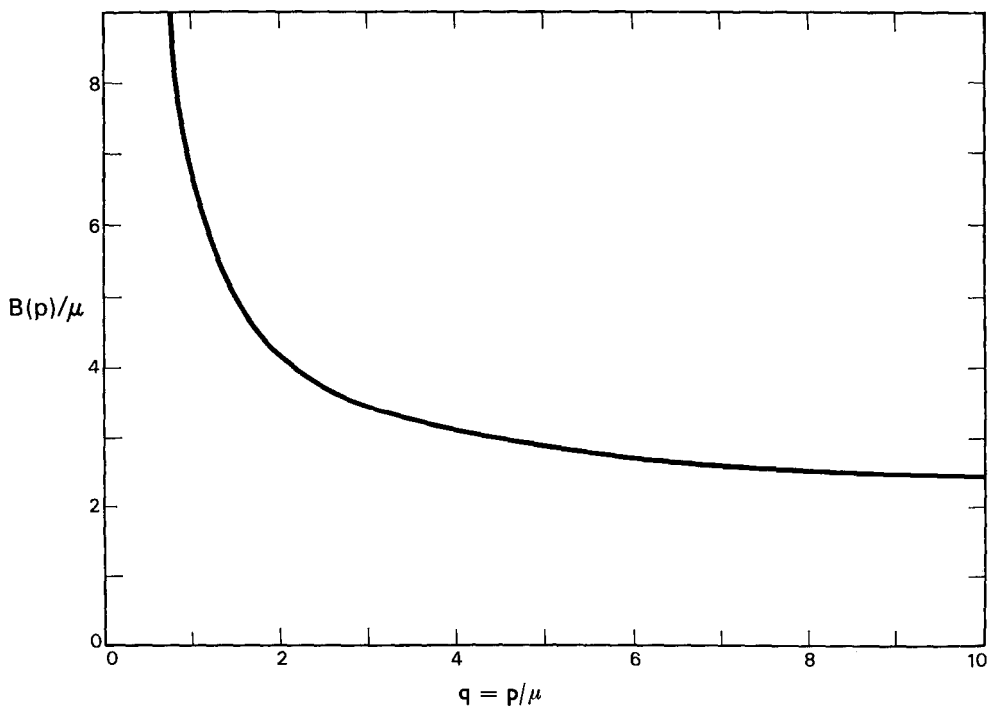

FIG. 1. The function ${}_yB(p)/\mu$ for the Yukawa potential.

strongly with the result from the Yukawa potential at small momenta—and it is just at small momenta that we expect the interaction should surely agree with the Yukawa interaction.

4. ANALYTIC PROPERTIES OF THE SCATTERING AMPLITUDE

We will first determine the region of analyticity for $\mathcal{A}(p, t)$ implied by the choice $B(p) = 2\mu$, and will then return to the questions of the behavior of $B(p)$ for very small momenta. We may therefore suppose that, for all interactions which "decrease like $O(e^{-\mu r})$,"

$$\|T\psi\|^2 \leq AG[R, \psi] + e^{-2\mu R}\langle C(p)\rangle_\psi. \quad (4.1)$$

It then follows [from Eqs. (3.8) and (3.10)] that for every such interaction,

$$K(p) \leq e^{-v\mu/p} = h(v). \quad (4.2)$$

In I we showed that $\mathcal{A}(p, t)$ is analytic in t for every fixed $p \geq 0$ in an ellipse with foci at $-4p^2$ and 0, and rightmost point $t_1 \geq \tilde{t}_1$, where

$$\tilde{t}_1(p) = [p(h - h^{-1})]^2 = [2p \sinh (v\mu/p)]^2, \quad (4.3)$$

where, as in Eq. (4.2), v is determined by $e^{-v\mu/p} = h(v)$. The fundamental feature of $\tilde{t}_1(p)$ is that it is a monotonically increasing function of p such that

$$\lim_{p\to\infty} \tilde{t}_1(p) = 4\mu^2.$$

Approximate formulas for $\tilde{t}_1(p)$ are given in the Appendix. The function $\tilde{t}_1(p)/\mu^2$ determined by Eq. (4.3) is plotted versus p/μ in Fig. 2 (labeled $B = 2\mu$). We see that $\tilde{t}_1$ always lies below the corresponding curve for the Yukawa potential, while it approaches it for very large values of p.

On the other hand the above curve for $\tilde{t}_1$, determined by the condition $B(p) = 2\mu$, gives altogether too small a region of analyticity for very small momenta. The reason for this has already been mentioned. Let us investigate the possibilities of another choice for $B(p)$ which would overcome this difficulty. From Eq. (3.13) we see that a simple function, which would approximate ${}_\nu B(p)$ for both large and small momenta and still remain everywhere smaller, would be $B(p) = 2\mu(1 + \mu/p)$. Repeating the above analysis for this case, we find that the corresponding $\tilde{t}_1(p)$ is a smooth function of p with one minimum such that $\lim \tilde{t}_1(p) = 4\mu^2$ as $p \to \infty$, as before, while for very small momenta (see the Appendix) we have that

$$\tilde{t}_1(p)/\mu^2 = [(ep/2\mu) \log (ep^2/2\mu^2)]^{-2}[1 + o(p)], \quad \text{for } p \text{ near zero.}$$

This function $\tilde{t}_1(p)/\mu^2$ is plotted versus p/μ in Fig. 2 [labeled $B = 2\mu(1 + 1/q)$]. We see therefore that by simulating the behavior of the actual Yukawa function ${}_\nu B(p)$ [i.e., by causing our choice of $B(p)$ to become infinite as $p \to 0$], we can derive a region of analyticity which more closely approximates that of the actual Yukawa interaction.

It should be noted at this point that, if we actually assume $B(p) = {}_\nu B(p)$, we will acquire the full region of analyticity of the Yukawa interaction. That is, if we assume that a characterization of an interaction of "exponential range" is that for all $R \geq 0$,

$$\|T\psi\|^2 \leq AG[R, \psi] + \langle C(p)e^{-R_\nu B(p)}\rangle_\psi,$$

then it follows that all such interactions are such that the squares of their amplitudes $\mathcal{A}(p, t)$ are analytic in an ellipse at least as large as the ellipse corresponding to the Yukawa interaction. We should recall that this was not the case in the study in I, where only specific sequences of wavepackets were used. In that case, we showed that, if we actually assumed that $P(s) = \bar{O}(e^{sw_0(k)})$ for all appropriate interactions, then this assumption only implied that the corresponding quantities $\mathcal{A}(p, t)$ were analytic (as distributions) in an ellipse, which for small momenta was much smaller than the ellipse of analyticity of the Yukawa interaction. In addition, because of the extreme "smoothness" of the Gaussian wavepackets, we were only able to show that for every fixed p and every $\epsilon > 0$, $\int_{p-\epsilon}^{p+\epsilon} dx\mathcal{A}(x, t)$ was an analytic function of t in a certain ellipse. However, in the present case, because our results are true for arbitrary wavefunctions, we have that, at least, $\tilde{\mathcal{A}}(p, t)$ is an analytic function for every fixed p, where $\tilde{\mathcal{A}}(p, t)$ is the quantity constructed from the functions

$$\tilde{A}_i(p) \equiv \lim_{\epsilon\to 0} \tfrac{1}{2}\epsilon^{-1}\int_{p-\epsilon}^{p+\epsilon} dxA_i(x).$$

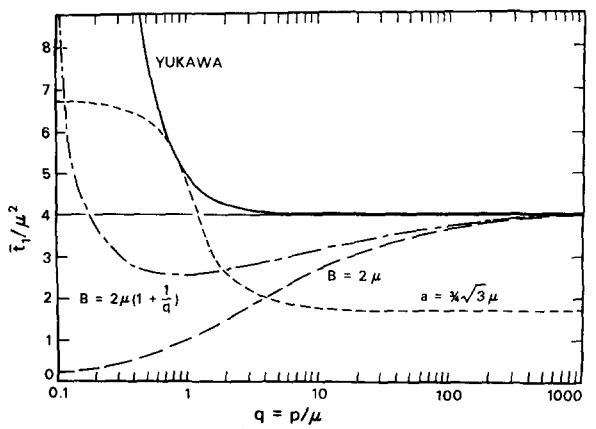

FIG. 2. The rightmost points of various ellipses of analyticity of $\mathcal{A}(p, t)$.

[See Ref. 3. We note that the functions $\tilde{A}_l(p)$ differ from the functions $A_l(p)$ at most on a set of measure zero.] In particular, if the functions $A_l(p)$ are continuous, then $\tilde{A}_l(p) = A_l(p)$, and so $\mathcal{A}(p, t)$ is itself analytic in t, for every fixed p. In the method using only specific sequences of wavefunctions, we showed that there were continuous functions $A_l(p)$ satisfying the necessary requirements, but such that only the corresponding function $\int_{p-\epsilon}^{p+\epsilon} dx\mathcal{A}(x, t)$ was analytic in t—i.e., not $\mathcal{A}(p, t)$ itself.

5. CONCLUSIONS

It was our desire to construct a possible characterization of "short-range" interactions by means of conditions on the total transition probability. In this paper, we generalized the notions in I to arbitrary wavepackets. We found that it is indeed possible to characterize "short-range" interactions by the requirement that, for every such interaction, there exist a constant $A > 0$ and positive functions $B(p)$ and $C(p)$ such that

$$\|T\psi\|^2 \leq AG[R, \psi] + \langle C(p)e^{-RB(p)}\rangle_\psi, \quad \text{for all } R \geq 0.$$

It is not totally clear how to pick the function $B(p) \geq 0$; however, the function ${}_1B(p)$, obtained by considering the Yukawa potential, should serve as a guide, especially for small momenta. This method gives information about $\mathcal{A}(p, t)$ (almost everywhere in p), rather than merely as a distribution in p.

In Fig. 2 is plotted a comparison of the results which we have obtained. The curves are the locations of the rightmost point $[\tilde{t}_1(p)/\mu^2]$ of the ellipse of analyticity of $\mathcal{A}(p, t)$ (considered as a function of t for fixed p) as a function of p/μ. We compare $\tilde{t}_1(p)$ of the type derived with the Gaussian wavepackets (for the choice $a = \frac{3}{4}(\sqrt{3})\mu$), $\tilde{t}_1(p)$ derived with the choice of $B = 2\mu$, and $\tilde{t}_1(p)$ derived with the choice of $B = 2\mu(1 + \mu/p)$, with the similar function $t_1(p)$ for the Yukawa interaction.

We would also like to say that the particular form of the inequality [Eq. (3.1)] which we have chosen to bound the total transition probability is not necessarily the best, but it has merits of simplicity. It could be pointed out, however, that the form of the function $\bar{B}(p)$ is unchanged if, instead of $G[R, \psi]$, we use some similar integral—for instance,

$$\int_{-\infty}^{+\infty} dt \int_{|\mathbf{x}| \leq R} d\mathbf{x} f(x)\, |\psi(\mathbf{x}, t)|^2,$$

where $f(x)$ is an everywhere-positive weighting function satisfying some restrictions on its decrease (or growth), and such that the integral still continues to exist. In particular, if $f(x)$ is bounded above and decreases no faster than some fixed inverse power of x (for $x > 0$)—$f(x)$ should not be singular at $x = 0$—then $f(x)$ would be suitable as a weighting function in the above integral. We have considerable freedom in our choice of the positive function $C(p)$.

It should also be pointed out that it is desirable to broaden the type of bound above so that it would include information concerning general displacements—timelike, for instance—of the wavefunctions as well as spacelike ones. (This might be done, for instance, by considering a function $G_t[R, \psi]$ where the time integration is performed only up to some fixed time t, rather than infinity.) Such types of bounds could presumably be used to obtain information about the analytic behavior of the scattering amplitude in all its variables simultaneously.

ACKNOWLEDGMENT

The author is deeply indebted to Professor E. H. Wichmann for encouragement and guidance during the course of this study.

APPENDIX: EVALUATION OF $\tilde{t}_1(p)$ FOR A YUKAWA POTENTIAL

We first study

$$h(v) = [v \exp(1 - v^2)^{\frac{1}{2}}]/[1 + (1 - v^2)^{\frac{1}{2}}]. \quad \text{(A1)}$$

We see immediately that $h \in [0, 1]$ for $v \in [0, 1]$, $h(0) = 0$, $h(1) = 1$, and $h(v) \geq v$. Furthermore,

$$\frac{dh}{dv} \geq 0.$$

One easily obtains the following formulas:

$$h(v) = \tfrac{1}{2}ev[1 - \tfrac{1}{4}v^2 + O(v^6)], \quad v \text{ near } 0, \quad \text{(A2)}$$

which implies

$$v = 2(h/e)[1 + (h/e)^2 + O(h^4)], \quad h \text{ near } 0, \quad \text{(A3)}$$

$$h(v) = 1 - \tfrac{1}{3}(1 - v^2)^{\frac{3}{2}} - \tfrac{1}{5}(1 - v^2)^{\frac{5}{2}} + O((1 - v^2)^3), \quad v \text{ near } 1, \quad \text{(A4)}$$

which implies

$$v = 1 - \tfrac{1}{2}[3(1 - h)]^{\frac{2}{3}} + O((1 - h)^{\frac{4}{3}}), \quad h \text{ near } 1. \quad \text{(A5)}$$

We may then use these formulas to study $\beta = \beta(K)$. From Eq. (3.12), we have that

$$\beta v = \log K^{-1} = \log h^{-1}. \quad \text{(A6)}$$

Therefore, from Eq. (A3), we have that

$$\beta = \tfrac{1}{2}(e/K)(\log K^{-1})[1 - (K/e)^2 + O(K^4)], \quad K \text{ near } 0. \quad \text{(A7)}$$

Similarly, from Eq. (A5), we have that

$$\beta = (1 - K)\{1 + \tfrac{1}{2}[3(1 - K)]^{\frac{2}{3}} + O(1 - K)\}, \quad K \text{ near } 1. \quad \text{(A8)}$$

Since

$$\frac{dh}{dv} \geq 0,$$

we see that

$$\frac{d\beta}{dK} < 0.$$

Therefore, $\beta = \beta(K)$ is a monotonically decreasing function such that $\beta(1) = 0$ and $\lim \beta(K) = \infty$ as $K \to 0$.

Let us now consider the case where K is taken to be ${}_vK$ [the $K(p)$ corresponding to the Yukawa potential]. It follows from Eq. (3.13) that

$${}_vK(p) = 1 - \mu/p + O(\mu^2/p^2), \quad \text{for} \quad p \to \infty,$$

while

$${}_vK(p) = p^2/\mu^2 + O(p^4/\mu^4), \quad \text{for } p \text{ near } 0.$$

Since ${}_vB(p) = 2p\beta[{}_vK(p)]$, we see that for the Yukawa potential

$${}_vB(p) = 2e(\mu^2/p)[\log(\mu/p)][1 + O(p^2/\mu^2)], \quad \text{for } p \text{ near } 0, \quad \text{(A9)}$$

and

$${}_vB(p) = 2\mu[1 + \tfrac{1}{2}(3\mu/p)^{\frac{2}{3}} + O(\mu/p)], \quad \text{for} \quad p \to \infty. \quad \text{(A10)}$$

For a given choice of $B(p)$, we may compute K and the corresponding $\tilde{t}_1$.

From Eqs. (2.7) and (3.13), we have that

$$\tilde{t}_1 = [p(K^{-1} - K)]^2.$$

Then, from Eq. (A6), we have that

$$\tilde{t}_1 = 4p^2 \sinh^2 v\beta = 4p^2 \sinh^2 (vB/2p). \quad \text{(A11)}$$

If we choose $B = 2\mu$, we have

$$\tilde{t}_1 = 4p^2 \sinh^2 (v\mu/p).$$

Since v is bounded,

$$\tilde{t}_1 = (2\mu v)^2[1 + O(\mu^2/p^2)], \quad \text{as} \quad p \to \infty.$$

From Eqs. (A4) and (A8), we have

$$v = 1 - \tfrac{1}{2}(3\beta)^{\frac{2}{3}} + O(\beta^{\frac{4}{3}}), \quad \beta \text{ near } 0. \quad \text{(A12)}$$

Therefore,

$$\tilde{t}_1 = (2\mu)^2[1 - (3\mu/p)^{\frac{2}{3}} + O(\mu/p)], \quad \text{as} \quad p \to \infty. \quad \text{(A13)}$$

Again, for small p, Eqs. (A2) and (A7) imply

$$v = \frac{1}{\beta}[\log(2\beta/e)]\left\{1 - \frac{\log[\log(2\beta/e)]}{\log(2\beta)} + O\left(\frac{\log\beta}{\beta}\right)\right\}, \quad \text{(A14)}$$

which implies

$$\tilde{t}_1 = [(e/2\mu)\log(ep/2\mu)]^{-2}[1 + o(p)], \quad \text{for} \quad p \text{ near } 0. \quad \text{(A15)}$$

Lastly, if we choose $B = 2\mu(1 + \mu/p)$, we have, from Eqs. (A11), (A12), and (A14),

$$\tilde{t}_1 = (2\mu)^2[1 - (3\mu/p)^{\frac{2}{3}} + O(\mu/p)], \quad \text{as} \quad p \to \infty, \quad \text{(A16)}$$

and

$$\tilde{t}_1 = \mu^2[(ep/2\mu)\log(ep^2/2\mu^2)]^{-2}[1 + o(p)], \quad \text{for} \quad p \text{ near } 0. \quad \text{(A17)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

V-Particle Decay in the Lee Model*

C. L. HAMMER AND T. A. WEBER
Institute for Atomic Research and Department of Physics
Iowa State University, Ames, Iowa

(Received 13 January 1969)

The formalism developed previously for scattering of wavepackets is applied to the $N + \theta$ sector of the Lee model. A single analysis suffices to discuss both the stable and unstable case. The rate at which N and θ particles are produced and the number of V particles produced are calculated as a function of time assuming that the incident wave is initially a semi-infinite, plane-wave train. An unstable V-particle state is constructed from the $|N, \theta\rangle$ state by requiring agreement with the scattering analysis. Its dependence on the production process is explicitly shown. The state can be made independent of the production process by requiring normalization. The time dependence of each channel is also calculated for this case. It is shown that both the unstable and stable V-particle states can be generated from the mathematical V-particle state, the only difference being the location of the pole which describes the resonance state or the bound state.

I. INTRODUCTION

The Lee model, which can be solved exactly for many of its sectors, is close enough to reality so that its investigation gives insight into real physical systems. The $N + \theta$ sector is of interest because not only can the N and θ particles have a bound state (the V particle) which leads to a discussion of mass and coupling constant renormalization, but they also can have resonant scattering states, which leads to a discussion of unstable states.

The bound-state problem is discussed extensively in several text books[1] and is well understood. The unstable V particle has been discussed by several authors who construct the V-particle wavefunction using somewhat different approaches and who reach somewhat diverse conclusions. Glaser and Källen[2] construct the unstable V-particle wavefunction from the stable V-particle wavefunction by replacing the bound-state energy with the complex energy $E + i\lambda$. They point out that this differs from an exact solution by terms which are of the order of the half-width of the state. To obtain the expected exponential decay, the limit $\lambda \to \infty$ is taken. There seems to be no justification for this procedure, especially since this limit reduces the state to the mathematical V-particle state.

Araki *et al.*[3] use an S-matrix approach to show that an exponential decay results when the mathematical V-particle state is used as the unstable state. Except for the fact that an exponential decay is obtained, it is not clear whether or not this is the proper choice since it is not apparent that this state is produced from the scattering of the N and θ particles—the only mechanism for the production of the unstable V particle in the Lee model. In addition, complex normalization constants are used which cause difficulties with the Hermitian properties of the renormalized fields.

Levy[4] constructs a general wavefunction using arbitrary coefficients for the mathematical states. These coefficients are presumed to be determined by the production process. He shows that simple choices for the coefficients, none of which correspond to those used by the authors previously cited, lead to as pure an exponential decay as is desired. Again no attempt is made to relate these coefficients to the scattering process. Levy also discusses the ambiguities that arise in the renormalization process for the unstable cases.

In view of the fact that the Lee model is one of the most simple field-theoretical models that involve unstable particles, it is clear that the "state of the art" regarding these particles is, at best, uncertain.

In a recent paper,[5] the authors presented a wavepacket formalism which extends the usual S-matrix approach to scattering theory to include the case of decaying states. The bound state and the time dependence of the scattering state can be calculated in a simple, straightforward way. This formalism is applied in this paper to the $N + \theta$ sector in an attempt to clarify the analysis of the unstable V particle. The number of V particles produced as a function of time is calculated from the observed number of final-state N and θ particles and shown to agree with the number obtained by projecting the $|N, \theta\rangle$ state onto the mathematical

* Work was performed in the Ames Laboratory of the U.S. Atomic Energy Commission, contribution No. 2477.

[1] See for example, S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson & Co., New York, 1961), Chap. 12, Sec. 12b, p. 352; C. Källen, *Lectures in Theoretical Physics, 1961: Brandeis Summer Institute*, M. E. Rose and E. C. G. Sudarshan, Eds. (W. A. Benjamin, Inc., New York, 1962). For more recent references see also L. M. Scarfone, J. Math. Phys. **9**, 346 (1968), and M. M. Broido, J. Math. Phys. **9**, 510 (1968).

[2] V. Glasser and C. Källen, Nucl. Phys. **2**, 706 (1956–57).

[3] H. Araki, Y. Munakata, M. Kawaguchi, and T. Goto, Progr. Theoret. Phys. (Kyoto) **17**, 419 (1957), Sec. 18b.

[4] M. M. Levy, Nuovo Cimento **13**, 115 (1959); **14**, 612 (1959).

[5] T. A. Weber and C. L. Hammer, USAEC Report IS-1795, 1969.

V-particle state. It is therefore clear that the part of the $|N, \theta\rangle$ state which is not the incident wave is the unstable V-particle state. Since this term also contains the amplitude of the incident wave, the dependence on the production process is immediately apparent. For the initial conditions being those assumed in the scattering problem, the expected time dependence is obtained. If it is assumed that the existence of the V-particle state has been determined by measurement after it has been produced, so that the state can be normalized to unity thereby making it independent of the production process, then the state reduces initially to the "dressed" mathematical V-particle state, which, apart from the renormalization factor, justifies the choice of Araki *et al.*[3] for the unstable V particle. If a bound-state singularity is assumed, rather than a pole in the second sheet, then the stable V particle is obtained.

II. REVIEW

The basic idea developed previously,[5] and which will be used in this analysis, is that, given an initial condition

$$|\psi(0)\rangle = \int_{-\infty}^{\infty} dp A(p)\,|p\rangle, \tag{1}$$

the solution to the Schrödinger equation

$$H\,|\psi(t)\rangle = i(\partial/\partial t)\,|\psi(t)\rangle \tag{2}$$

can be written as

$$|\psi(t)\rangle = \int_{\mathrm{C}} dp A(p)\,|\psi_+\rangle \exp(-iEt), \tag{3}$$

where $|\psi_+\rangle$ is the stationary state solution

$$\begin{aligned} |\psi_+\rangle &= [1 + (E - H + i\eta)^{-1}H']\,|p\rangle, \\ H\,|\psi_+\rangle &= E\,|\psi_+\rangle, \end{aligned} \tag{4}$$

and C is a contour in the complex p plane which is chosen so that

$$\int_{\mathrm{C}} dp A(p)(E - H + i\eta)^{-1}H'\,|p\rangle = 0. \tag{5}$$

Here it is assumed that

$$H = H_0 + H' \tag{6}$$

and that $|p\rangle$ is the eigenstate

$$H_0\,|p\rangle = E\,|p\rangle. \tag{7}$$

As a consequence of Eq. (5), for $t = 0$, $|\psi(t)\rangle$ as expressed in Eq. (3) reduces to $|\psi(0)\rangle$ so that the initial conditions are automatically satisfied.

The Hamiltonian for the Lee model can be expressed as

$$H_0 = M_V \int d\mathbf{p}\psi_V^\dagger(\mathbf{p})\psi_V(\mathbf{p}) + M_N \int d\mathbf{p}\psi_N^\dagger(\mathbf{p})\psi_N(\mathbf{p}) + \tfrac{1}{2}\int d\mathbf{k} a^\dagger(\mathbf{k})a(\mathbf{k}); \tag{8}$$

$$H' = -g_0(2\pi)^{-\frac{3}{2}} M_V M_N \int \frac{d\mathbf{p}}{E_V} \int \frac{d\mathbf{k}}{2\omega E_N} f(\omega) \times [\psi_V^\dagger(\mathbf{p})\psi_N(\mathbf{p} - \mathbf{k})a(\mathbf{k}) + a^\dagger(\mathbf{k})\psi_N^\dagger(\mathbf{p} - \mathbf{k})\psi_V(\mathbf{p})], \tag{9}$$

where the commutation rules are

$$\begin{aligned} [a(\mathbf{k}), a^\dagger(\mathbf{q})] &= 2\omega\delta(\mathbf{k} - \mathbf{q}), \\ \{\psi_V(\mathbf{p}), \psi_V^\dagger(\mathbf{q})\} &= (E_V/M_V)\delta(\mathbf{p} - \mathbf{q}), \\ \{\psi_N(\mathbf{p}), \psi_N^\dagger(\mathbf{p})\} &= (E_N/M_N)\delta(\mathbf{p} - \mathbf{q}), \end{aligned} \tag{10}$$

and

$$\omega^2 = k^2 + \mu^2.$$

The function $f(\omega)$ is the usual "cutoff" function; $f(\omega) \sim 0$ for large ω, and g_0 is a real coupling constant. As is customary in Lee-model calculations, because of the sharp cutoff due to $f(\omega)$, the recoil of the heavy V and N particles is ignored. Thus E_V and E_N will be replaced by M_V and M_N.

The eigenstates of H will be written as $|V, N, \theta\rangle$, corresponding to the three particles of the Lee model, and the eigenstates of H_0 will be written as

$$|p, p - k, k\rangle,$$

corresponding to a V particle of momentum $\mathbf{p}$, an N particle of momentum $\mathbf{p} - \mathbf{k}$, and a θ particle of momentum $\mathbf{k}$. With the normalization used in the preceding equations, it follows that the configuration representation for the state vectors of H_0 is

$$\langle x_1, x_2, 0\,|\,0, p, k\rangle = u_N(\mathbf{p}, \mathbf{x}_2)u_\theta(\mathbf{k}, \mathbf{x}_1), \tag{11}$$

where

$$u(\mathbf{q}, \mathbf{y}) = (2\pi)^{-\frac{3}{2}} \exp(i\mathbf{q}\cdot\mathbf{y}). \tag{12}$$

III. THE N, θ STATE

The solution for the N, θ sector of the Lee model is well known and has been reproduced in many places.[1] However, to use the contour integration technique described in the previous section and to discuss decaying states, it is more convenient to rederive the solution by using the idea of the level shift operator. The stable V particle and the N, θ scattering state can then be discussed from a single point of view.

The starting point is the Lippman–Schwinger form

$$|0, N, \theta\rangle = [1 + (E - H + i\eta)^{-1}H'](2\omega)^{-\frac{1}{2}} \times |0, p - k, k\rangle \tag{13}$$

$$= \{1 + (E - H_0 + i\eta)^{-1} \times [H' + H'(E - H + i\eta)^{-1}H']\}(2\omega)^{-\frac{1}{2}} \times |0, p - k, k\rangle, \tag{14}$$

where $E = M_N + \omega$ and where the factor $(2\omega)^{-\frac{1}{2}}$ is required by the normalization

$$\langle p' - k', k', 0 \,|\, 0, p - k, k\rangle = 2\omega\delta(\mathbf{p}' - \mathbf{k}' - \mathbf{p} + \mathbf{k})\delta(\mathbf{k}' - \mathbf{k}).$$

With the help of the relation

$$H' \,|0, p - k, k\rangle = -(2\pi)^{-\frac{3}{2}} g_0 f(\omega) \,|p, 0, 0\rangle, \quad (15)$$

this equation reduces to

$$(2\omega)^{\frac{1}{2}} \,|0, N, \theta\rangle = |0, p - k, k\rangle - g_0(2\pi)^{-\frac{3}{2}} f(\omega)[(E - M_V + i\eta)^{-1} + (E - H_0 + i\eta)^{-1} H'(E - H + i\eta)^{-1}] \,|p, 0, 0\rangle. \quad (16)$$

The product $H'(E - H + i\eta)^{-1}$ can be written in terms of the level-shift operator R as in Goldberger and Watson[6]:

$$H'(E - H + i\eta)^{-1} \,|p, 0, 0\rangle = \frac{R \,|p, 0, 0\rangle}{E - M_V - \langle 0, 0, p| \,(R) \,|p, 0, 0\rangle + i\eta}, \quad (17)$$

where R is defined as

$$R \,|p, 0, 0\rangle = [H' + H' P_V (E - H_0 - P_V H' P_V)^{-1} P_V H'] \,|p, 0, 0\rangle. \quad (18)$$

P_V is the projection operator

$$P_V = 1 - \int d\mathbf{p} \,|p, 0, 0\rangle \langle 0, 0, p| \quad (19)$$

which projects out the V particle, and the reduced matrix element is defined in general by

$$\langle k, p' - k, p'| \,\mathcal{O} \,|p, p - k, k\rangle = \delta(\mathbf{p}' - \mathbf{p}) \,\langle k, p - k, p| \,(\mathcal{O}) \,|p, p - k, k\rangle.$$

For the N, θ sector, the only other state is $|0, p - k, k\rangle$ so that

$$P_V = \int d\mathbf{p} \frac{d\mathbf{k}}{2\omega} \,|0, p - k, k\rangle \langle k, p - k, 0|. \quad (20)$$

To complete the solution the matrix elements $\langle k, p - k, 0| \,(R) \,|p, 0, 0\rangle$ and $\langle 0, 0, p| \,(R) \,|p, 0, 0\rangle$ must be evaluated. The first follows directly from Eq. (18) and the commutation relations. This gives

$$\langle k, p - k, 0| \,(R) \,|p, 0, 0\rangle = -(2\pi)^{-\frac{3}{2}} g_0 f(\omega). \quad (21)$$

The second follows from the dispersion relation

$$\begin{aligned} R_p(E) &\equiv \langle 0, 0, p| \,(R) \,|p, 0, 0\rangle \\ &= D(E) - iI(E) \qquad (22) \\ &= \langle 0, 0, p| \,(H) \,|p, 0, 0\rangle - \pi^{-1} \int_\mu^\infty \frac{d\omega \, I(M_N + \omega)}{\omega + M_N - E - i\eta} \qquad (23) \end{aligned}$$

where, as is shown in Goldberger and Watson,[6]

$$I(E) = \lim_{\eta \to 0} \eta \int \frac{d\mathbf{k}}{2\omega} \frac{|\langle k, p - k, 0| \,(R) \,|p, 0, 0\rangle|^2}{(E - M_N - \omega)^2 + \eta^2}, \quad E \geq M_N + \omega, \quad (24a)$$

$$I(E) = 0, \quad E < M_N + \omega,$$

and

$$D(E) = \langle 0, 0, p| \,(H') \,|p, 0, 0\rangle - \pi^{-1} P \int_\mu^\infty d\omega \frac{I(M_N + \omega)}{\omega + M_N - E}. \quad (24b)$$

From Eq. (21) it then follows that

$$I(E) = (4\pi)^{-1} g_0^2 k f^2(\omega), \quad E \geq M_M + \omega,$$
$$I(E) = 0, \quad E < M_N + \omega. \quad (25)$$

Since the matrix element of H' in Eq. (23) vanishes, Eq. (25) leads to the result

$$R_p(E) = -(2\pi)^{-2} g_0^2 \int_\mu^\infty d\omega \frac{k f^2(\omega)}{\omega + M_N - E - i\eta} \quad (26a)$$

or, for the continuation into the second sheet,

$$R_p(E) = -(2\pi)^{-2} g_0^2 \int_C d\omega \frac{k f^2(\omega)}{\omega + M_N - E}, \quad (26b)$$

where C extends from μ to ∞ going below the pole in the integrand. Except for a sign, the function $F(E - M_N)$ defined by Källen[1] is expressed by Eq. (26a).

By interposing a complete set of states to the left of R_p in Eq. (16) and using Eqs. (21) and (22), the N, θ solution is found to be

$$(2\omega)^{\frac{1}{2}} \,|0, N, \theta\rangle = |0, p - k, k\rangle - \frac{(2\pi)^{-\frac{3}{2}} g_0 f(\omega)}{E - M_V - R_p(E) + i\eta} \Big[|p, 0, 0\rangle - (2\pi)^{-\frac{3}{2}} g_0 \times \int \frac{(d\bar{\mathbf{k}}/2\bar{\omega}) f(\bar{\omega})}{E - M_N - \bar{\omega} + i\eta} \,|0, p - \bar{k}, \bar{k}\rangle \Big], \quad (27)$$

where $\bar{\omega} = \omega(\bar{k})$.

[6] Marvin L. Goldberger and Kenneth M. Watson, *Collision Theory* (John Wiley & Sons, Inc., New York, 1964), Chap. 8, pp. 437–452.

IV. THE STABLE V PARTICLE

It is readily seen from Eq. (26) that $R_p(E)$ is real if $E < \mu + M_N$. Recalling that $E = \omega + M_N$, this implies that $\omega < \mu$ or that k is imaginary for these values. This corresponds to a bound-state pole in the state vector $|0, N, \theta\rangle$ if $E = M_V + R_p(E)$ in this region. If one assumes this is the case, then, since Eq. (27) represents a solution for all k, it follows that the contour integral

$$|V, 0, 0\rangle = \oint_C dk B(k) |0, N, \theta\rangle \tag{28}$$

is also a solution, where C is a contour, arbitrarily small, encircling the pole. If E_V is the value of E at the pole,

$$E_V = M_V + R_p(E_V). \tag{29}$$

Evaluation of Eq. (28) gives

$$|V, 0, 0\rangle = Z^{\frac{1}{2}}(E_V) |p, 0, 0\rangle - (2\pi)^{-\frac{3}{2}} g \times \int \frac{(d\bar{\mathbf{k}}/2\bar{\omega}) f(\bar{\omega}) |0, p - \bar{k}, \bar{k}\rangle}{E_V - M_N - \bar{\omega} + i\eta}, \tag{30}$$

where $Z(E_V)$ is the normalization constant

$$Z^{-1}(E_V) = 1 + (2\pi)^{-2} g_0^2 \int_\mu^\infty \frac{d\omega k f^2(\omega)}{(E_V - M_N - \omega)^2} = 1 - \frac{dR_p(E_V)}{dE_V} \tag{31}$$

and g is the renormalized coupling constant

$$g^2 = Z(E_V) g_0^2. \tag{32}$$

This is the usual stable V-particle solution[1] with $R_p(E_V)$ corresponding to the mass renormalization term

$$\delta m = E_V - M_V = -Z^{-1}(E_V) g^2 (2\pi)^{-2} \int_\mu^\infty \frac{d\omega k f^2(\omega)}{\omega + M_N - E_V}, \tag{33}$$

as can be seen from Eqs. (26) and (32).

V. N, θ RESONANT SCATTERING

If the resonance has a narrow width, then, at the appropriate energies, an unstable state will be formed which will subsequently decay into an N particle and an θ particle. The probability that those particles are observed at large distances from the scattering region should depend upon time according to the usual decay rule $[1 - \exp(-\Gamma t')]$, where t' is the time at which the unstable state decays. To show that this is indeed the case, it is appropriate to use the configuration representation.

If $\mathbf{x}_2$ and $\mathbf{x}_1$ are the position vectors of the θ and N particles, respectively, then the probability that these particles are at these positions can be obtained, according to Eq. (11), from the amplitude

$$\varphi(\mathbf{x}_2, \mathbf{x}_1) \equiv \langle x_2, x_1, 0 \mid 0, N, \theta\rangle = u_N(\mathbf{p} - \mathbf{k}, \mathbf{x}_1) u_\theta(\mathbf{k}, \mathbf{x}_2)(2\omega)^{-\frac{1}{2}} + \frac{(2\pi)^{-3}(2\omega)^{-\frac{1}{2}} g_0^2 f(\omega)}{E - M_V - R_p(E) + i\eta} \times \int \frac{(d\bar{\mathbf{k}}/2\bar{\omega}) f(\bar{\omega}) u_N(\mathbf{p} - \bar{\mathbf{k}}, \mathbf{x}_1) u_\theta(\bar{\mathbf{k}}, \mathbf{x}_2)}{E - M_N - \bar{\omega} + i\eta}. \tag{34}$$

The algebra is greatly simplified without loss of physical content by choosing the wavepacket for the N particle considerably smaller in spatial extent than the wavepacket of the θ particle. With this approximation, $\mathbf{x}_1$ can be taken equal to zero so that the N particle is "nailed down" at the origin. After completing the angular integration, the scattering solution becomes

$$\varphi(\mathbf{x}_2, 0) = (2\pi)^{-\frac{3}{2}}(2\omega)^{-\frac{1}{2}} \times \left[u_\theta(\mathbf{k}, \mathbf{x}_2) + \frac{(2\pi)^{-\frac{7}{2}} g_0^2 f(\omega)(ix_2)^{-1}}{E - M_V - R_p(E) + i\eta} \times \int_{-\infty}^{\infty} \frac{(d\bar{k}/2\bar{\omega}) \bar{k} f(\bar{\omega}) \exp(i\bar{k}x_2)}{E - M_N - \bar{\omega} + i\eta} \right], \tag{35}$$

where $x_2 = |\mathbf{x}_2|$. The solution as a function of time is

$$\psi(\mathbf{x}_2, t) = \int_C d\mathbf{k} a(\mathbf{k}) \varphi(\mathbf{x}_2, 0) \exp(-iEt), \tag{36}$$

where C must be chosen to eliminate the second term of Eq. (35) when $t = 0$.

The contour which will accomplish this purpose is shown in Fig. 1 along with the singularities of $\varphi(\mathbf{x}_2, 0)$. The discussion which follows is based on the assumption that the θ particle is nonrelativistic, as is appropriate since recoil of the heavy particles is ignored. The resonant and bound-state singularities, assumed here to be simple poles, arise from the vanishing of the

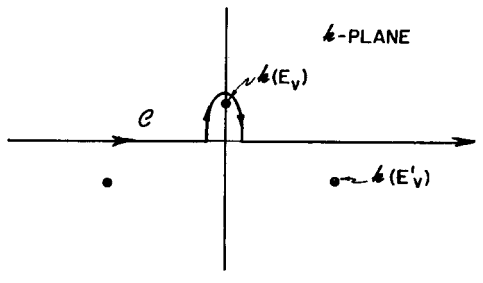

FIG. 1. The contour C and the singularities of $\varphi(\mathbf{x}_2, 0)$ in the k plane.

denominator $E - M_V - R_p(E)$. By convention, $E + i\eta$ is chosen on the first sheet of $R_p(E)$ and hence on the first sheet of $\varphi(\mathbf{x}_2, 0)$. This corresponds to values of k in the upper k plane. If $E = M_V + R_p(E)$ for values (real E) $> M_N + \omega$, it is easily shown that $\varphi(\mathbf{x}_2, 0)$ has singularities, corresponding to the resonance singularities, in the second sheet at

$$E = M_V + D(E) \pm iI(E)$$

or in the third and fourth quadrant of the k plane as shown in Fig. 1. Except for these singularities, these functions are analytic. The contour must also exclude any singularities of $f(\omega)$ which are in the upper k plane. Also, to eliminate the second term of $\varphi(\mathbf{x}_2, 0)$ at $t = 0$ requires the assumption that $f(\omega)$ has appropriate properties along the infinite arc. It should be noted that the quantity η in Eq. (36) can be set equal to zero with the understanding that E corresponds to values of k in the upper k plane.

For the scattering solution, only terms proportional to $(1/x_2)$ need be retained. Therefore only the contribution from the simple pole at $\bar{k} = k + i\eta$ in the integrand of Eq. (35) need be considered. For large x_2, $\varphi(\mathbf{x}_2, 0)$ then is

$$\varphi(\mathbf{x}_2, 0)_{x_2\to\infty} = (2\pi)^{-3}(2\omega)^{-\frac{1}{2}} \times \left[\exp(i\mathbf{k}\cdot\mathbf{x}_2) - \frac{(4\pi x_2)^{-1} g_0^2 f^2(\omega)\exp(ikx_2)}{E - M_V - R_p(E) + i\eta}\right], \tag{37}$$

or, in terms of spherical harmonics,

$$\varphi(\mathbf{x}_2, 0)_{x_2\to\infty} = 2(2\pi)^{-2}(2\omega)^{-\frac{1}{2}}\Sigma_{lm} Y^*_{lm}(\hat{k}) Y_{lm}(\hat{x}_2) i^l \times \left[j_l(kx_2) - \frac{\delta_{l0} g_0^2 f^2(\omega)(4\pi x_2)^{-1}\exp(ikx_2)}{E - M_V - R_p(E) + i\eta}\right]. \tag{38}$$

It is readily shown[5] that the initial wavepacket can be expanded in terms of radial functions according to

$$\psi(\mathbf{x}_2, 0) = \Sigma_{lm}\int_{-\infty}^{\infty} dk k^2 a_{lm}(\mathbf{k}, \mathbf{k}_0, \mathbf{x}_0) j_l(kx_2) Y_{lm}(\hat{x}_2), \tag{39}$$

where $\mathbf{x}_0$ and $\mathbf{k}_0$ are the average position and momentum of the packet at $t = 0$. If the over-all factors of Eq. (38) are absorbed in the coefficient a_{lm}, the scattering solution at any time t is

$$\psi(\mathbf{x}_2, t)_{x_2\to\infty} = \Sigma_{lm}\int_{-\infty}^{\infty} k^2\, dk a_{lm}(\mathbf{k}, \mathbf{k}_0, \mathbf{x}_0) j_l(kx_2) \times Y_{lm}(\hat{x}_2)\exp(-iEt) - g_0^2(4\pi x_2)^{-1}\Sigma_{lm}\delta_{l,0} \times \int_C k^2\, dk f^2(\omega) a_{lm}(\mathbf{k}, \mathbf{k}_0, \mathbf{x}_0)\frac{Y_{lm}(\hat{x}_2)\exp i(kx_2 - Et)}{E - M_V - R_p(E)}. \tag{40}$$

Therefore the scattered packet is given by

$$\psi_s(\mathbf{x}_2, t) = -g_0^2(4\pi)^{-\frac{3}{2}} x_2^{-1} \times \int_C \frac{k^2\, dk f^2(\omega) a_{00}\exp i(kx_2 - Et)}{E - M_V - R_p(E)}. \tag{41}$$

The denominator in the integrand vanishes for $R_p(E)$ real only if there are bound states and the singularities are on the imaginary k axis. If this is the case, the mathematical mass M_V, which is not a measurable quantity, can be expressed as a function of the bound-state energy through Eqs. (26) and (29):

$$E - M_V - R_p(E) = E - E_V + R_p(E_V) - R_p(E), = (E - E_V)Z^{-1}(E), \tag{42}$$

where

$$Z^{-1}(E) = 1 + g_0^2(2\pi)^{-2} \times \int_\mu^\infty \frac{d\bar{\omega}\bar{k} f^2(\bar{\omega})}{(\bar{\omega} + M_N - E_V)(\bar{\omega} + M_N - E - i\eta)}. \tag{43}$$

Note that $Z(E_V)$ is the residue of the denominator in Eq. (41) which accounts for the second equation given in Eq. (31).

In terms of renormalized quantities, the scattered packet is

$$\psi_s(\mathbf{x}_2, t) = -(4\pi)^{-\frac{3}{2}} g^2 x_2^{-1} \times \int_C \frac{k^2\, dk f^2(\omega) a_{00}}{E - E_V}[Z(E)/Z(E_V)]\exp i(kx_2 - Et). \tag{44}$$

Because of the ratio $Z(E)/Z(E_V)$, Eq. (44) is finite in the point-interaction limit, $f(\omega) \to 1$.[7]

The denominator in the integrand of Eq. (41) vanishes for $R_p(E)$ complex only if there are resonant states. If the singularity is a simple pole, the denominator vanishes when

$$E'_V = M_V + R_p(E'_V), \tag{45}$$

where E'_V is the complex eigenvalue of H,[8]

$$E'_V = E_R - i(\tfrac{1}{2}\Gamma); \tag{46}$$

E_R is the center of the resonance; $\frac{1}{2}\Gamma$ is the observed half-width; and the definition in Eq. (26b) must be used to define $R_p(E'_V)$. In parallel to the bound-state case, Eq. (45) can be used to eliminate M_V in terms of the measurable quantities E_R and Γ. Thus,

$$E - M_V - R_p(E) = Z'^{-1}(E)(E - E'_V), \tag{47}$$

[7] F. J. Yndurain, J. Math. Phys. 7, 1133 (1966). In this paper it is shown that if the Lee model is analyzed with relativistic kinematics, the point-coupling limit is better behaved.

[8] If in Eq. (28) the contour encircles the pole at E'_V, the state that is obtained is a solution of H but it is not normalizable.

where here E_V' replaces E_V in Eq. (43) giving

$$Z'^{-1}(E) = 1 + g_0^2(2\pi)^{-2} \times \int_C \frac{d\omega k f^2(\omega)}{(\omega + M_N - E_V')(\omega + M_N - E)}, \quad (48)$$

where C extends from $\omega = \mu$ to $\omega = \infty$ and goes below both simple poles in the integrand. The scattered solution now becomes

$$\psi_s(\mathbf{x}_2, t) = -(4\pi)^{-\frac{3}{2}} g^2 x_2^{-1} \exp[i\delta(E_V')] \times \int_C \frac{dk k^2 f^2(\omega) a_{00}}{E - E_V'} [Z'(E)/Z'(E_V')] \exp i(kx_2 - Et), \quad (49)$$

where

$$g^2 = g_0^2 |Z'(E_V')| \quad (50)$$

and δ is defined by

$$Z'(E) = |Z'(E)| \exp i\delta(E). \quad (51)$$

Because the scattering solution is independent of the phase δ, the measured coupling constant will depend only on the absolute value as defined in Eq. (50). As was previously the case for $Z(E_V)$, $Z'(E_V')$ is the residue of the denominator in Eq. (41):

$$Z'(E_V') = 1 - \frac{dR_p(E_V')}{dE_V'}.$$

Thus, as $E_V' \to E_V$, it follows that $Z'(E_V') \to Z(E_V)$, as is to be expected if a single analysis is to describe both the stable and unstable cases. The solution given in Eq. (49) is simply the analytic continuation of the stable solution given in Eq. (44).

Equations (45) and (47) can be used to show that $I(E)$ is related to the half-width by the relation

$$I(E) = \tfrac{1}{2}\Gamma |Z'(E)|^{-1} \cos \delta(E). \quad (52)$$

For scattering in the vicinity of a very narrow resonance, it is usually assumed that $I(E)$ is slowly varying and that it can be replaced by

$$I(E_R) = \tfrac{1}{2}\Gamma_0,$$

the unrenormalized half-width. In this limit, to lowest order in g_0^2,

$$\tfrac{1}{2}\Gamma = \tfrac{1}{2}\Gamma_0 |Z'(E_V')|, \quad (53)$$

as is to be expected. To show the exponential decay of the scattering solution, it is necessary to consider Eq. (49) in this limit. As a result, $\psi_s(\mathbf{x}_2, t)$ becomes

$$\psi_s(\mathbf{x}_2, t) = -(4\pi)^{-\frac{3}{2}} g^2 x_2^{-1} \times \int_C \frac{dk k^2 f(\omega) a_{00} \exp i(kx_2 - Et)}{E - E_R + \frac{1}{2}i\Gamma} + O(\mu\Gamma/k_R^2). \quad (54)$$

The simplest incident wave that satisfies the assumptions already made with regard to spatial extent is a semi-infinite one-dimensional wavetrain

$$\psi(\mathbf{x}_2, 0) = \Theta(\hat{\mathbf{x}}_0 \cdot \mathbf{x}_2 - x_0) \exp[ik_0(x_0 - \hat{\mathbf{x}}_0 \cdot \mathbf{x}_2)]. \quad (55)$$

To establish this function as the initial condition for the scattering solution given in Eq. (35) requires that

$$\psi(\mathbf{x}_2, 0) = \int_C d\mathbf{k} A(\mathbf{k}) \langle x_2, x_1, 0 \mid 0, N, \theta \rangle = \int d\mathbf{k} (2\omega)^{-\frac{1}{2}} A(\mathbf{k}) \langle x_2, x_1, 0 \mid 0, p - k, k \rangle. \quad (56)$$

For $\mathbf{x}_1 = 0$ and $\mathbf{x}_0$ chosen in the $-\hat{z}$ direction,

$$A(\mathbf{k}) = \frac{i(2\pi)^2 (2\omega)^{\frac{1}{2}} \delta(k_x)\delta(k_y) \exp(ik_z x_0)}{k_z - k_0 + i\epsilon}, \quad (57)$$

or, after integration over k_x and k_y,

$$\psi(\mathbf{x}_2, 0) = -(2\pi i)^{-1} \int_{-\infty}^{\infty} dk (k - k_0 + i\epsilon)^{-1} \times \exp[ik(x_0 - \hat{\mathbf{x}}_0 \cdot \mathbf{x}_2)] = 2i\Sigma_{lm}(-i)^l Y_{lm}^*(\hat{x}_0) Y_{lm}(\hat{x}_2) \times \int_{-\infty}^{\infty} \frac{dk j_l(kx_2) \exp(ikx_0)}{k - k_0 + i\epsilon}. \quad (58)$$

Identification with Eq. (39) shows that

$$a_{00}(\mathbf{k}, \mathbf{k}_0, \mathbf{x}_0) = 2i(4\pi)^{-\frac{1}{2}} k^{-2} (k - k_0 + i\epsilon)^{-1} \exp(ikx_0). \quad (59)$$

The resonant scattering solution then takes the form

$$\psi_s(\mathbf{x}_2, t) = -4\mu i x_2^{-\frac{1}{2}} g^2 (4\pi)^{-2} \times \int_C \frac{dk f^2(\omega) \exp i[k(x_2 + x_0) - Et]}{(k - k_1)(k - k_2)(k - k_0 + i\epsilon)},$$

where k_1 and k_2 are defined by

$$(\tfrac{1}{2}\mu)(k - k_1)(k - k_2) = (k^2/2\mu) + M_N + \mu - E_R + \tfrac{1}{2}i\Gamma.$$

This integral can now be evaluated using the asymptotic integration procedure developed previously.[9] The contour C can be distorted from the real k axis to the contour shown in Fig. 2. The 45° line represents the line of steepest descent, giving negligible contribution to the original integral. Consequently, the only contribution arises from the encircled poles at k_0 and k_1.

[9] T. A. Weber, D. M. Fradkin, and C. L. Hammer, Ann. Phys. (N.Y.) 27, 362 (1964); C. L. Hammer and T. A. Weber, J. Math. Phys. 6, 1591 (1965).

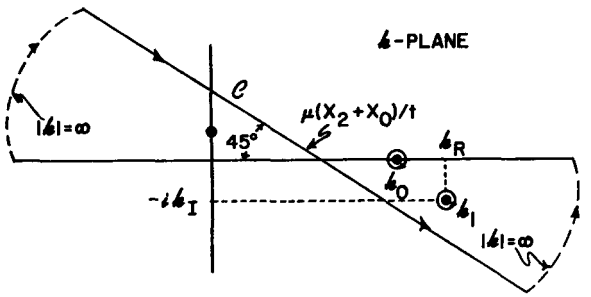

FIG. 2. The distorted contour in the k plane.

The result is

$$\psi_s(x_2, t) = -\frac{(4\pi x_2)^{-\frac{1}{2}} g^2 f^2(\omega_0)}{E_0 - E_R + \frac{1}{2}i\Gamma} \times \{\exp i[k_0(x_2 + x_0) - E_0 t]\Theta((k_0/\mu)t - x_2 - x_0) - \exp[(-\tfrac{1}{2}\Gamma(t' - t_0)] \exp i[k_R(x_2 + x_0) - E_R t] \times \Theta((k_R/\mu)t - x_2 - x_0)\} + x_2^{-\frac{1}{2}} O(\mu\Gamma/k_R^2), \quad (60)$$

where the retarded time t' is given by

$$t' = t - \mu x_2/k_R \quad (61)$$

and

$$\omega_0 = \omega(k_0), \quad E_0 = E(k_0),$$
$$t_0 = (\mu x_0/k_R) \cong (\mu x_0/k_0). \quad (62)$$

In deriving this equation use is made of the relation $|k_0 - k_R| \sim \mu\Gamma/k_R$, which is required because of the assumptions made in arriving at Eq. (53). The asymptotic procedure used to solve the integral in Eq. (54) does not strictly apply when the 45° contour is very close to the poles. Thus k_0 should be replaced by k_R in the argument of the first step function to be consistent with the approximation already made.

The second term of Eq. (60),[10] which clearly represents the decay of the unstable V particle, implicitly contains the scattering distance $x_2 + x_0$ in the damping exponential. To be consistant with the scattering solution, $\exp -\frac{1}{2}\Gamma(t' - t_0) \gg x_2^{-1}$. Substitution for t in this expression from the argument of the appropriate step function gives the largest value of the exponential (i.e., the smallest value for t) as

$$\exp -[(\mu^2\Gamma^2/4k_R^3)(x_2 + x_0)] \gg x_2^{-1}. \quad (63)$$

Rearranging this expression gives

$$(2\mu\Gamma/k_R^2)^2 \ll 16[k_R(x_2 + x_0)]^{-1} \ln x_2 \quad (64)$$

as a measure of the validity of ψ_s. For larger values of Γ, the second term of Eq. (60) should be dropped.

The scattering cross section, which can be calculated from the incident and scattered currents in the usual way, is, for $t > t_0$,

$$\frac{d\sigma}{d\Omega} = \frac{(4\pi)^{-2} g^4 f^4(\omega_0)}{(E_0 - E_R)^2 + \frac{1}{4}\Gamma^2} \times \{1 - 2\cos[(E_0 - E_R)(t' - t_0)] \times \exp[-\tfrac{1}{2}\Gamma(t' - t)] + \exp -\Gamma(t' - t_0) + O(\mu\Gamma/k_R^2)\}, \quad (65)$$

where

$$\cos[(k_0 - k_R)(x_2 + x_0) - (E_0 - E_R)t] \cong \cos[(E_0 - E_R)(t' - t_0)].$$

For the limit $(t' - t_0) \to \infty$, this result is what one expects for an S-matrix calculation for N, θ scattering where the cross section is defined as[11]

$$d\sigma = \iint \frac{(2\pi)^4 \delta(p_f - p_i)\,|R_{fi}|^2}{\rho_N \rho_\theta\,|\mathbf{v}_N - \mathbf{v}_\theta|} \frac{d\mathbf{p}}{(2\pi)^3 \rho_N} \frac{d\mathbf{k}}{(2\pi)^3 \rho_\theta}, \quad (66)$$

where the particle densities are

$$\rho_\theta = 2\omega(2\pi)^{-3},$$
$$\rho_N = (E_N/M_N)(2\pi)^{-3}, \quad (67)$$

and

$$S_{fi} = -(2\pi)^4 i\delta(p_f - p_i) R_{fi}. \quad (68)$$

The time dependence of Eq. (65) is exactly that for the decay of a prepared unstable state of mean life Γ^{-1} as calculated by Goldberger and Watson.[6] Thus the cross section represents the elastic scattering of N, θ particles via the formation of an intermediate unstable state which subsequently decays.

Since it is assumed that the resonance width is very narrow, it is appropriate in finding the transition rate per unit volume to integrate the incident beam energies over the resonance spectrum

$$\frac{dN}{dt} = \rho_N \int d\Omega \int d\omega_0 W(\omega_0) \rho_\theta \frac{k}{\omega_0} d\sigma/d\Omega, \quad (69)$$

where $W(\omega_0)$ is the spectral distribution of the incident θ particles. The result is

$$\frac{dN}{dt} = W(\omega_R)\rho_N k_R g^4 f^4(\omega_R)(2\pi)^{-3}\Gamma^{-1} \times \{1 - \exp -[\Gamma(t' - t_0)]\}, \quad (70)$$

[10] C. L. Hammer and T. A. Weber, J. Math. Phys. **8**, 494 (1967). This term is shown to exist for any isolated pole in the second sheet, independent of the model or the wavepacket shape.

[11] Stephen Gasciorowicz, *Elementary Particle Physics* (John Wiley & Sons, Inc., New York, 1960), p. 142.

where the integral relations

$$\int (d\omega_0/\omega_0) k\rho_\theta W(\omega_0)[(E_0 - E_R)^2 + (\tfrac{1}{2}\Gamma)^2]^{-1} \cong 2k_R\Gamma^{-1}W(\omega_R)(2\pi)^{-2},$$

$$\int \frac{(d\omega_0/\omega_0)k\rho_\theta W(\omega_0)\cos[(E_0 - E_R)(t' - t_0)]}{(E_0 - E_R)^2 + (\tfrac{1}{2}\Gamma)^2} \cong 2k_R\Gamma^{-1}W(\omega_R)(2\pi)^{-2}\exp[-\tfrac{1}{2}\Gamma(t' - t_0)] \quad (71)$$

have been used.

The quantity Γ^{-1} can be eliminated by using Eqs. (25) and (52), so that

$$\frac{dN}{dt} = \rho_N W(\omega_R) g^2 f^2(\omega_R)(2\pi)^{-2}[1 - \exp -\Gamma(t' - t_0)], \quad (72)$$

or, in terms of matrix elements,

$$\frac{dN}{dt} = 2\pi W(\omega_R)\rho_N |\langle k_R, p - k_R, 0|\,(R)\,|p, 0, 0\rangle_D|^2 \times [1 - \exp -\Gamma(t' - t_0)], \quad (73)$$

where $|p, 0, 0\rangle_D$ is the dressed mathematical state

$$|p, 0, 0\rangle_D = |Z'(E_V')|^{-\frac{1}{2}}\,|p, 0, 0\rangle.$$

This is the result for the rate of formation of decay products that should be expected on classical grounds for a beam of particles which begins to form unstable states of mean life Γ^{-1} at the time t_0 at the constant rate $2\pi W(\omega_R)\rho_N |\langle k_R, p - k_R, 0|\,(R)\,|p, 0, 0\rangle_D|^2$—a direct parallel to the quantum-mechanical calculation done here for a semi-infinite wavetrain representing the θ particle which begins interacting with the N particle at t_0, interacting continuously thereafter to form unstable V-particle states.

Since the rate of decay of the V particle $d\mathcal{N}/dt$ is related to the rate of formation of the N and θ particles dN/dt by

$$\frac{dN}{dt} = -\frac{d\mathcal{N}}{dt} = \Gamma\mathcal{N},$$

the number of V particles per unit volume present at the time t' is deduced as

$$\mathcal{N} = 2\pi\rho_N W(\omega_R)\Gamma^{-1}|\langle k_R, p - k_R, 0|\,(R)|p, 0, 0\rangle_D|^2 \times [1 - \exp -\Gamma(t' - t_0)]. \quad (74)$$

VI. CONFIGURATION REPRESENTATION FOR THE UNSTABLE V PARTICLE

The probability that a V particle will be at the position $\mathbf{x}_1$ at the time t can be obtained from the amplitude

$$\psi_V(\mathbf{x}_1, t) = \int d\mathbf{k} A(\mathbf{k}) \exp(-iEt)\langle 0, 0, x_1 \,|\, 0, N, \theta\rangle. \quad (75)$$

Substitution for $|0, N, \theta\rangle$ from Eq. (27) and for $A(\mathbf{k})$ from Eq. (57) gives

$$\psi_V(\mathbf{x}_1, t) = (2\pi)^{-3}(2\pi i)^{-1} g_0 \exp(i\mathbf{p}\cdot\mathbf{x}_1) \times \int_{\mathcal{C}} \frac{dk f(\omega)\exp i(kx_0 - Et)}{(k - k_0 + i\epsilon)[E - M_V - R_p(E)]}, \quad (76)$$

where $\mathcal{C}$ is the contour used previously. It should be noted here that $\psi_V(\mathbf{x}_2, 0) = 0$, as is appropriate. This expression can be evaluated for large t and x_0 (recall $\mathbf{x}_0$ is the initial position of the leading edge of the incident wave and is therefore asymptotic) with x_0/t finite in a manner similar to that used for the scattered wave in the previous section. The result is, for small (Γ/k_R^2),

$$\psi_V(\mathbf{x}_1, t) = -\frac{(2\pi)^{-3} g f(\omega_R)}{E_0 - E_R + \frac{1}{2}i\Gamma}\exp(i\mathbf{p}\cdot\mathbf{x}_1) \times \left\{\Theta\left(\left(\frac{k_R}{\mu}\right)t - x_0\right)\exp i(k_0 x_0 - E_0 t) - \Theta\left(\left(\frac{k_R}{\mu}\right)t - x_0\right)\exp[-\tfrac{1}{2}\Gamma(t - t_0)] \times \exp i(k_R x_0 - E_R t)\right\} + O\left(\frac{\mu\Gamma}{k_R^2}\right). \quad (77)$$

The probability density for $t > t_0$ is

$$\psi_V^\dagger\psi_V = \frac{(2\pi)^{-3}\rho_N g^2 f^2(\omega_R^2)}{(E_0 - E_R)^2 + \frac{1}{4}\Gamma^2} \times \{1 - \cos[(E_R - E_0)(t - t_0)] \times \exp[-\tfrac{1}{2}\Gamma(t - t_0)] + \exp[-\Gamma(t - t_0)]\}. \quad (78)$$

The number of V particles per unit volume follows as

$$\mathcal{N} = \int d\omega_0 W(\omega_0)\psi_V^\dagger\psi_V = 2\pi\rho_N W(\omega_R)\,|\langle k_R, p - k_R, 0|\,(R)\,|p, 0, 0\rangle_D|^2\,\Gamma^{-1} \times \{1 - \exp[-\Gamma(t - t_0)]\}, \quad (79)$$

in agreement with the result deduced in Eq. (74).

VII. THE UNSTABLE V-PARTICLE STATE

As pointed out by Levy,[4] any state that represents the unstable V particle will contain coefficients that depend upon the method of production. On the other hand, once the existence of this state is made

known through a subsequent measurement, the probability for the system being in that state at that instant is unity. Thus the state should be renormalized, thereby making the coefficients independent of the production process.

In the previous section, for a particular set of initial conditions, the appropriate time dependence is obtained for both the production of N and θ particles and unstable V particles from that part of the $|0, N, \theta\rangle$ state which is not the incident wave, that is, the second and third terms of Eq. (27). Also, only this part of $|0, N, \theta\rangle$ contributes to the stable V-particle state derived in Eqs. (28) and (30). This clearly indicates that the state vector for the V particle is

$$|V(t)\rangle = (2\pi)^{-\frac{3}{2}}\int d\mathbf{p}\int_{\mathcal{C}} d\mathbf{k} \times \frac{A(\mathbf{k},\mathbf{p})(2\omega)^{-\frac{1}{2}}f(\omega)\exp(-iEt)\{|p,0,0\rangle}{E - M_V - R_p(E)} + g_0(2\pi)^{-\frac{3}{2}}\int\frac{(d\bar{\mathbf{k}}/2\bar{\omega})f(\bar{\omega})}{E - M_N - \bar{\omega}}|0, p-\bar{k}, \bar{k}\rangle\}, \tag{80}$$

where $A(\mathbf{k}, \mathbf{p})$ depends upon the shape of the incident wavepackets for the N and θ particles, that is, the way in which the N and θ particles are initially produced, and $\mathcal{C}$ is chosen such that $|V(0)\rangle$ vanishes. The rate of decay into N and θ particles is then obtained from $\langle x_2, x_1, 0 \mid V(t)\rangle$ or $\langle k, p-k, 0 \mid V(t)\rangle$, whereas the number of unstable V particles is obtained from $\langle 0, 0, x_1 \mid V(t)\rangle$ or $\langle 0, 0, p \mid V(t)\rangle$. It is obvious that this choice for $|V(t)\rangle$ leads to the solutions previously derived when Eq. (57) is used for $A(\mathbf{k})$:

$$A(\mathbf{k}, p') = A(\mathbf{k})\delta(\mathbf{p}' - \mathbf{p}).$$

Since the V particle considered here is not formed for $t < 0$, it is not surprising that $\langle V(0) \mid V(0)\rangle = 0$. This means, however, that the state cannot be normed in the usual manner. An appropriate normalization, which is very physical, is also considered by Levy.[4] Since the state can be observed only for asymptotic t, that is, a time long compared to the interaction time, it should be normalized in the asymptotic time limit. For example, if $E - M_V - R(E)$ has a zero only at $E = E_V$ (the bound-state pole) and

$$A(\mathbf{k}, \mathbf{p}') = i[4\pi\omega Z(E_V)]^{-\frac{1}{2}}[kf(\omega)]^{-1}a(\mathbf{p}), \tag{81}$$

where

$$\int d\mathbf{p}\,|a(\mathbf{p})|^2 = 1,$$

then the choice of $\mathcal{C}$ shown in Fig. 1 leads to

$$\langle V(t) \mid V(t)\rangle_{t\to\infty} = 1. \tag{82}$$

It is easier to see this result if the integration variable used in Eq. (80) is changed from k to ω. The contour $\mathcal{C}'$ in the ω plane which corresponds to $\mathcal{C}$ is shown in Fig. 3(a). The state $|V(t)\rangle$ then is

$$|V(t)\rangle = -(2\pi i)^{-1}Z^{-\frac{1}{2}}(E_V) \times\int d\mathbf{p}a(\mathbf{p})\int_{\mathcal{C}'} d\omega\,\frac{\exp(-iEt)}{E - M_V - R_p(E)} \times\{|p,0,0\rangle + g_0(2\pi)^{-\frac{3}{2}}\int\frac{(d\bar{\mathbf{k}}/2\bar{\omega})f(\bar{\omega})}{E - M_N - \bar{\omega}} \times |0, p-\bar{k}, \bar{k}\rangle\}. \tag{83}$$

The asymptotic limit is obtained by rotating the contour 90° in a clockwise direction, as shown in Fig. 3(b). The integration about the branch point in $R_p(E)$ leads to terms with inverse powers of t so that only the pole term contributes in the large-time limit:

$$|V(t)\rangle_{t\to\infty} = \int d\mathbf{p}a(\mathbf{p})\exp(-iE_Vt)\Big[Z^{\frac{1}{2}}(E_V)\,|p,0,0\rangle + g(2\pi)^{-\frac{3}{2}}\int\frac{(d\mathbf{k}/2\omega)f(\omega)\,|0, p-k, k\rangle}{E_V - M_N - \omega}\Big] + O(t^{-\frac{3}{2}}),$$

$$|V(t)\rangle = \int d\mathbf{p}a(\mathbf{p})\,|V_p, 0, 0\rangle\exp(-iE_Vt) + O(t^{-\frac{3}{2}}). \tag{84}$$

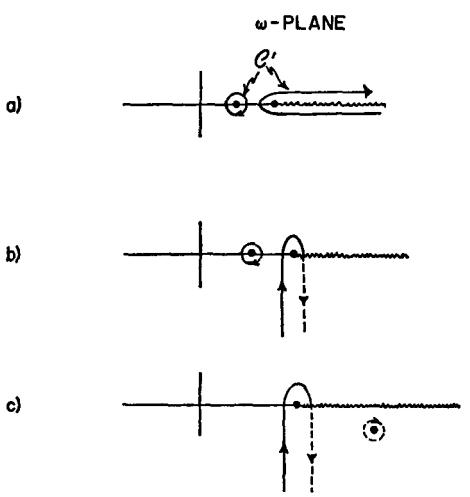

FIG. 3. Various contours in the ω plane: (a) corresponds to the contour $\mathcal{C}$ in the k plane when a bound-state pole is present; (b) the contour used to obtain the asymptotic time equations for the stable case; (c) the contour used to obtain the asymptotic time equations for the unstable case. The pole in the second sheet is explicitly shown.

Therefore, after a long time t, the system finds itself in a stable V-particle state. Equation (82) then follows directly from the normalization of $|V_p, 0, 0\rangle$.

If $E - M_V - R(E)$ has a zero at $E = E_V'$, the pole moves to the second sheet as shown in Fig. 3(c). Correspondingly, if $|Z'^{-\frac{1}{2}}(E_V')|$ is used in Eq. (81) instead of $Z^{-\frac{1}{2}}(E_V)$, the result is

$$\langle 0, 0, p' \mid V(t)\rangle_{t\to\infty} = [Z'(E_V')/|Z'^{\frac{1}{2}}(E_V')|]a(\mathbf{p}') \times \exp(-iE_R t)\exp[-\tfrac{1}{2}\Gamma t] + O(t^{-\frac{3}{2}}). \quad (85)$$

Thus, for small Γt, the sum over all "dressed" momentum states has the expected exponential decay

$$\int d\mathbf{p}' \,|_D\langle 0, 0, p' \mid V(t)\rangle|^2 = \exp(-\Gamma t). \quad (86)$$

The projection onto the state $|0, p' - k', k'\rangle$ is more complicated because of the extra pole that appears on the real axis in the expression

$$\langle k', p' - k', 0 \mid V(t)\rangle = -(2\pi i)^{-1}\,|Z'^{-\frac{1}{2}}(E_V')|\, a(\mathbf{p}')(2\pi)^{-\frac{3}{2}} g_0 f(\omega') \times \int_{C'} \frac{d\omega \exp(-iEt)}{(\omega - \omega')[E - M_V - R_p(E)]}. \quad (87)$$

For large t, this becomes

$$\langle k', p' - k', 0 \mid V(t)\rangle_{t\to\infty} = \frac{(2\pi)^{-\frac{3}{2}} a(\mathbf{p}') g_0 f(\omega')\,|Z'^{-\frac{1}{2}}(E_V')|}{\omega' + M_N - E_R - i(\Gamma/2)} \times \{Z'(\omega' + M_N)\exp[-i(\omega' + M_N)t] - Z'(E_V') \times \exp(-iE_R t)\exp(-\tfrac{1}{2}\Gamma t)\} + O(t^{-\frac{3}{2}}). \quad (88)$$

For Γ/E_R small, the probability for the decay into an N and θ particle is

$$\int d\mathbf{p}' \int (d\mathbf{k}'/2\omega')\,|\langle k', p' - k', 0 \mid V(t)\rangle|^2 \cong (\Gamma/2\pi)\int d\omega \times \frac{[1 - 2\cos(E - E_R)t\exp(-\tfrac{1}{2}\Gamma t) + \exp(-\Gamma t)]}{(E - E_R)^2 + (\tfrac{1}{2}\Gamma)^2} \cong I - \exp(-\Gamma t), \quad (89)$$

where the last integral is evaluated similarly to those in Eq. (71). The normalization of $|V(t)\rangle$ follows from Eqs. (86) and (89), since, to lowest order in g_0^2 (small Γ), $|p, 0, 0\rangle$ can be replaced by $|p, 0, 0\rangle_D$ in the completeness relation.

It is interesting to note that, for both the stable and unstable case, the state defined by Eq. (83) [with $Z'(E_V')$ for the unstable case] is

$$|V(t)\rangle = \int d\mathbf{p} a(\mathbf{p}) \exp(-iHt)\,|p, 0, 0\rangle_D, \quad (90)$$

where H is the Lee-model Hamiltonian. This is readily seen by expanding $|p, 0, 0\rangle$ in terms of the complete set of states $|0, N, \theta\rangle$ and $|V_p, 0, 0\rangle$. Thus, one has, for the stable case,

$$|V(t)\rangle = I_1 + I_2,$$

where

$$I_1 = \int d\mathbf{p} a(\mathbf{p})\,|V_p, 0, 0\rangle \exp(-iE_V t) \quad (91)$$

and

$$I_2 = -(2\pi)^{-\frac{3}{2}} g_0 Z^{-\frac{1}{2}}(E_V)\int d\mathbf{p} a(\mathbf{p}) \times \int \frac{(d\mathbf{k}/2\omega) f(\omega)}{E - M_V - R_p^*(E) - i\eta}\,|0, N, \theta\rangle \exp(-iEt). \quad (92)$$

The term I_1 is the result expressed in Eq. (84) which arose from Eq. (83) by integrating in a clockwise direction around the bound-state pole. Thus, if I_2 can be shown to be the integration around the branch point shown in Fig. 3(a), Eq. (90) will be proven.

Substitution from Eq. (27) for $|0, N, \theta\rangle$ gives

$$I_2 Z^{\frac{1}{2}}(E_V) = -(2\pi)^{-\frac{3}{2}} g_0 \int d\mathbf{p} a(\mathbf{p}) \times \int \frac{(d\mathbf{k}/2\omega)\,|0, p - k, k\rangle \exp(-iEt)}{E - M_V - R_p^*(E) - i\eta} + \pi^{-1}\int d\mathbf{p} a(\mathbf{p}) \int_\mu^\infty d\omega \times \frac{I(E)\exp(-iEt)}{[E - M_V - R_p^*(E) - i\eta][E - M_V - R_p(E) + i\eta]} \times \left(|p, 0, 0\rangle - (2\pi)^{-\frac{3}{2}} g_0 \int \frac{(d\bar{\mathbf{k}}/2\bar{\omega})\,|0, p - \bar{k}, \bar{k}\rangle}{E - M_N - \bar{\omega} + i\eta}\right), \quad (93)$$

where Eq. (25) is used to obtain $I(E)$ in the second term. The ω integration in the second term of Eq. (93) can be broken into two pieces using

$$\frac{-2iI(E)}{[E - M_V - R_p^*(E)][E - M_V - R_p(E)]} = \frac{1}{E - M_V - R_p(E)} - \frac{1}{E - M_V - R_p^*(E)}. \quad (94)$$

The first piece corresponds to an integral from μ to ∞ above the cut in $R_p(E)$. The second piece corresponds to an integral from ∞ to μ below the cut in

$R_p(E)$ if E in the denominator of the $\bar{\omega}$ integration in Eq. (93) is continued to values below the real $\bar{\omega}$ axis. This continuation gives rise to a contribution from the $\bar{\omega}$ integration which cancels the first term in I_2, giving finally the cut contribution to Eq. (83):

$$I_2 = -(2\pi i)^{-1} Z^{-\frac{1}{2}}(E_V) \int d\mathbf{p} \int \frac{d\omega a(\mathbf{p})}{E - M_V - R_p(E)} \times \left(|p,0,0\rangle - (2\pi)^{-\frac{3}{2}} g_0 \int \frac{(d\bar{k}/2\bar{\omega}) f(\bar{\omega}) |0, p-\bar{k}, \bar{k}\rangle}{E - M_N - \bar{\omega}} \right), \tag{95}$$

where C is that part of $\mathcal{C}'$ which is the contour around the cut in $R_p(E)$ shown in Fig. 3(a). The proof of Eq. (90) for the unstable case follows similarly, except that $I_1 \equiv 0$ and $|Z'^{-\frac{1}{2}}(E'_V)|$ replaces $Z^{-\frac{1}{2}}(E_V)$.

In parallel to the state $|V(t)\rangle$, there is a state $|N\theta(t)\rangle$, which can be defined as

$$|N\theta(t)\rangle \doteq \int d\mathbf{p} \int d\mathbf{k} (2\omega)^{-\frac{1}{2}} A(\mathbf{k}, \mathbf{p}) \times \exp(-iHt) |0, p-k, k\rangle, \tag{96}$$

which in the asymptotic limit becomes

$$|N\theta(t)\rangle_{t\to\infty} = \int d\mathbf{p} \int d\mathbf{k} A(\mathbf{k}, \mathbf{p}) \exp(-iEt) |0, N, \theta\rangle. \tag{97}$$

The proof of this last equation follows similarly to the proofs used to show Eqs. (84) and (90).

VIII. CONCLUSIONS

In view of Eqs. (84), (86), (89), (90), and (97) there appears to be no difficulty in defining $|\text{in}\rangle$ states for the Lee model even for the case in which no stable V particle exists. However, to obtain the resonant state, the infinite time limit must be avoided.

The ambiguities pointed out by Levy[4] in the definition of the renormalization constant $Z'(E'_V)$ are removed by requiring a pure exponential decay for the V particle as shown in Eq. (86). Then, for the unstable case as expressed by Eq. (84), the probability that the V particle is "initially" in the "dressed" momentum state $|p, 0, 0\rangle_D$ is

$$\int d\mathbf{p} \, |\langle V(t) \mid p, 0, 0\rangle_D|^2 = 1,$$

the same result obtained from Eq. (84) for the stable case.

To what extent the results discussed in this paper apply for isolated resonant states in general is not completely clear because, although the level shift operator approach is itself quite general, the specifics presented here depend heavily upon the fact that only two free-particle states are needed to describe the $N + \theta$ sector of the Lee model. However, the assertions proven here will be very useful for the examinations of more general models.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Irreducibility of the Ladder Representations of $U(2, 2)$ when Restricted to the Poincaré Subgroup

G. MACK*
Center for Theoretical Studies, University of Miami, Coral Gables, Florida
AND
I. TODOROV†
Institute for Advanced Studies, Princeton, New Jersey

(Received 12 March 1969)

It is shown that the most degenerate discrete series of unitary irreducible representations of $U(2, 2)$, the so-called ladder representations, remain irreducible when restricted to representations of the Poincaré subgroup $ISL(2, C)$. They correspond to representations of this subgroup with mass zero and arbitrary integer or half-integer helicity λ. The basis vectors of the canonical basis are calculated as functions of a lightlike 4-vector, which is formed by the simultaneous eigenvalues of the generators of the subgroup of translations.

1. INTRODUCTION

The group of pseudo-unitary transformations $SU(2, 2)$ has entered physics in various different ways. First, it appeared as covering group of the conformal group of space–time transformations (including the Poincaré group, dilatations, and special conformal transformations—see Appendix). With this interpretation, it was considered as a symmetry of the wave equations for massless particles.[1–6] Secondly, the Lie algebra of $SU(2, 2)$ was considered as the spectrum-generating algebra for the (spinless) hydrogen atom.[7] Thirdly, it was also used in hadron physics as an algebra acting on the space spanned by the states of an infinite multiplet of particles at rest, or acting on the indices of an infinite-component field.[8] In this case, the so-called ladder representations of $SU(2, 2)$ were used, which can be simply described in terms of creation and annihilation operators.[9]

In the main part of the present paper, we show that all the ladder representations of $U(2, 2)$ [or $SU(2, 2)$] remain irreducible when restricted to representations of its Poincaré subgroup[10] $ISL(2, C)$. They correspond to representations of this subgroup characterized by zero mass and arbitrary integer and half-odd integer helicity λ. The first-order Casimir operator of $U(2, 2)$, which labels the irreducible representations of the ladder series, is linearly related to the helicity λ.

We also calculate the basis vectors of the canonical basis as functions of a lightlike 4-vector ξ_μ, which is formed by the simultaneous eigenvalues of the generators of the subgroup of translations [Eq. (4.4) and following]. The form of the generators of $SU(2, 2)$ when acting on functions $f(\xi_\mu)$ is given in Eq. (3.11).

It is amusing to find that one is led to the same set of irreducible representations of $SU(2, 2)$ for all three physical interpretations of this group mentioned above (Sec. 5). In particular, the representation used for the group-theoretic description of the H atom is equivalent to the one used for the description of massless spin-0 particles.

An appendix is added which deals with the conformal group of space–time in quantum field theory. A comment is included on the connection between the conformal structure of space–time and infinite-component field theories of the type investigated recently.[8]

* Supported in part by U.S. Air Force under contract No. AF-AFOSR 69-1628.

† On leave of absence from the Joint Institute for Nuclear Research, Dubna, USSR, and the Physical Institute of the Bulgarian Academy of Sciences, Sofia, Bulgaria.

[1] P. A. M. Dirac, Ann. Math. **37**, 429 (1936).

[2] H. A. Kastrup, Phys. Rev. **142**, 1060 (1966); **143**, 1041 (1966); **150**, 1183 (1966); Ann. Physik **9**, 388 (1962). The last reference contains a historical survey on physical applications of the conformal group.

[3] A. Salam and G. Mack, Ann. Phys. (N.Y.) (to be published).

[4] G. Mack, Nucl. Phys. **B5**, 499 (1968).

[5] M. Flato and D. Sternheimer, Compt. Rend. **263**, 935 (1966).

[6] L. Gross, J. Math. Phys. **8**, 1931 (1967), and references reviewed in Ref. 3.

[7] I. A. Malkin and V. I. Manko, J. Nucl. Phys. (USSR) **3**, 372 (1966); Y. Nambu, Progr. Theoret. Phys. (Kyoto) Suppl. **37, 38**, 368 (1966); Phys. Rev. **160**, 1171 (1967); A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967); **157**, 1180 (1967).

[8] A. O. Barut and H. Kleinert, Phys. Rev. **161**, 1464 (1967); A. O. Barut, *Lectures in Theoretical Physics* (Gordon & Breach Publ. Co., New York, 1968), Vol. 10, Part B; C. Fronsdal (report of work prior to publication).

[9] See B. Kurşunoglu, *Modern Quantum Theory* (W. H. Freeman and Co., San Francisco, 1962), p. 257. This description became popular after the paper of Dothan, M. Gell-Mann, and Ne'eman, Phys. Letters **17**, 148 (1965). The relation between the ladder representations of $U(p, q)$ and the Gel'fand–Graev discrete series of representations of this group is described in I. Todorov, ICTP, Trieste, Preprint IC/66/71, 1966. Yao uses the term "exceptional degenerate discrete series": T. Yao, J. Math. Phys. **8**, 1931 (1967); **9**, 1615 (1968).

[10] This problem of irreducibility has been stated in D. Sternheimer, J. Math. Pure Appl. **47**, 289 (1968). The authors are indebted to Dr. D. Sternheimer for bringing this reference to their attention.

2. THE LADDER REPRESENTATIONS OF $U(2, 2)$

We start by recalling the definition of the ladder representations. We present it here in a basis-independent way. This is not only demanded by the canon of mathematical aesthetics, but is also convenient for practical computations as it allows one to choose the most convenient basis for the treatment of any given problem.

We define $U(2, 2)$ as the group of linear transformations in the complex 4-dimensional space C_4 which preserves the Hermitian form

$$\bar{u}\beta v \equiv \sum_{\alpha=1}^{4} \bar{u}_\alpha v^\alpha. \tag{2.1}$$

Here and in the following, a bar stands for complex conjugation. β is a Hermitian matrix with two positive and two negative eigenvalues[11] satisfying the $U(2, 2)$-invariant normalization $\det \beta = 1$. By virtue of the invariance of Eq. (2.1), the generators $J_{AB} = -J_{BA} = \gamma_{AB}$ of the defining representation of $SU(2, 2)$ obey

$$\beta\gamma_{AB}\beta^{-1} = \gamma_{AB}^*. \tag{2.2}$$

They admit of the following commutation relations (CR):

$$[J_{KL}, J_{MN}] = i(g_{KN}J_{LM} + g_{LM}J_{KN} - g_{KM}J_{LN} - g_{LN}J_{KM}), \tag{2.3}$$

where $g_{KK} = (+ - - - - +)$ and $g_{KM} = 0$ otherwise. Capital Roman letters run over the values 0, 1, 2, 3, 5, 6.

The remaining generator C_1 of $U(2, 2)$ commutes with all the J_{AB} and is represented by unity in the defining representation.

Let the Dirac matrices [i.e., spin affinors $\gamma_\mu = (\gamma_\mu)_\beta^\alpha$] be defined in the usual way, satisfying covariant anticommutation relations for $\mu, \nu = 0, \cdots, 3$:

$$\{\gamma_\mu, \gamma_\nu\} = 2g_{\mu\nu}.$$

Then we may choose the matrices γ_{AB} of the defining representation as follows:

$$\begin{aligned} &\gamma_{\mu\nu} = \tfrac{1}{4}i[\gamma_\mu, \gamma_\nu], \quad \gamma_{56} = \tfrac{1}{2}\gamma_5 \equiv \tfrac{1}{2}\gamma_0\gamma_1\gamma_2\gamma_3, \\ &\gamma_{\mu 5} = \tfrac{1}{2}i\gamma_\mu\gamma_5, \qquad \gamma_{\mu 6} = \tfrac{1}{2}\gamma_\mu. \end{aligned} \tag{2.4}$$

They satisfy Eq. (2.3). A matrix β obeying Eq. (2.2) always exists. Its precise form depends on the choice of basis for the γ-matrices.[12]

[11] We remark that this property of β is invariant because of the inertia law of quadratic forms.

[12] It has been conventional to choose γ_0 and $i\gamma_j$ Hermitian; in that case $\beta = \pm\gamma_0$. This identification is, however, not invariant under an arbitrary change of basis, but only under those induced by a unitary transformation.

The ladder representations are now constructed as follows: We define the operator-valued 4-component spinor, $\varphi \equiv (\varphi^\alpha)$ and $\tilde{\varphi} = \varphi^*\beta \equiv (\tilde{\varphi}_\alpha)$, $\alpha = 1, 2, 3, 4$, and impose the canonical commutation relations

$$[\varphi^\alpha, \tilde{\varphi}_\beta] = \delta_\beta^\alpha, \quad [\varphi^\alpha, \varphi^\beta] = 0. \tag{2.5}$$

Each (star) representation of the canonical commutation relations for the φ's gives rise to a unitary (infinite-dimensional) representation of $U(2, 2)$ generated by

$$C_1 = \tilde{\varphi}\varphi, \quad J_{AB} = \tilde{\varphi}\gamma_{AB}\varphi. \tag{2.6}$$

There are two important inequivalent realizations of the canonical commutation relations (2.5) and, correspondingly, two series of inequivalent ladder representations of $U(2, 2)$. For the first, we define an $SU(2) \otimes SU(2)$ invariant vector Ψ_0 by

$$\Pi_+\varphi\Psi_0 = \tilde{\varphi}\Pi_-\Psi_0 = 0, \quad \Pi_\pm \equiv \tfrac{1}{2}(1 \pm \gamma_0). \tag{2.7}$$

For the second one, we put

$$\Pi_-\varphi\Psi_0 = \tilde{\varphi}\Pi_+\Psi_0 = 0. \tag{2.8}$$

To be consistent with the positivity of the metric in the representation space, we have to assume in each case the positive-definiteness of the 4×4 matrix:

$$\begin{aligned} &(\Psi_0, (\tilde{\varphi}\Pi_\pm)^*(\tilde{\varphi}\Pi_\pm)\Psi_0) \\ &\quad = (\Psi_0, (\beta\Pi_\pm\varphi)(\tilde{\varphi}\Pi_\pm)\Psi_0) = \beta\Pi_\pm \end{aligned} \tag{2.9}$$

[the last equality is a consequence of Eqs. (2.5) and (2.7) or (2.8)]. This leads to a different sign of β in the two cases. Taking a basis in which γ_0 is Hermitian and γ_j are anti-Hermitian, we obtain $\beta = \gamma_0$ in the case (2.7) and $\beta = -\gamma_0$ in the case (2.8). In the first case, we obtain the so-called $\mathfrak{L}$ series of most degenerate representations of $U(2, 2)$; in the second case, we arrive at the $\mathfrak{L}^*$ series (see Ref. 9); the representations of these two series are conjugate to each other. Each series contains a denumerable set of (unitary) irreducible representations of $U(2, 2)$, labeled by the (integer) value of the first-order Casimir operator C_1. The canonical basis is defined in terms of the eigenvectors of the (maximal) set of commuting operators

$$-\tfrac{1}{2}C_1 - 1, \quad J_{06}, \quad \mathbf{M}^2 = J_{12}^2 + J_{23}^2 + J_{31}^2, \quad M_3 = J_{12},$$

with eigenvalues λ, n, $s(s + 1)$, and m. They change in the range $\lambda = 0, \pm\tfrac{1}{2}, \cdots$; $\pm n = |\lambda| + 1, |\lambda| + 2, \cdots$ [+ sign for the case (2.7) and − sign for the case (2.8)]; $s = |\lambda|, |\lambda| + 1, \cdots, |n| - 1$; $-s \leq m \leq s$. The representation used for the group-theoretic description of the nonrelativistic hydrogen atom is contained in the $\mathfrak{L}$ series for $\lambda = 0$ ($C_1 = -2$). All these representations are known to be integrable

to unitary representations of the universal covering group.[13]

3. REDUCTION TO REPRESENTATIONS OF THE POINCARÉ SUBALGEBRA

It is convenient to introduce an alternative set of generators in which the Poincaré subalgebra is displayed explicitly. We define, for $\mu, \nu = 0, \cdots, 3$,

$$M_{\mu\nu} = J_{\mu\nu}, \quad P_\mu = J_{\mu 6} + J_{\mu 5},$$
$$K_\mu = J_{\mu 6} - J_{\mu 5}, \quad D = J_{56}. \tag{3.1}$$

Their CR follow from Eq. (2.3) and are given in Eq. (A2) of the Appendix. In particular, the generators $M_{\mu\nu}$ and P_μ satisfy the CR of the Poincaré algebra (the same is true for the set $M_{\mu\nu}$ and K_ν).

First, we show that

$$P_\mu P^\mu = K_\mu K^\mu = 0, \tag{3.2}$$

i.e., that the ladder representations contain only zero-mass representations of the Poincaré subgroup.[14]

Let $B_{\alpha\beta}$ be the antisymmetric tensor defined by

$$B\gamma_\mu B^{-1} = \gamma_\mu^T, \quad (B^{-1})^{\alpha\beta} = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}B_{\gamma\delta}$$

and related to the charge-conjugation matrix $C_{\alpha\beta}$ by $B = iC\gamma_5$, where ϵ is the completely antisymmetric tensor with $\epsilon_{1234} = 1$. Using the identity

$$(\gamma_\mu)^\alpha_\beta(\gamma^\mu)^\gamma_\delta = \delta^\alpha_\beta\delta^\gamma_\delta + (\gamma_5)^\alpha_\beta(\gamma_5)^\gamma_\delta + 2\epsilon^{\alpha\gamma\sigma\tau}B_{\sigma\beta}B_{\tau\gamma}, \tag{3.3}$$

we readily establish that

$$R_\pm \equiv \tfrac{1}{4}[\gamma_\mu(1 \pm i\gamma_5)]^\alpha_\beta[\gamma^\mu(1 \pm i\gamma_5)]^\gamma_\delta$$

is traceless ($R^{\alpha\sigma}_{\sigma\delta} = 0$) and antisymmetric ($R^{\alpha\gamma}_{\beta\delta} = -R^{\gamma\alpha}_{\beta\delta}$). This implies Eq. (3.2) by virtue of the CR (2.5); for instance,

$$P_\mu P^\mu = \tilde{\varphi}_\alpha\varphi^\beta\tilde{\varphi}_\gamma\varphi^\delta R^{\alpha\gamma}_{\beta\delta} = \tilde{\varphi}_\alpha R^{\alpha\sigma}_{\sigma\delta}\varphi^\delta = 0.$$

One can check (3.3) by applying both sides to any of the quantities $(\gamma_{AB})^\delta_\gamma$, cf. (2.4) (sum over γ and δ).

It remains to show that these representations remain actually irreducible when restricted to representations of the Poincaré group and to determine the helicities. (Recall that the zero-mass representations of the Poincaré group which are used in physics are labeled by the helicity.)

Let us choose the Dirac matrices, such that γ_5 is diagonal, and the matrix β satisfying Eq. (2.2) as follows:

$$\gamma_0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}, \quad \gamma_i = \begin{pmatrix} 0 & -\sigma_i \\ \sigma_i & 0 \end{pmatrix}, \quad i\gamma_5 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & \sigma_0 \end{pmatrix},$$
$$\beta = \pm\gamma_0, \tag{3.4}$$

where σ_0 is the 2×2 unit matrix and the σ_k are Pauli matrices. We shall take σ_3 to be diagonal.

To carry out the reduction, we re-express the generators as differential operators acting on functions of a lightlike 4-vector ξ^μ which is formed by the simultaneous eigenvalues of the generators P^μ of translations.

Consider the Hilbert space of those complex functions $f(z_1, z_2)$ of two complex variables z_1, z_2 which have finite norm (f, f), where the scalar product is defined as

$$(f, g) = \iint \overline{f(z_1, z_2)}g(z_1, z_2)\, d^2z_1\, d^2z_2. \tag{3.5}$$

Here $d^2z = d\,\mathrm{Re}\, z\, d\,\mathrm{Im}\, z = \frac{1}{2}i\, dz\, d\bar{z}$. The polynomials in z, $\bar{z}$, $\partial/\partial z$, $\partial/\partial\bar{z}$ form an irreducible set of operators in this Hilbert space. Furthermore,

$$(\partial/\partial z)^* = -\partial/\partial\bar{z} \quad \text{and} \quad z^* = \bar{z}.$$

We may then proceed to writing down the operator-valued 4-component spinor φ in the form

$$\varphi = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \pm\partial/\partial z_1 \\ \pm\partial/\partial z_2 \end{pmatrix}$$

and

$$\tilde{\varphi} = \left(-\frac{\partial}{\partial\bar{z}_1}, -\frac{\partial}{\partial\bar{z}_2}, \pm z_1, \pm z_2\right), \tag{3.6}$$

where again the + sign refers to the case (2.7) and the − sign to the case (2.8). In both cases, the normalized $SU(2) \otimes SU(2)$ invariant vector is given by $\Psi_0 = (2/\pi)e^{-2\bar{z}}$. The ladder representation has been displayed in this form in Ref. 15.

The explicit form of the generators is found to be

$$\pm P_\mu = z\sigma_\mu\bar{z}, \quad D = \frac{i}{2}\left(z\frac{\partial}{\partial z} + \frac{\partial}{\partial\bar{z}}\bar{z}\right),$$
$$M_k = \tfrac{1}{2}\epsilon_{ijk}J_{ij} = \frac{1}{2}\left(z\sigma_k\frac{\partial}{\partial z} - \frac{\partial}{\partial\bar{z}}\sigma_k\bar{z}\right),$$
$$\pm K_\mu = -\frac{\partial}{\partial\bar{z}}\sigma^\mu\frac{\partial}{\partial z}, \tag{3.7}$$
$$N_k = J_{0k} = -\frac{i}{2}\left(z\sigma_k\frac{\partial}{\partial z} + \frac{\partial}{\partial\bar{z}}\sigma_k\bar{z}\right),$$
$$C_1 = z\frac{\partial}{\partial z} - \frac{\partial}{\partial\bar{z}}\bar{z}.$$

[13] R. L. Anderson, J. Fischer, and R. Rączka, Proc. Roy. Soc. (London) A302, 49 (1968). Moreover, it has been shown by Itzykson and Bargmann that there exists a pair of irreducible representations of the covering group of $Sp(8, R)$ generated by all Hermitian quadratic combinations of φ_α, φ^*_α. These representations are acting on the same Hilbert space (i.e., in our Fock space) and reduce, when restricted to the subgroup $U(2, 2)$, into direct sums of ladder representation with either only integer, or only half-odd integer λ. C. Itzykson, Commun. Math. Phys. 4, 92 (1967); V. Bargmann, *Group Representation on Hilbert Spaces of Analytic Functions*, lectures at the International Symposium on Analytic Methods in Mathematical Physics, Indiana University, 1968 (to be published).

[14] This observation has been made earlier by B. Kurşunoğlu, J. Math. Phys. 8, 1694 (1967).

[15] D. Tz. Stoyanov and I. Todorov, J. Math. Phys. 9, 2146 (1968).

Here and in the following, $\sigma^\mu = g^{\mu\nu}\sigma_\nu$ and the summation convention for $\mu, \nu = 0, \cdots, 3$ is adopted.

We now introduce a new set of four real independent variables ξ^1, ξ^2, ξ^3, and α, related to the complex variables z_1, z_2 by

$$-\xi^j = z\sigma_j\bar{z}, \quad j = 1, 2, 3, \quad -\infty < \xi^j < +\infty,$$
$$\alpha = \arg(z_1) + \arg(z_2), \quad 0 \leq \alpha < 4\pi. \tag{3.8}$$

All points with $\xi \equiv 0$ are identified; then, the mapping is bijective. arg (z) stands for the phase of z as usual. It is convenient to introduce, in addition,

$$\xi^0 = |\xi| \equiv z\sigma_0\bar{z}, \quad \xi_\mu = g_{\mu\nu}\xi^\nu. \tag{3.9}$$

Let us now re-express the functions $f(z_1, z_2)$ which form our Hilbert space as functions of the new variables ξ^j, α:

$$f(z_1, z_2) \equiv F(\xi^1, \xi^2, \xi^3; \alpha). \tag{3.10}$$

Making use of the chain rule

$$\frac{\partial}{\partial z_a} = \sum_{j=1}^{3} \frac{\partial \xi^j}{\partial z_a}\frac{\partial}{\partial \xi^j} + \frac{\partial \alpha}{\partial z_a}\frac{\partial}{\partial \alpha}$$
$$= \sum_{j=1}^{3} -(\sigma_j)_{a\dot{b}}\,\bar{z}_{\dot{b}}\frac{\partial}{\partial \xi^j} - \frac{i}{2}\frac{1}{z_a}\frac{\partial}{\partial \alpha}$$

and its analog for $\partial/\partial\bar{z}$, one finds the form of the generators acting on the new functions F as

$$\pm P_\mu = \xi_\mu \quad (+ \text{ for the } \mathfrak{L} \text{ series and } - \text{ for the } \mathfrak{L}^* \text{ series}),$$

$$M_{23} = -i\left(\xi^2\frac{\partial}{\partial \xi^3} - \xi^3\frac{\partial}{\partial \xi^2}\right) + |\xi|\,\xi^1\xi_\perp^{-2} i\frac{\partial}{\partial \alpha},$$

$$M_{31} = -i\left(\xi^3\frac{\partial}{\partial \xi^1} - \xi^1\frac{\partial}{\partial \xi^3}\right) + |\xi|\,\xi^2\xi_\perp^{-2} i\frac{\partial}{\partial \alpha},$$

$$M_{12} = -i\left(\xi^1\frac{\partial}{\partial \xi^2} - \xi^2\frac{\partial}{\partial \xi^1}\right),$$

$$M_{01} = +i|\xi|\frac{\partial}{\partial \xi^1} - \xi^2\xi^3\xi_\perp^{-2} i\frac{\partial}{\partial \alpha},$$

$$M_{02} = +i|\xi|\frac{\partial}{\partial \xi^2} + \xi^1\xi^3\xi_\perp^{-2} i\frac{\partial}{\partial \alpha},$$

$$M_{03} = +i|\xi|\frac{\partial}{\partial \xi^3},$$

$$D = i\left(1 + \xi^j\frac{\partial}{\partial \xi^j}\right),$$

$$\pm K_0 = -|\xi|\,\Delta_\xi - 2i\xi^3\xi_\perp^{-2}\left(\xi^2\frac{\partial}{\partial \xi^1} - \xi^1\frac{\partial}{\partial \xi^2}\right)$$
$$\times\, i\frac{\partial}{\partial \alpha} - |\xi|\,\xi_\perp^{-2}\frac{\partial^2}{\partial \alpha^2},$$

$$\pm K_1 = +\xi^1\Delta_\xi - 2\xi^j\frac{\partial}{\partial \xi^j}\frac{\partial}{\partial \xi^1} - 2\frac{\partial}{\partial \xi^1}$$
$$+ 2i|\xi|\,\xi^2\xi_\perp^{-2}\frac{\partial}{\partial \xi^3} i\frac{\partial}{\partial \alpha} - \xi^1\xi_\perp^{-2}\frac{\partial^2}{\partial \alpha^2},$$

$$\pm K_2 = +\xi^2\Delta_\xi - 2\xi^j\frac{\partial}{\partial \xi^j}\frac{\partial}{\partial \xi^2} - 2\frac{\partial}{\partial \xi^2}$$
$$- 2i|\xi|\,\xi^1\xi_\perp^{-2}\frac{\partial}{\partial \xi^3} i\frac{\partial}{\partial \alpha} - \xi^2\xi_\perp^{-2}\frac{\partial^2}{\partial \alpha^2},$$

$$\pm K_3 = +\xi^3\Delta_\xi - 2\xi^j\frac{\partial}{\partial \xi^j}\frac{\partial}{\partial \xi^3} - 2\frac{\partial}{\partial \xi^3}$$
$$- 2i|\xi|\,\xi_\perp^{-2}\left(\xi^2\frac{\partial}{\partial \xi^1} - \xi^1\frac{\partial}{\partial \xi^2}\right) i\frac{\partial}{\partial \alpha} + \xi^3\xi_\perp^{-2}\frac{\partial^2}{\partial \alpha^2}. \tag{3.11}$$

Here we have used the abbreviations

$$\Delta_\xi = \sum_{j=1}^{3}\frac{\partial^2}{\partial(\xi^j)^2} \quad \text{and} \quad \xi_\perp^2 = (\xi^1)^2 + (\xi^2)^2.$$

Summation over repeated indices j is to be carried out from 1 to 3. The first-order Casimir operator C_1 of $U(2, 2)$ takes the form

$$C_1 = -2i\frac{\partial}{\partial \alpha} - 2. \tag{3.12}$$

Finally, one finds for the invariant scalar product from Eq. (3.5)

$$(F, G) = \frac{1}{16}\int_0^{4\pi} d\alpha \int \frac{d^3\xi}{|\xi|}\,\overline{F(\xi; \alpha)}G(\xi; \alpha). \tag{3.13}$$

For an irreducible representation of $U(2, 2)$, C_1 is diagonal and takes integer values. Consequently,

$$i\frac{\partial}{\partial \alpha} = \lambda, \quad \lambda = 0, \pm\tfrac{1}{2}, \pm 1, \pm\tfrac{3}{2}, \cdots,$$

and

$$C_1 + 2 = -2\lambda. \tag{3.14}$$

Substituting this into Eq. (3.11), we see that P_μ, $M_{\mu\nu}$ are then generators of an irreducible representation of the Poincaré group with zero mass and helicity λ as given by Shirokov.[16] Evidently, the remaining generators D, K_μ act on the same irreducible representation space of the Poincaré group. The first-order Casimir operator C_1 is linearly related to the helicity λ by Eq. (3.13). This completes the proof of our statements.

As a final remark, we note that the decomposition of an arbitrary vector F into a sum of vectors that transform according to an irreducible representation

[16] Iu. M. Shirokov, Zh. Eksp. Teor. Fiz. 33, 1208 (1957) [Sov. Phys.—JETP 6, 929 (1958)]. (There is a sign error in Shirokov's formula for $M_{23} \equiv M_1$.)

of $U(2,2)$ is effected by carrying out a Fourier transform of F with respect to the variable α over the interval $[0, 4\pi)$:

$$F(\xi, \alpha) = \frac{2}{\pi} \sum_{\lambda=0,\pm\frac{1}{2},\cdots} F_\lambda(\xi) e^{-i\lambda\alpha}.$$

Here the helicity λ runs through all integer and half-integer numbers. The scalar product (3.13) may then be rewritten as

$$(F, G) = \sum_\lambda \int \frac{d^3\xi}{2|\xi|} \overline{F_\lambda(\xi)} G_\lambda(\xi). \tag{3.15}$$

4. THE CANONICAL BASIS

The canonical basis $\Psi_{\lambda,n,s,m}$ introduced in Sec. 2 is defined in terms of the eigenvectors of a complete set of commuting operators of the maximal compact subgroup $U(2) \otimes U(2)$:

$$\begin{aligned} (C_1 + 2 + 2\lambda)\Psi_{\lambda,n,s,m} &= 0, \\ (J_{06} - n)\Psi_{\lambda,n,s,m} &= 0, \\ [\mathbf{M}^2 - s(s+1)]\Psi_{\lambda,n,s,m} &= 0, \\ (M_3 - m)\Psi_{\lambda,n,s,m} &= 0. \end{aligned} \tag{4.1}$$

Upon introduction of the parametrization

$$\begin{aligned} z_1 &= (\rho/2)^{\frac{1}{2}} e^{(i/2)(\alpha+\varphi)} \sin\theta/2, \\ z_2 &= -(\rho/2)^{\frac{1}{2}} e^{(i/2)(\alpha-\varphi)} \cos\theta/2, \end{aligned} \tag{4.2}$$

Eqs. (4.1) reduce to the following system of ordinary differential equations:

$$\left(i\frac{\partial}{\partial\alpha} - \lambda\right)\Psi_{\lambda,n,s,m} = 0, \quad \left(-i\frac{\partial}{\partial\varphi} - m\right)\Psi_{\lambda,n,s,m} = 0,$$

$$\left[-\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}(\lambda^2 + m^2 - 2m\lambda\cos\theta) - s(s+1)\right] \times \Psi_{\lambda,n,s,m} = 0,$$

$$\left[\frac{\partial^2}{\partial\rho^2} + \frac{2}{\rho}\frac{\partial}{\partial\rho} - \left\{\frac{1}{4} + \frac{s(s+1)}{\rho^2} - \frac{n}{\rho}\right\}\right]\Psi_{\lambda,n,s,m} = 0. \tag{4.3}$$

The normalized simultaneous solution of this system of equations is given by

$$\Psi_{\lambda,n,s,m} = \left\{(2s+1)\frac{(n-s-1)!}{(n+s)!}\right\}^{\frac{1}{2}} \times e^{-\frac{1}{2}\rho}\rho^s L_{n-s-1}^{(2s+1)}(\rho) e^{im\varphi} d_{\lambda m}^{(s)}(\theta) e^{-i\lambda\alpha}. \tag{4.4}$$

Here $d_{\lambda m}^{(s)}$ are the well-known generalized spherical functions associated with the $(2s+1)$-dimensional representation of $SU(2)$. They are simply related to the Jacobi polynomials.[17]

The parameters $\rho/2$, θ, φ are the polar coordinates of the vector ξ [cf. Eqs. (3.8) and (4.2)]:

$$\xi^1 = \tfrac{1}{2}\rho\sin\theta\cos\varphi, \quad \xi^2 = \tfrac{1}{2}\rho\sin\theta\sin\varphi, \quad \xi^3 = \tfrac{1}{2}\rho\cos\theta.$$

$L_n^{(\alpha)}$ is the Laguerre polynomial defined by[18]

$$L_n^{(\alpha)}(\rho) = \sum_{r=0}^{n} \binom{n+\alpha}{n-r} \frac{(-\rho)^r}{r!} \equiv \binom{\alpha+n}{n} {}_1F_1(-n, \alpha+1; \rho).$$

Equation (4.4) is the desired expression for the canonical basis.

5. TRANSFORMATION LAW OF LORENTZ COVARIANT FIELDS

In the present section we shall present proof that the ladder representations are unitarily equivalent to the representations of $SU(2,2)$ used in Refs. 1–6 for the description of massless particles (cf. Appendix). This will involve establishing the transformation law under $SU(2,2)$ of Lorentz covariant, local, massless, free fields. As is well known, such a field is associated with a pair of zero-mass representations of the Poincaré group with helicity λ and $-\lambda$, describing particles and antiparticles. The result is given in Eq. (5.9) below.

Consider the Fock space $\mathcal{H}$ created from a conformal invariant "vacuum" $|0\rangle$ by applying polynomials in smeared creation operators $a^*(\mathbf{p}, \pm\lambda)$, which satisfy either the usual Bose or Fermi rules for integer or half-odd integer λ, respectively:

$$\begin{aligned} [a(\mathbf{p}, \lambda), a^*(\mathbf{p}', \lambda')]_\pm &= \delta_{\lambda\lambda'} 2|\mathbf{p}|\,\delta^3(\mathbf{p} - \mathbf{p}'), \\ a(\mathbf{p}, \lambda)|0\rangle = 0, \quad J_{AB}|0\rangle &= 0. \end{aligned} \tag{5.1}$$

We may identify the Hilbert space of wavefunctions $F_\lambda(\mathbf{p})$, considered in Sec. 3 with that subspace of $\mathcal{H}$ which consists of "one-particle states" with helicity λ, by virtue of the isometry

$$F_\lambda \to |F_\lambda\rangle = \int \frac{d^3\mathbf{p}}{2|\mathbf{p}|} F_\lambda(\mathbf{p}) a^*(\mathbf{p}, \lambda)|0\rangle, \tag{5.2}$$

whence

$$F_\lambda(p) = \langle \mathbf{p}, \lambda \mid F_\lambda\rangle, \text{ where } \langle p, \lambda| = \langle 0|\, a(\mathbf{p}, \lambda). \tag{5.3}$$

[17] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Application* (Pergamon Press, Ltd., London, 1963) (translated from the Russian), p. 85.

[18] A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill Book Co., New York, 1953), Vol. 1.

Let us now impose on $a(p, \lambda)$ the following transformation law:

$$[a(p, \lambda), J_{AB}]_- = \partial^{(\lambda)}_{AB} a(p, \lambda). \tag{5.4a}$$

$\partial^{(\lambda)}_{AB}$ are the differential operators, defined by the right-hand side of Eq. (3.11) with ξ replaced by p, which implement the action of the generators J_{AB} in the irreducible representation space of functions $F_\lambda(\mathbf{p})$ considered in Sec. 3. The corresponding differential operator for a representation of the $\mathfrak{L}^*$ series will be denoted by $\partial^{(\lambda^*)}_{AB}$. Equation (5.4a) defines $J_{AB} = J^*_{AB}$ as (unbounded) operators in $\mathfrak{H}$, because of Eq. (5.1). They form a representation of the algebra of $SU(2, 2)$; and its restriction to the subspace of $\mathfrak{H}$ consisting of vectors of the form (5.2) is unitarily equivalent to the $\mathfrak{L}$ representation with helicity λ, which was considered in Sec. 3. This follows immediately from Eq. (5.3).

From Eqs. (5.4) and (3.11), we find the further relation

$$\begin{aligned}[a^*(p, -\lambda), J_{AB}]_- &= -\bar{\partial}^{(-\lambda)}_{AB} a^*(p, -\lambda) \\ &= \partial^{(\lambda^*)}_{AB} a^*(\mathbf{p}, -\lambda).\end{aligned} \tag{5.4b}$$

The problem of constructing Lorentz covariant, local, free massless fields from creation and annihilation operators has been solved by Weinberg.[19,20] By definition, a Lorentz covariant (finite component) field $\chi_\sigma(x)$ transforms under Poincaré transformations $x^\mu \to \Lambda^\mu_\nu x^\nu + a^\mu$ according to

$$U(a, \Lambda)\chi_\sigma(x)U(a, \Lambda)^{-1} = \sum_{\sigma'} D_{\sigma\sigma'}[\Lambda^{-1}]\chi_{\sigma'}(\Lambda x + a), \tag{5.5}$$

where $D_{\sigma\sigma'}$ is some finite-dimensional representation of $SL(2, C)$. We may restrict ourselves to irreducible representations $D^{(j_1, j_2)}_{\sigma\sigma'}$; the result will carry over to the general case immediately.

Weinberg shows that the only such massless fields (acting in a Hilbert space with positive-definite metric) are fields transforming according to a representation $D^{(j_1, j_2)}_{\sigma\sigma'}$ with $j_2 - j_1 = \lambda$. Moreover, he proves that all these fields may be written as suitable derivatives of fields transforming according to $(j, 0)$ and $(0, j)$ for $\lambda = \mp j$, respectively. We may, therefore, restrict our attention to this case. Then, the field is given by

$$\chi_\sigma(x) = (2\pi)^{-\frac{3}{2}} \int d^4p \delta_+(p^2) u_\sigma(\mathbf{p}, \lambda) \times [a(\mathbf{p}, \lambda)e^{-ipx} + a^*(p, -\lambda)e^{ipx}],$$

with $\delta_+(p^2) = \delta(p^2)\theta(p^0)$ and

$$\begin{aligned}u_\sigma(\mathbf{p}, \lambda) &= [2\,|\mathbf{p}|]^j D^{(j)}_{\sigma\lambda}[R(\hat{\mathbf{p}})] \\ &= [2\,|\mathbf{p}|]^j D^{(j)}_{\sigma\lambda}(\varphi + \tfrac{1}{2}\pi, \theta, 0), \\ &\qquad j = |\lambda|.\end{aligned} \tag{5.6}$$

$R(\hat{\mathbf{p}})$ is a (suitably standardized) 3 rotation which takes the z axis into the unit vector $\hat{\mathbf{p}} = \mathbf{p}/|\mathbf{p}|$ with polar coordinates

$$\hat{\mathbf{p}}^1 = \sin\theta\cos\varphi, \quad \hat{\mathbf{p}}^2 = \sin\theta\sin\varphi, \quad \hat{\mathbf{p}}^3 = \cos\theta. \tag{5.7}$$

Its Euler angles may be chosen as $\varphi + \frac{1}{2}\pi$, θ, 0, whence the second equation for u_σ, $D^{(j)}_{\sigma\lambda}$ are the rotation functions for the $(2j + 1)$-dimensional irreducible representation of $SU(2)$.

The field satisfies the equations

$$\left(\mathbf{J}^{(j)} \cdot \nabla + \lambda \frac{\partial}{\partial x^0}\right)\chi(x) = 0, \tag{5.8a}$$

$$\Box\chi(x) = 0. \tag{5.8b}$$

$\mathbf{J}^{(j)}$ is the usual $(2j + 1)$-dimensional representation of angular momentum

$$\begin{aligned}[J_1 \pm iJ_2]_{\sigma\sigma'} &= \delta_{\sigma', \sigma\pm1}[(j \mp \sigma)(j \pm \sigma + 1)]^{\frac{1}{2}}, \\ [J_3]_{\sigma\sigma'} &= \sigma\delta_{\sigma,\sigma'}.\end{aligned}$$

The representations of $SU(2, 2)$ which have been used for the description of massless particles are defined by the following field transformation law:

$$\begin{aligned}[\chi_\sigma(x), P_\mu]_- &= i\partial_\mu\chi_\sigma(x), \\ [\chi_\sigma(x), M_{\mu\nu}]_- &= \{i(x_\mu\partial_\nu - x_\nu\partial_\mu)\delta_{\sigma\sigma'} + [\Sigma_{\mu\nu}]_{\sigma\sigma'}\}\chi_{\sigma'}(x), \\ [\chi_\sigma(x), D]_- &= i(l - x_\nu\partial^\nu)\chi_\sigma(x), \\ [\chi_\sigma(x), K_\mu]_- &= \{i(-2lx_\mu + 2x_\mu x_\nu\partial^\nu - x^2\partial_\mu)\delta_{\sigma\sigma'} \\ &\qquad + 2x^\nu[\Sigma_{\mu\nu}]_{\sigma\sigma'}\}\chi_{\sigma'}(x).\end{aligned} \tag{5.9}$$

Here $\Sigma_{\mu\nu}$ is the generator of $SL(2, C)$ in the (i_1, j_2) representation, i.e., in the present case of $(0, j)$ and $(j, 0)$ fields,

$$\Sigma_{ij} = \epsilon_{ijk}J_k, \quad \Sigma_{0k} = -i(\lambda/j)J_k, \tag{5.10}$$

and

$$l = -1 - j.$$

It has been shown[3] that Eq. (5.9) is, in fact, the most general *local* $SU(2, 2)$-transformation law for a field that transforms under Lorentz transformations according to a finite-dimensional, *irreducible* representation $D^{(j_1, j_2)}$ of $SL(2, C)$ (see the Appendix).

Now we proceed to the equivalence between the transformation law (5.9) and (5.4) for a field defined

[19] S. Weinberg, Phys. Rev. **134**, B882 (1964).
[20] S. Weinberg, Phys. Rev. **138**, B988 (1965).

by Eq. (5.6). In other words, we shall show that

$$[\chi_\sigma(x), J_{AB}]_- = (2\pi)^{-\frac{3}{2}}\int d^4p\delta_+(p^2)u_\sigma(\mathbf{p},\lambda) \times [e^{-ipx}\partial^{(\lambda)}_{AB}a(\mathbf{p},\lambda) + e^{ipx}\partial^{(\lambda^*)}_{AB}a^*(\mathbf{p},-\lambda)], \quad (5.11)$$

where the left-hand side is defined by Eqs. (5.9) and (3.1).

For the generators P_μ, Eq. (5.11) is obviously true. It suffices then to check Eq. (5.11) for the generator $K_0 = J_{06} - J_{05}$, since all other generators can be expressed in terms of multiple commutators of these. The proof is then an elementary, though somewhat tedious, exercise in properties of rotation functions. For the benefit of the reader let us sketch the calculation for the annihilation part $\chi^{(+)}_\sigma$ of a field, for positive helicity $\lambda = +j$. From Eq. (5.9) we find, using well-known properties of the δ distribution,

$$[\chi^{(+)}_\sigma(x), K_0]_- = (2\pi)^{-\frac{3}{2}}\int d^4p\delta_+(p^2)e^{-ipx} \times \left\{\left(-|\mathbf{p}|\Delta - \frac{2j}{|\mathbf{p}|}\right)\delta_{\sigma\sigma'} + 2[J_k]_{\sigma\sigma'}\frac{\partial}{\partial p^k}\right\} \times u_{\sigma'}(p,j)a(p,j).$$

As a consequence of (5.6), the u_σ satisfy

$$(\mathbf{J}\cdot\mathbf{p} - j|\mathbf{p}|)u(\mathbf{p},j) = 0,$$

$$\left(|\mathbf{p}|\frac{\partial}{\partial|\mathbf{p}|} - j\right)u_\sigma(\mathbf{p}\cdot j) = 0,$$

$$\frac{\partial}{\partial\theta}u_\sigma(\mathbf{p},j) = -i[-\sin\varphi J_1 + \cos\varphi J_2]_{\sigma\sigma'}u_{\sigma'} = \frac{1}{\sin\theta}(j\cos\theta - \sigma)u_\sigma(\mathbf{p},j), \quad (5.12)$$

$$\frac{\partial}{\partial\varphi}u_\sigma(\mathbf{p},j) = -i[J_3]_{\sigma\sigma'}u_{\sigma'} = -i\sigma u_\sigma(\mathbf{p},j).$$

Using these relations, it is readily established that

$$[\chi^{(+)}_\sigma(x), K_0]_- = (2\pi)^{-\frac{3}{2}}\int d^4p e^{-ipx}\delta_+(p^2)u_\sigma(\mathbf{p},j) \times \left\{-|\mathbf{p}|\Delta + 2i\frac{\cos\theta}{|\mathbf{p}|\sin^2\theta}j\frac{\partial}{\partial\varphi} + \frac{1}{|\mathbf{p}|\sin^2\theta}j^2\right\}a(\mathbf{p},j).$$

The differential operator in the braces agrees, indeed, with (3.11) and (3.14), which completes the proof of our statement.

Finally, let us say a word about the vector potential A_μ for a massless spin-1 particle. As discussed in detail by Weinberg, it is not a Lorentz covariant field in the sense used above, but its Lorentz transformation law differs from (5.5) by a gauge transformation. The vector potential $A_\mu(x)$ is defined in terms of the field strengths, up to a gauge transformation, by

$$\partial_\nu A_\mu - \partial_\mu A_\nu = F_{\mu\nu}. \quad (5.13)$$

$F_{\mu\nu}$ is expressed in terms of the electric and magnetic field strengths $\mathbf{E}$ and $\mathbf{B}$ in the usual way. $\mathbf{E} - i\mathbf{B}$ and $\mathbf{E} + i\mathbf{B}$ are the massless fields with helicity $\lambda = -1$ and $\lambda = +1$, respectively, as were considered above, and Eqs. (5.8a) are Maxwell's equations. From Eq. (5.13) and the transformation law (5.9) for $F_{\mu\nu}$, one checks that $A_\mu(x)$ also transforms according to Eq. (5.9) up to a gauge transformation, with $l = -1$. $\Sigma_{\mu\nu}$ are then the generators of $SL(2, C)$ in the $(\frac{1}{2}, \frac{1}{2})$ representation.

An analogous statement holds for the spin-2 (gravitational) potential, again $l = -1$. We refer the reader to Weinberg's paper[20] for a detailed discussion of the gravitational potential.

ACKNOWLEDGMENTS

The authors are grateful to Professor A. Salam and Professor P. Budini and to the IAEA for the hospitality extended to them at the International Center for Theoretical Physics, Trieste, where the main result of this paper was obtained. The second named author (I. T.) would like to thank Dr. C. Kaysen for his kind hospitality at the Institute for Advanced Study (Princeton), where part of the work was done. The authors are also indebted to Professor B. Kurşunoğlu for his hospitality at the Center for Theoretical Studies of the University of Miami, where the paper was completed.

APPENDIX: THE CONFORMAL GROUP OF SPACE-TIME IN QUANTUM FIELD THEORY

The conformal group of space-time is compounded from coordinate transformations as follows:

(a) inhomogeneous Lorentz transformations,
(b) dilatations $x'_\mu = \rho x_\mu$, $\rho > 0$,
(c) special conformal transformations[21]

$$x'_\mu = \sigma^{-1}(x)(x_\mu - c_\mu x^2),$$

where

$$\sigma(x) = 1 - 2cx + c^2x^2.$$

This is the largest continuous group which leaves the light cone invariant. The generators D of dilatation,

[21] They may be written as $x'_\mu = Rt_cRx_\mu$, where R is the inversion $Rx_\mu = -x_\mu/x^2$ and t_c stands for the translation $x'_\mu = x_\mu + c_\mu$. We stress, however, that R does not belong to the proper conformal group.

K_μ of special conformal transformations, and P_μ, $M_{\mu\nu}$ of the Poincaré group satisfy the following CR:

$$\begin{aligned}
[D, P_\mu] &= iP_\mu, \\
[D, M_{\mu\nu}] &= 0, \\
[D, K_\mu] &= -iK_\mu, \\
[K_\mu, K_\nu] &= 0, \\
[K_\mu, P_\nu] &= 2i(g_{\mu\nu}D - M_{\mu\nu}), \\
[K_\lambda, M_{\mu\nu}] &= i(g_{\lambda\mu}K_\nu - g_{\lambda\nu}K_\mu), \\
[P_\lambda, M_{\mu\nu}] &= i(g_{\lambda\mu}P_\nu - g_{\lambda\nu}P_\mu), \\
[M_{\kappa\lambda}, M_{\mu\nu}] &= i(g_{\lambda\mu}M_{\kappa\nu} - g_{\kappa\mu}M_{\lambda\nu} \\
&\quad - g_{\lambda\nu}M_{\kappa\mu} + g_{\kappa\nu}M_{\lambda\mu}).
\end{aligned} \tag{A1}$$

Note that the special conformal transformations do not transform momentum eigenstates into momentum eigenstates, since $[K_\mu, P_\rho]$ does not commute with the momenta P_ρ. The relation

$$e^{i\alpha D}P^2e^{-i\alpha D} = e^{-2\alpha}P^2 \tag{A2}$$

implies that the mass-squared spectrum contained in a unitary representation of $SU(2, 2)$ either covers (at least) a whole real semiaxis or consists of the zero point only.

Let us now consider a (quantum) field which transforms according to a representation of the conformal group, i.e.,

$$(T(g)\chi)_\alpha(x) = S_{\alpha\beta}(g, x)\chi_\beta(g^{-1}x). \tag{A3}$$

Here g acts on x as indicated in (a)–(c).

The little group which leaves $x = 0$ invariant is given by dilatations, special conformal transformations, and homogeneous Lorentz transformations. As is seen from Eq. (A1), it is isomorphic to an inhomogeneous Lorentz group plus dilatations, i.e.,

$$(SL(2, C) \otimes \{D\}) \cdot T_4. \tag{A4}$$

The "translations" T_4 are the image of the special conformal transformations. Let the generators of this little group be denoted by δ, κ_μ, and $\Sigma_{\mu\nu}$, respectively. By the standard theory of induced representations, one finds[3]

$$\begin{aligned}
P_\mu\chi(x) &= i\partial_\mu\chi(x), \\
M_{\mu\nu}\chi(x) &= \{i(x_\mu\partial_\nu - x_\nu\partial_\mu) + \Sigma_{\mu\nu}\}\chi(x), \\
D\chi(x) &= \{-ix^\nu\partial_\nu + \delta\}\chi(x), \\
K_\mu\chi(x) &= \{i(2x_\mu x_\nu\partial^\nu - x^2\partial_\mu \\
&\quad + 2ix^\nu[g_{\mu\nu}\delta - \Sigma_{\mu\nu}]) + \kappa_\mu\}\chi^{(x)}.
\end{aligned} \tag{A5}$$

The action of finite transformations is given by Eq. (A3) with

$$S_{\alpha\beta}(g, x) = D_{\alpha\beta}(t_x g t_x^{-1}).$$

Here D is the representation of the little group (A4) whose generators are δ, κ_μ, $\Sigma_{\mu\nu}$. The translation which takes the point 0 of Minkowski space into x is t_x. Two choices of the representation of the little group (A4) seem to be of particular interest:

(1) finite-dimensional representations with $\kappa_\mu \equiv 0$ and $\delta = il1$,

(2) infinite-dimensional unitary representations with $\kappa_\mu \neq 0$.

We wish to make a comment on case (2) first. In this case, $\chi(x)$ is an infinite component field. Recently, infinite-component field theories have been investigated[8] which have the following property: for fixed space–time coordinate x, $x = 0$, say, the components of the field span an irreducible representation space of the algebra of $SU(2, 2)$. The representations used are the ladder representations. The auxiliary (index $-$) $SU(2, 2)$ is, of course, not identical with the conformal group of space–time, as it does not act on the space–time coordinates. On the other hand, let us assume that the conformal structure of space–time reflects itself in the fact that the fields $\chi(x)$ form a representation space for the algebra of the conformal group of space–time. The resulting little group which acts on the indices only is then smaller than $SU(2, 2)$ and is given by (A5). However, the result of the present paper tells us that the ladder representations of the above-mentioned index $-$ $SU(2, 2)$ remain irreducible when restricted to the subgroup (A4),[22] because this group contains an inhomogeneous Lorentz group. Thus, the conformal structure of space–time provides a new motivation for using fields that transform according to a reducible representation of the "spin" $-$ $SL(2, C)$. Moreover, for a special choice of representations of the little group (A4) (corresponding to $\kappa_\mu\kappa^\mu = 0$), we arrive at precisely the same reducible representations of the "spin" $-$ $SL(2, C)$ as have been used previously with the H atom as motivation.

Case (1) has been investigated in some detail.[1–6] Here we have to do with ordinary finite-component fields. This is the case which we discussed in Sec. 5. It is seen from Eq. (A5) and the CR of $\Sigma_{\mu\nu}$, κ_μ, δ that we must have $\delta = c\mathbf{1}$, $\kappa_\mu = 0$, if $\Sigma_{\mu\nu}$ form an irreducible representation of the $SL(2, C)$ algebra. This is a consequence of Schur's lemma.

[22] They are, however, not quite the only representations of $SU(2, 2)$ with this property. L. Castell [Nucl. Phys. **B4**, 343 (1967)] has shown that the same is true for some other discrete degenerate representations. They belong to $\kappa_\mu\kappa^\mu \neq 0$ and spin 0. The unitary ray representations of the group (A5) have been investigated by U. Ottoson, Arkiv Fysik **33**, 523 (1967). The authors are grateful to Dr. L. Castell for helpful discussions on this and related points.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Matrices of Finite Lorentz Transformations in a Noncompact Basis. I. Discrete Series of $O(2, 1)$

N. MUKUNDA
Tata Institute of Fundamental Research, Bombay, India

(Received 10 August 1968)

We consider the problem of obtaining the matrices that represent finite group elements in unitary irreducible representations of the group $O(2, 1)$, in a basis in which the "noncompact" generator of an $O(1, 1)$ subgroup is diagonal. The discrete series of representations is treated and expressions obtained for the matrix elements of group elements belonging both to the $O(2)$ subgroup and the other $O(1, 1)$ subgroup.

INTRODUCTION

A detailed study of the structure and properties of the representations of the group $O(2, 1)$ is of considerable importance in connection with the current uses of noncompact group representations in elementary particle physics. Of these uses we may mention here the theory of infinite-component wave equations and of dynamical and noninvariance groups,[1] the development of various types of partial wave expansions at fixed momentum transfer,[2] and the representations of local current algebras.[3] Among the semisimple noncompact Lie groups, the group $O(2, 1)$ is the simplest to study; at the same time it is contained as a subgroup in every other semisimple noncompact Lie group. In this respect, its position is similar to that of the three-dimensional rotation group among all compact semisimple Lie groups.

The group $O(2, 1)$ has the same Lie algebra as the group $SU(1, 1)$ of pseudounitary complex matrices in two dimensions, the homomorphism from the latter to the former being two-to-one. All the single-valued nontrivial unitary irreducible representations (UIR's) of $SU(1, 1)$ were determined a long time ago by Bargmann.[4] These UIR's are all infinite-dimensional. The method used by Bargmann was to first determine all the Hermitian irreducible representations of the Lie algebra of $SU(1, 1)$, and then explicitly demonstrate that to each such representation there did correspond a UIR of the entire group. This analysis was carried out in a basis in the representation space, in which the generator of the compact subgroup $O(2)$ of $O(2, 1)$ was diagonal. The spectrum of eigenvalues of this generator is discrete and any two eigenvalues in a given UIR differ by an integer. Bargmann also determined, in this basis, the matrices that represent finite elements of the group. Analogous to the Eulerian decomposition of rotations in three-dimensional space, each element of $O(2, 1)$ can be written as a product of three elements, the first and third belonging to the compact subgroup $O(2)$, the second belonging to a noncompact subgroup $O(1, 1)$. Consequently, the problem of computing the matrix that represents an arbitrary element of the group, in the above-mentioned basis, reduces to that of computing the matrix that represents an arbitrary element of a given $O(1, 1)$ subgroup of $O(2, 1)$. The resulting matrix elements are closely related to the d-functions of angular-momentum theory.

Elsewhere, we have examined the structure of the UIR's of $O(2, 1)$ in a basis in which the (hyperbolic) generator of an $O(1, 1)$ subgroup is diagonalized.[5] Our considerations were essentially restricted to the Lie algebra, and the manner in which the generators act in this "noncompact" basis was elucidated. The spectrum of such a noncompact generator always consists of the entire real line and the corresponding eigenvectors are nonnormalizable "ideal" vectors. The UIR's of $SU(1, 1)$ can be broadly separated into two kinds, the discrete and the continuous; each of these can be further split up into various subclasses. These two kinds of UIR's can be very simply characterized in the following manner: in every UIR of the discrete kind, the $O(1, 1)$ generator possesses *one* eigenvector for each eigenvalue, while in every UIR

[1] See, for instance, "Session on Infinite Representations of Particles" in *Proceedings of the 1967 International Conference on Particles and Fields, University of Rochester, Rochester, N.Y.* (Interscience Publishers, New York, 1967).

[2] M. Toller, CERN Preprints TH 770 and 780, 1967.

[3] H. Bebie and H. Leutwyler, Phys. Rev. Letters **19**, 618 (1967); M. Gell-Mann, D. Horn, and J. Weyers, *Proceedings of the International Conference on Particle Physics, Heidelberg, September, 1967.*

[4] V. Bargmann, Ann. Math. **48**, 568 (1947). Other recent papers on this group include: A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) **A287**, 532 (1965); A. Kihlberg, Arkiv Fysik **30**, 121 (1965); W. J. Holman III and L. C. Biedenharn, Ann. Phys. **39**, 1 (1966); N. Mukunda, J. Math. Phys. **8**, 2210 (1967); **9**, 417 (1968); J. G. Kuriyan, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **9**, 2100 (1968); A. O. Barut and E. C. Phillips, Commun. Math. Phys. **8**, 52 (1968).

[5] N. Mukunda, Ref. 4. We shall refer to the first of these papers as (A).

of the continuous kind, it possesses *two* linearly independent eigenvectors for each eigenvalue.[6]

We here extend the above-mentioned work and determine explicitly the "matrices" that represent finite elements of $SU(1, 1)$ in a noncompact basis in which an $O(1, 1)$ generator is diagonal. Let us refer to the three independent elements of the Lie algebra as J_0, J_1, and J_2, where the first one is the generator of the $O(2)$ subgroup, while the last two are the two independent $O(1, 1)$ generators. We shall compute the matrices that represent elements on the two one-parameter subgroups generated by J_0 and J_1, respectively, in a basis with J_2 diagonal. It is obvious that, if an element in $O(2, 1)$ can be written in either the form $e^{i\zeta J_2}e^{i\nu J_1}e^{i\zeta' J_2}$ or the form $e^{i\xi J_2}e^{i\mu J_0}e^{i\xi' J_2}$, then its representative matrix in the noncompact basis would be particularly simple. It is interesting to note that the sets of elements which can be expressed in these two ways are mutually exclusive (except for the case $\nu = \mu = 0$), and, taken together, they do not exhaust the entire group. That is to say, there is a set of group elements which cannot be expressed in either of the two ways given above. To compute their representative matrices, one could go back to the "Euler" decomposition valid for *all* elements and then express their matrices essentially as the product of two matrices representing elements of the type $e^{i\mu J_0}$ together with an integration over a complete set of eigenstates of J_2.

As we have said earlier, there is a characteristic difference in the eigenvalue and eigenvector structure of J_2 in the discrete UIR's, on the one hand, and the continuous UIR's, on the other. The former are in one sense more simple, and in another sense more complicated, than the latter. They are more simple because in them J_2 possesses only one eigenvector corresponding to each eigenvalue. They are more complicated in that their "natural" realizations in function spaces involves Hilbert spaces of analytic functions or boundary values of analytic functions, whereas the continuous UIR's can be "naturally" realized in Hilbert spaces of functions which are square-integrable (L_2) over a suitable domain.[7] Consequently, it is natural to consider these two cases separately. In the present paper, we consider the discrete UIR's, and in a succeeding one we deal with the continuous UIR's, treating both non-exceptional and exceptional types. Some of the results of this paper, namely, the matrices representing the elements $e^{i\mu J_0}$ in the discrete UIR's, have been obtained recently by Barut and Phillips.[8] However, since the methods of derivation are somewhat different and so as not to destroy the continuity of the presentation, these results are included in this paper.

In Sec. 1, we review briefly the structure of the group $SU(1, 1)$ and its Lie algebra and then go on to determine the sets of elements decomposable in one of the two forms given earlier. In Sec. 2 we determine the matrices of the elements $e^{i\nu J_1}$ in the basis with J_2 diagonal; Sec. 3 is devoted to the identical problem for the elements $e^{i\mu J_0}$. It is convenient to work in a specific realization of the discrete UIR's and we choose the one given by Gel'fand *et al.*[9] This allows a uniform treatment of all the single-valued discrete UIR's of $SU(1, 1)$, i.e., $D_k^{\pm}$, for $k = \frac{1}{2}, 1, \frac{3}{2}, \cdots$, in the notation of Bargmann. We will consider only D_k^+ in detail, the results for D_k^- being obtained from these essentially by complex conjugation.

Note Added in Proof: After completion of this paper, the book *Special Functions and the Theory of Group Representations*, by N. J. Vilenkin (American Mathematical Society, Providence, Rhode Island, 1968), has come to the author's attention. The problems treated in this and the following paper are also considered in Chap. VII of this book.

1. PARAMETRIZATION OF GROUP ELEMENTS

Elements of the group $SU(1, 1)$ are in one-to-one correspondence with two-dimensional complex pseudounitary matrices in the following way:

$$g \to \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}, \quad |\alpha|^2 - |\beta|^2 = 1. \tag{1.1}$$

The Lie algebra of $SU(1, 1)$ is spanned by the three elements J_0, J_1, J_2 obeying the following commutation rules:

$$\begin{aligned} -i[J_0, J_1] &= J_2, \\ -i[J_0, J_2] &= -J_1, \\ -i[J_1, J_2] &= -J_0. \end{aligned} \tag{1.2}$$

In every UIR of $SU(1, 1)$, the J's are linear self-adjoint operators and the Casimir invariant Q,

$$Q = J_1^2 + J_2^2 - J_0^2, \tag{1.3}$$

becomes equal to a real multiple of the identity operator. The matrices representing the generators J_j and the one parameter subgroups $e^{i\mu J_0}$, $e^{i\nu J_1}$, and

[6] The more familiar distinction between the discrete and the continuous UIR's is the statement that in the former the quadratic Casimir invariant is quantized, while in the latter it can assume a continuous set of values.

[7] See, for instance, V. Bargmann, Ref. 4.

[8] A. O. Barut and E. C. Phillips, Ref. 4.

[9] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions*, (Academic Press Inc., New York and London, 1966), Vol. 5, Chap. VII.

$e^{i\zeta J_2}$, in the defining two-dimensional representation of $SU(1, 1)$, have been listed in (A).

From (1.1) it is evident that a simple way to parametrize the elements of $SU(1, 1)$ is to introduce three real variables ξ, ϕ, ψ in the following way:

$$\alpha = \cosh \xi \cdot e^{i\phi}, \quad \beta = \sinh \xi \cdot e^{i\psi}. \tag{1.4}$$

If we choose their ranges to be

$$0 \leq \xi < \infty, \quad -\pi < \phi, \psi \leq \pi, \tag{1.5}$$

we certainly obtain all the group elements. In order to obtain them just once each, we restrict ψ to be zero when ξ vanishes. (This parametrization is trivially related to the Euler-type decomposition.[5]) So we may write an element g of $SU(1, 1)$ as

$$g = g(\xi, \phi, \psi). \tag{1.6}$$

Let us now consider the set of elements $k(\zeta, \nu, \zeta')$, where

$$k(\zeta, \nu, \zeta') = e^{i\zeta J_2} e^{i\nu J_1} e^{i\zeta' J_2}. \tag{1.7}$$

Each such element $k(\zeta, \nu, \zeta')$ is equal to a $g(\xi, \phi, \psi)$ for some set of values of ξ, ϕ, ψ. We wish to determine what part of the domain of variation of ξ, ϕ, ϕ is covered by considering the elements $k(\zeta, \nu, \zeta')$ as each of the parameters ζ, ν, ζ' varies independently from $-\infty$ to $+\infty$. To discover this, all that has to be done is to equate the two-dimensional matrix representing the element in (1.7) to that representing the element in (1.6) and to solve for ξ, ϕ, and ψ. One can check easily that the range of ξ, ϕ, ψ that is covered is symmetric under reflections in the ϕ and ψ axes separately, i.e., under $\psi \to -\psi$ and $\phi \to -\phi$. It is, therefore, sufficient to consider the region $0 \leq \phi, \psi \leq \pi$. One then finds that, for each value of ξ, the values of ϕ and ψ attainable by elements of the form $k(\zeta, \nu, \zeta')$ are bounded by a curve determined by ξ:

$$0 \leq \phi \leq \phi_0 = \sin^{-1} \tanh \xi, \quad \phi_0 < \pi/2,$$

$$-[1 - \sin^2 \phi \coth^2 \xi]^{\frac{1}{2}} \leq \cos \psi \leq +[1 - \sin^2 \phi \coth^2 \xi]^{\frac{1}{2}}. \tag{1.8}$$

This can be represented as in Fig. 1; the shaded region, together with its reflections in the ϕ and ψ axes, consists of the set of elements of the form $k(\zeta, \nu, \zeta')$. It is interesting to note, among other things, that this region does not contain any nontrivial element of the form $e^{i\mu J_0}$.

Next we consider the set of elements $h(\zeta, \mu, \zeta')$, defined as

$$h(\zeta, \mu, \zeta') = e^{i\zeta J_2} e^{i\mu J_0} e^{i\zeta' J_2}, \tag{1.9}$$

and the region in the ξ, ϕ, ψ space covered by them.

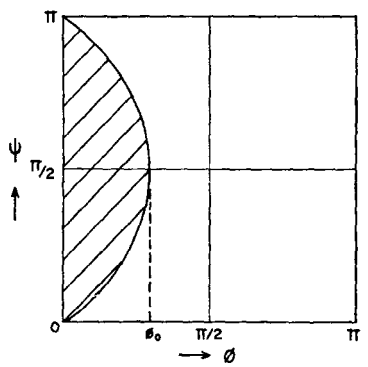

FIG. 1: ELEMENTS OF THE FORM $k(\xi, \nu, \xi')$.
($\phi_0 = \sin^{-1} \tanh \xi$)

It is clear that this region and that covered by the elements $k(\zeta, \nu, \zeta')$ must be nonoverlapping since, as we remarked earlier, the latter does not contain any nontrivial element of the form $e^{i\mu J_0}$. As before, one can check that the region involved is symmetric under both replacements $\phi \to -\phi$, $\psi \to -\psi$, and one may restrict oneself to the domain $0 \leq \phi, \psi \leq \pi$. A routine analysis then shows that the area covered by the elements $h(\zeta, \mu, \zeta')$ is bounded in the following way: For each value of ξ, all values of ϕ are attainable, while for given ξ and ϕ, ψ is constrained by the following inequalities:

$$\max(0, 1 - \sin^2 \phi \coth^2 \xi) \leq \cos^2 \psi \leq \min(1, \cos^2 \phi \coth^2 \xi). \tag{1.10}$$

(Note that these limits on $\cos^2 \psi$ are always compatible.) Thus, for ϕ lying either in the range 0 to $\sin^{-1} \tanh \xi$ or in the range $\pi - \sin^{-1} \tanh \xi$ to π, $\cos^2 \psi$ is bounded below by $1 - \sin^2 \phi \coth^2 \xi$; and for ϕ in the range $\frac{1}{2}\pi - \sin^{-1} \tanh \xi$ to $\frac{1}{2}\pi + \sin^{-1} \tanh \xi$, $\cos^2 \psi$ is bounded above by $\cos^2 \phi \coth^2 \xi$. This region can be depicted as in Fig. 2; to the shaded region must be added those obtained by the reflections $\phi \to -\phi$, $\psi \to -\psi$ in order to exhibit all the elements of the form $h(\zeta, \mu, \zeta')$.

From these considerations it is clear that, as far as elements of the two forms $k(\zeta, \nu, \zeta')$ and $h(\zeta, \mu, \zeta')$ are concerned, their representative matrices in a basis wherein J_2 is diagonal reduce essentially to those for the elements $e^{i\nu J_1}$ and $e^{i\mu J_0}$, respectively (and these will be computed in the following sections). However, if we have an element $g \in SU(1, 1)$, which is not of one of these two forms, the computation of its representative matrix involves more work. In principle, we could express such an element as

$$g = e^{i\mu J_0} e^{i\zeta J_2} e^{i\mu' J_0} \tag{1.11}$$

(which is actually a form valid for *all* elements g).

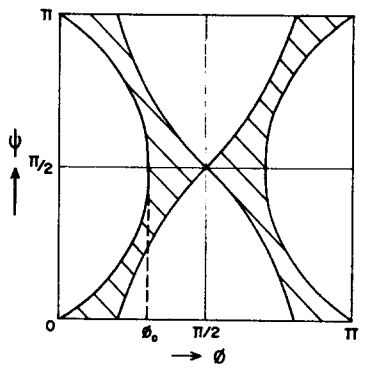

FIG.2: ELEMENTS OF THE FORM $h(\xi, \mu, \xi')$. ($\emptyset_0 = \sin^{-1}\tanh\xi$)

Then the matrix representing g appears as the integral, over intermediate eigenstates of J_2, of the product of the matrices corresponding to $e^{i\mu J_0}$ and $e^{i\mu' J_0}$, together with the factor $e^{i\zeta J_2}$, which, of course, reduces to a "plane wave." In this paper, we shall not compute the matrix corresponding to such a general element g.

2. MATRICES OF $O(1, 1)$ TRANSFORMATIONS

The single-valued discrete UIR's of $SU(1, 1)$ are generally denoted by D_k^+ and D_k^-, k running over the values $\frac{1}{2}, 1, \frac{3}{2}, \cdots$. Both in D_k^+ and in D_k^-, the Casimir invariant Q takes on the value $k(1-k)$. However, in the former, the spectrum of J_0 consists of the numbers $k, k+1, k+2, \cdots$, while in the latter it consists of the numbers $-k, -k-1, -k-2, \cdots$. The representation D_k^- is essentially the complex conjugate of D_k^+, and the same is then true for the representation matrices. We may, therefore, restrict ourselves to the UIR's of type D_k^+ only. Note that k and the eigenvalues of J_0 are simultaneously integral or simultaneously half-odd integral. Except for the fact that, at certain points, the case $k = \frac{1}{2}$ has to be treated separately, the actual calculation of representation matrices goes through in a uniform manner for all k.[10]

In the UIR D_k^+, the spectrum of the operator J_2 consists of the entire real line, and, as stated in the Introduction, there is just one eigenvector for each eigenvalue.[11] Thus we may introduce a basis in the following way:

$$J_2 |k_+; p\rangle = p\,|k_+; p\rangle,$$
$$\langle k_+; p' | k_+; p\rangle = \delta(p' - p),$$
$$-\infty < p,\quad p' < \infty. \qquad (2.1)$$

(Note that, in contrast to the generator J_0, the range of the eigenvalues of J_2 is independent of k.) In this basis, the one-parameter subgroups generated by J_1 and J_0 are represented by "matrices" with the continuous variables p' and p playing the role of row and column indices. We define

$$\langle k_+; p' | e^{i\nu J_1} | k_+; p\rangle = \mathcal{F}^{(k_+)}(p', p; \nu),$$
$$\langle k_+; p' | e^{i\mu J_0} | k_+; p\rangle = \mathcal{G}^{(k_+)}(p', p; \mu). \qquad (2.2)$$

In this section, we deal with the determination of the functions $\mathcal{F}$; in the next we deal with $\mathcal{G}$.

The function $\mathcal{F}$ obeys the following second-order differential equation with respect to the variable ν[12]:

$$\left[\frac{d^2}{d\nu^2} + \coth\nu\,\frac{d}{d\nu} + k(1-k) + \frac{p'^2 + p^2 - 2p'p\cosh\nu}{\sinh^2\nu}\right]\mathcal{F}^{(k_+)}(p', p; \nu) = 0. \qquad (2.3)$$

This equation possesses two linearly independent solutions which we shall call ϕ_1 and ϕ_2[13]:

$$\phi_1(k; p', p; \nu) = (\cosh^2\tfrac{1}{2}\nu)^{i(p'+p)/2}(\sinh^2\tfrac{1}{2}\nu)^{i(p'-p)/2} \times F(k+ip', 1-k+ip'; 1+ip'-ip; -\sinh^2\tfrac{1}{2}\nu),$$
$$\phi_2(k; p', p; \nu) = (\cosh^2\tfrac{1}{2}\nu)^{i(p+p')/2}(\sinh^2\tfrac{1}{2}\nu)^{i(p-p')/2} \times F(k+ip, 1-k+ip; 1+ip-ip'; -\sinh^2\tfrac{1}{2}\nu),$$
$$\phi_2(k; p', p; \nu) = \phi_1(k; p, p'; \nu) = \phi_1(k; p', p; \nu)^* \qquad (2.4)$$

(where F is the hypergeometric function). $\mathcal{F}$ is a linear combination of these two solutions. To determine it, we turn to the specific realization of the UIR's D_k^+ as given by Gel'fand *et al.*[9]

The Hilbert space $\mathcal{H}_k$, in which the UIR D_k^+ will be set up, consists of functions $f(x)$, which are boundary values on the real axis of functions that are analytic in the upper half-plane, and for which the inner product is defined in the following way for $k = \frac{1}{2}$:

$$(h, f)_{\frac{1}{2}} = (2\pi)^{-1}\int_{-\infty}^{\infty} dx\, h(x)^* f(x);$$

for $k > \frac{1}{2}$ we have

$$(h, f)_k = \frac{(2\pi)^{-2}e^{-i\pi k}}{\Gamma(2k-1)}\int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx\, h(x')^*(x'-x)^{2k-2} \times \ln(x'-x-i\epsilon) f(x). \qquad (2.5)$$

[10] Recall that the UIR's $D_{\frac{1}{2}}^{\pm}$ are not needed in the Plancherel formula for $SU(1, 1)$, whereas the remaining discrete UIR's are needed.

[11] For details, see (A).

[12] This may be derived by using, for example, the methods described by J. F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, "Partial Wave Analysis (Part I)," ICTP Preprint IC/67/9, Trieste, 1967.

[13] Note that both functions ϕ_1 and ϕ_2 are, by definition, *even* functions of ν.

If we pass to the Fourier transforms of $f(x)$ and $h(x)$ defined via

$$f(x) = \int_0^\infty d\lambda e^{i\lambda x}\tilde{f}(\lambda), \tag{2.6}$$

then the scalar product has a uniform appearance for all k:

$$(h,f)_k = \int_0^\infty d\lambda \lambda^{1-2k}\tilde{h}(\lambda)^*\tilde{f}(\lambda). \tag{2.7}$$

In the UIR D_k^+, the element g, corresponding to a matrix as in (1.1), is represented by a unitary operator $U(g)$ which acts on any vector $f \in \mathcal{H}_k$ as follows[14]:

$$[U(g)f](x) = (\alpha_r + \beta_r - (\alpha_i + \beta_i)x)^{-2k} \times f\left(\frac{\alpha_i - \beta_i + (\alpha_r - \beta_r)x}{\alpha_r + \beta_r - (\alpha_i + \beta_i)x}\right). \tag{2.8}$$

(α_r and α_i are the real and imaginary parts of α, similarly for β.)

We can now use Eq. (2.8) to determine the normalized eigenfunctions of J_2, corresponding to the kets defined in (2.1). They turn out to be

$$\begin{aligned} |k_+; p\rangle \to \varphi_p(x) &= (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}p\pi} |\Gamma(k + ip)| (x + i\epsilon)^{-k-ip}, \\ \tilde{\varphi}_p(\lambda) &= (2\pi)^{-\frac{1}{2}} e^{i[\eta_k(p) - \frac{1}{2}k\pi]} \lambda^{k+ip-1}, \\ \eta_k(p) &= \arg \Gamma(k - ip). \end{aligned} \tag{2.9}$$

Employing these explicit expressions and using (2.8), we can write down an integral representation for the matrix $\mathcal{F}$. We have

$$\begin{aligned} \mathcal{F}^{(k_+)}(p', p; v) = (2\pi)^{-2} e^{i[\frac{1}{2}k\pi - \eta_k(p')] - \frac{1}{2}p\pi} |\Gamma(k + ip)| \\ \times \int_0^\infty d\lambda \lambda^{-k-ip'} \int_{-\infty}^\infty dx e^{-i\lambda x} \\ \times (\cosh \tfrac{1}{2}v - (x + i\epsilon) \sinh \tfrac{1}{2}v)^{-2k} \\ \times \left(\frac{x \cosh \frac{1}{2}v - \sinh \frac{1}{2}v}{\cosh \frac{1}{2}v - x \sinh \frac{1}{2}v} + i\epsilon\right)^{-k-ip}. \end{aligned} \tag{2.10}$$

In this expression, the $i\epsilon$'s have been retained to remind us that in x space the functions we are dealing with are boundary values of functions analytic in the upper half-plane. Now, rather than explicitly evaluate this integral, it is much easier to compute the asymptotic value of $\mathcal{F}$ as, say, $v \to +\infty$, and then find out what linear combination of ϕ_1 and ϕ_2 reproduces this asymptotic behavior.[15] From (2.10) we obtain

$$\mathcal{F}^{(k_+)}(p', p; v) \xrightarrow[v \to +\infty]{} \frac{e^{\frac{1}{2}\pi(p-p')}}{2\pi} \times \frac{|\Gamma(k + ip')\Gamma(k + ip)|}{\Gamma(2k)} 4^k e^{-kv}. \tag{2.11}$$

(Note that the asymptotic form of $\mathcal{F}$ as $v \to -\infty$ is different from the one above.) After some algebra, we find that the proper linear combination of ϕ_1 and ϕ_2 for $v \geq 0$ is the following:

$$\begin{aligned} \mathcal{F}^{(k_+)}(p', p; v) \\ = e^{\frac{1}{2}\pi(p-p')}(2\pi)^{-1} \\ \times [e^{i[\eta_k(p) - \eta_k(p')]}\Gamma(ip - ip')\phi_1(k; p', p; v) \\ + e^{i[\eta_k(p') - \eta_k(p)]}\Gamma(ip' - ip)\phi_2(k; p', p; v)]. \end{aligned} \tag{2.12}$$

For negative values of v, one may either obtain the asymptotic form from (2.10) and repeat the procedure described above or, more simply, one may use the formula

$$\mathcal{F}^{(k_+)}(p', p; v) = [\mathcal{F}^{(k_+)}(p, p'; -v)]^*, \tag{2.13}$$

which follows from the unitarity of the operator $U(e^{ivJ_1})$. One then obtains, for $v \leq 0$,

$$\begin{aligned} \mathcal{F}^{(k_+)}(p', p; v) \\ = \frac{e^{\frac{1}{2}\pi(p'-p)}}{2\pi} [e^{i[\eta_k(p) - \eta_k(p')]}\Gamma(ip - ip')\phi_1(k; p', p; v) \\ + e^{i[\eta_k(p') - \eta_k(p)]}\Gamma(ip' - ip)\phi_2(k; p', p; v)]. \end{aligned} \tag{2.14}$$

This completes the evaluation of the "matrix elements" of the operator $U(e^{ivJ_1})$ in the basis with J_2 diagonal. One may check from (2.12) and (2.14) that, as $v \to 0$ either via positive or negative values, one has the right boundary value for $\mathcal{F}$:

$$\lim_{v \to 0} \mathcal{F}^{(k_+)}(p', p; v) = \delta(p' - p). \tag{2.15}$$

The identity needed to establish this is

$$\lim_{R \to \infty} a^{-1} \sin aR = \pi\delta(a). \tag{2.16}$$

Combining (2.11), (2.12), and (2.14), we can express the behavior of $\mathcal{F}$ as $v \to \pm\infty$ in the single formula

$$\mathcal{F}^{(k_+)}(p', p; v) \xrightarrow[v \to \pm\infty]{} \frac{e^{\pm\frac{1}{2}\pi(p-p')}}{2\pi} \frac{|\Gamma(k + ip')\Gamma(k + ip)|}{\Gamma(2k)} 4^k e^{-k|v|}. \tag{2.17}$$

[14] In Ref. 9, the formulas are appropriate to a discussion of the group $SL(2, R)$. We have transcribed them so as to express things in the language of the group $SU(1, 1)$. These two groups are, of course, isomorphic. Note that the parameter S in Ref. 9 is related to our k via $S = 1 - 2k$.

[15] For details concerning the hypergeometric functions, see, for instance, *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., New York, 1953), Vol. 1, Chap. II; N. N. Lebedev *Special Functions and their Applications* (Prentice-Hall, Inc. Englewood Cliffs, N.J., 1965).

We denote the corresponding "matrix elements" of $e^{i\nu J_1}$ in the UIR D_k^- by $\mathcal{F}^{(k_-)}(p', p; \nu)$. These may be taken to be just the complex conjugates of $\mathcal{F}^{(k_+)}(p', p; \nu)$ together with a change in the sign of ν. However, from (2.12) and (2.14), we see that the latter functions are real. Hence, we have

$$\mathcal{F}^{(k_-)}(p', p; \nu) = \mathcal{F}^{(k_+)}(p', p; -\nu). \tag{2.18}$$

3. MATRICES OF $O(2)$ TRANSFORMATIONS

We now turn to the evaluation of the matrix elements of $e^{i\mu J_0}$ namely, the functions $\mathcal{G}^{(k_+)}(p', p; \mu)$. Similar to (2.3), one can establish the following differential equation for $\mathcal{G}$[12]:

$$\left[\frac{d^2}{d\mu^2} + \cot\mu\,\frac{d}{d\mu} - k(1-k) + \frac{p'^2 + p^2 - 2p'p\cos\mu}{\sin^2\mu}\right]\mathcal{G}^{(k_+)}(p', p; \mu) = 0. \tag{3.1}$$

Let us define two linearly independent solutions ψ_1 and ψ_2 to be[16]

$$\begin{aligned}
&\psi_1(k; p', p; \mu)\\
&= (\cos^2\tfrac{1}{2}\mu)^{\frac{1}{2}i(p'+p)}(\sin^2\tfrac{1}{2}\mu)^{\frac{1}{2}i(p'-p)}\\
&\quad\times F(k + ip', 1 - k + ip'; 1 + ip' - ip; \sin^2\tfrac{1}{2}\mu),\\
&\psi_2(k; p', p; \mu)\\
&= (\cos^2\tfrac{1}{2}\mu)^{\frac{1}{2}i(p+p')}(\sin^2\tfrac{1}{2}\mu)^{\frac{1}{2}i(p-p')}\\
&\quad\times F(k + ip, 1 - k + ip; 1 + ip - ip'; \sin^2\tfrac{1}{2}\mu),\\
&\psi_2(k; p', p; \mu) = \psi_1(k; p, p'; \mu) = \psi_1(k; p', p; \mu)^*.
\end{aligned} \tag{3.2}$$

Then, $\mathcal{G}$ is a certain linear combination of ψ_1 and ψ_2.

Let us now use the realization of the UIR's D_k^+ described in the previous section. We can show quite easily that

$$U(e^{i\pi J_0})\,|k_+; p\rangle = e^{i\pi k}\,|k_+; -p\rangle, \tag{3.3}$$

which is essentially what we would expect to find. Using this result, we can show that

$$\mathcal{G}^{(k_+)}(p', p; \pi + \mu) = e^{i\pi k}\mathcal{G}^{(k_+)}(p', -p; \mu). \tag{3.4}$$

This means that it suffices to compute $\mathcal{G}$ for the range $0 \le \mu \le \pi$; its values elsewhere can be obtained by use of (3.4). Analogous to (2.10), we can write down an integral representation for $\mathcal{G}$:

$$\begin{aligned}
\mathcal{G}^{(k_+)}(p', p; \mu) &= (2\pi)^{-2}e^{i[\frac{1}{2}k\pi - \eta_k(p')] - \frac{1}{2}p\pi}\,|\Gamma(k + ip)|\\
&\quad\times\int_0^\infty d\lambda\lambda^{-k-ip'}\int_{-\infty}^{\infty}dx e^{-i\lambda x}\\
&\quad\times(\cos\tfrac{1}{2}\mu - (x + i\epsilon)\sin\tfrac{1}{2}\mu)^{-2k}\\
&\quad\times\left(\frac{x\cos\frac{1}{2}\mu + \sin\frac{1}{2}\mu}{\cos\frac{1}{2}\mu - x\sin\frac{1}{2}\mu} + i\epsilon\right)^{-k-ip}
\end{aligned} \tag{3.5}$$

We could now use this representation to extract the "asymptotic behavior" of $\mathcal{G}$; but this time, we have to go outside of the group to do this. Guided by the fact that the factors appearing on the right-hand side of (3.5) are boundary values for real x of functions analytic in the *upper* half-plane, we make $\cos\frac{1}{2}\mu$ and $\sin\frac{1}{2}\mu$ go to infinity in the following way:

$$\cos\tfrac{1}{2}\mu = \Lambda, \quad \sin\tfrac{1}{2}\mu \cong +i\Lambda, \quad \Lambda \to +\infty. \tag{3.6}$$

This analytic continuation corresponds to the following: We start with a "sensible" value of μ in the range $0 < \mu < \pi$, such that $0 < \Lambda < 1$. We now make Λ go to infinity along the positive real axis avoiding, however, the point $\Lambda = 1$ by making a small detour into the *lower* half of the complex plane around $\Lambda = 1$. This specifies the way in which $\sin\frac{1}{2}\mu$ is to be continued and is such that the functions of x appearing on the right-hand side in (3.5) always have singularities only in the *lower* half-plane. With this prescription, (3.5) yields

$$\mathcal{G}^{(k_+)}(p', p; \mu) \xrightarrow[\Lambda\to+\infty]{} \frac{1}{2\pi}\,\frac{|\Gamma(k + ip')\Gamma(k + ip)|}{\Gamma(2k)}\,\Lambda^{-2k}. \tag{3.7}$$

As before, we can now determine the proper linear combination of the functions ψ_1 and ψ_2 that will reproduce this asymptotic behavior.[15] In this way we find, for $0 \le \mu \le \pi$,

$$\begin{aligned}
\mathcal{G}^{(k_+)}(p', p; \mu) &= \frac{e^{\frac{1}{2}\pi(p'-p)}}{2\pi}\,e^{i[\eta_k(p) - \eta_k(p')]}\\
&\quad\times\Gamma(ip - ip')\psi_1(k; p', p; \mu)\\
&\quad+ (2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p') - \eta_k(p)]}\\
&\quad\times\Gamma(ip' - ip)\psi_2(k; p', p; \mu).
\end{aligned} \tag{3.8}$$

At the limiting values $\mu = 0$ and π, we recover the expected boundary values of $\mathcal{G}$:

$$\begin{aligned}
\mathcal{G}^{(k_+)}(p', p; 0) &= \delta(p' - p),\\
\mathcal{G}^{(k_+)}(p', p; \pi) &= e^{i\pi k}\delta(p' + p).
\end{aligned} \tag{3.9}$$

To evaluate $\mathcal{G}$ for other values of μ, we now use (3.4).

[16] Both ψ_1 and ψ_2 are, by definition, *even* functions of μ.

We obtain, for $\pi \leq \mu \leq 2\pi$,

$$\begin{aligned}\mathfrak{G}^{(k_+)}(p', p; \mu) = {} & e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu). \quad (3.10)\end{aligned}$$

This completes the evaluation of $\mathfrak{G}$ for the cases when k is integral, for then the relevant range of μ is just $0 \leq \mu \leq 2\pi$. For the cases when k is half an odd integer, we need only realize that increasing μ by 2π amounts to an over-all sign change of $\mathfrak{G}$ in such UIR's.

Last of all, we evaluate the $\mathfrak{G}$ functions for the UIR's D_k^-. These are just complex conjugates of what we have obtained above. So, for $0 \leq \mu \leq \pi$, we find

$$\begin{aligned}\mathfrak{G}^{(k_-)}(p', p; \mu) = {} & (2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + (2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu)\end{aligned}$$

and, for $\pi \leq \mu \leq 2\pi$,

$$\begin{aligned}\mathfrak{G}^{(k_-)}(p', p; \mu) = {} & e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu). \quad (3.11)\end{aligned}$$

It is interesting to see that, whereas the spectra of the noncompact generators J_1, J_2, [but not the matrices representing the $O(1, 1)$ elements $e^{i\nu J_1}$] do not distinguish between the two kinds of discrete UIR's D_k^+ and D_k^-, there is a characteristic difference in the matrices representing the elements $e^{i\mu J_0}$ of the compact $O(2)$ subgroup. Such a difference is definitely expected, because the generator J_0 has different eigenvalues in the two cases.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Matrices of Finite Lorentz Transformations in a Noncompact Basis. II. Continuous Series of $O(2, 1)$

N. MUKUNDA
Tata Institute of Fundamental Research, Bombay, India

(Received 19 August 1968)

The representation matrices in a noncompact basis for finite elements of the group $O(2, 1)$ are determined in the continuous classes of unitary irreducible representations. Integral as well as half-integral, and exceptional as well as nonexceptional, representations are treated.

INTRODUCTION

In the previous paper,[1] referred to hereafter as I, we have determined the matrices that represent finite elements of the three-dimensional Lorentz group $O(2, 1)$ in any unitary irreducible representation (UIR) of the discrete classes, and in a basis in which a "noncompact" (hyperbolic) generator of an $O(1, 1)$ subgroup is diagonal. Here we carry out a similar determination of the representation matrices in the remaining classes of UIR's, namely, the continuous nonexceptional and the continuous exceptional classes.

In any nontrivial UIR of $O(2, 1)$, the generator of the $O(1, 1)$ subgroup has a continuous spectrum, every real number appearing as an eigenvalue. Correspondingly, the eigenvectors are all nonnormalizable. In the discrete class UIR's there is just one eigenvector corresponding to each eigenvalue, while in the continuous class UIR's there are two linearly independent eigenvectors for each eigenvalue. The determination of representation matrices that was carried out in (I) suffers from the inherent ambiguity that results from the freedom to change at will the relative phases of the orthonormal eigenvectors of the noncompact $O(1, 1)$ generator. We therefore chose to carry out the calculations using a specific realization of the UIR's involved. In the UIR's of the continuous classes, this type of ambiguity is increased because we have two independent eigenvectors for each eigenvalue of the hyperbolic generator, and one is therefore free to make two-dimensional unitary transformations mixing the pair of eigenvectors that go with each eigenvalue; one could go even further and make these transformations dependent on the eigenvalue. We resolve this ambiguity by once again working with specific realizations

[1] N. Mukunda, J. Math. Phys. **10**, 2086 (1969), preceding paper. See this paper for further references.

We obtain, for $\pi \leq \mu \leq 2\pi$,

$$\begin{aligned}\mathfrak{G}^{(k_+)}(p', p; \mu) = {} & e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu). \quad (3.10)\end{aligned}$$

This completes the evaluation of $\mathfrak{G}$ for the cases when k is integral, for then the relevant range of μ is just $0 \leq \mu \leq 2\pi$. For the cases when k is half an odd integer, we need only realize that increasing μ by 2π amounts to an over-all sign change of $\mathfrak{G}$ in such UIR's.

Last of all, we evaluate the $\mathfrak{G}$ functions for the UIR's D_k^-. These are just complex conjugates of what we have obtained above. So, for $0 \leq \mu \leq \pi$, we find

$$\begin{aligned}\mathfrak{G}^{(k_-)}(p', p; \mu) = {} & (2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + (2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu)\end{aligned}$$

and, for $\pi \leq \mu \leq 2\pi$,

$$\begin{aligned}\mathfrak{G}^{(k_-)}(p', p; \mu) = {} & e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p'-p)}e^{i[\eta_k(p)-\eta_k(p')]} \\ & \times \Gamma(ip - ip')\psi_1(k; p', p; \mu) \\ & + e^{2\pi ik}(2\pi)^{-1}e^{\frac{1}{2}\pi(p-p')}e^{i[\eta_k(p')-\eta_k(p)]} \\ & \times \Gamma(ip' - ip)\psi_2(k; p', p; \mu). \quad (3.11)\end{aligned}$$

It is interesting to see that, whereas the spectra of the noncompact generators J_1, J_2, [but not the matrices representing the $O(1, 1)$ elements $e^{i\nu J_1}$] do not distinguish between the two kinds of discrete UIR's D_k^+ and D_k^-, there is a characteristic difference in the matrices representing the elements $e^{i\mu J_0}$ of the compact $O(2)$ subgroup. Such a difference is definitely expected, because the generator J_0 has different eigenvalues in the two cases.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Matrices of Finite Lorentz Transformations in a Noncompact Basis. II. Continuous Series of $O(2, 1)$

N. MUKUNDA
Tata Institute of Fundamental Research, Bombay, India

(Received 19 August 1968)

The representation matrices in a noncompact basis for finite elements of the group $O(2, 1)$ are determined in the continuous classes of unitary irreducible representations. Integral as well as half-integral, and exceptional as well as nonexceptional, representations are treated.

INTRODUCTION

In the previous paper,[1] referred to hereafter as I, we have determined the matrices that represent finite elements of the three-dimensional Lorentz group $O(2, 1)$ in any unitary irreducible representation (UIR) of the discrete classes, and in a basis in which a "noncompact" (hyperbolic) generator of an $O(1, 1)$ subgroup is diagonal. Here we carry out a similar determination of the representation matrices in the remaining classes of UIR's, namely, the continuous nonexceptional and the continuous exceptional classes.

In any nontrivial UIR of $O(2, 1)$, the generator of the $O(1, 1)$ subgroup has a continuous spectrum, every real number appearing as an eigenvalue. Correspondingly, the eigenvectors are all nonnormalizable. In the discrete class UIR's there is just one eigenvector corresponding to each eigenvalue, while in the continuous class UIR's there are two linearly independent eigenvectors for each eigenvalue. The determination of representation matrices that was carried out in (I) suffers from the inherent ambiguity that results from the freedom to change at will the relative phases of the orthonormal eigenvectors of the noncompact $O(1, 1)$ generator. We therefore chose to carry out the calculations using a specific realization of the UIR's involved. In the UIR's of the continuous classes, this type of ambiguity is increased because we have two independent eigenvectors for each eigenvalue of the hyperbolic generator, and one is therefore free to make two-dimensional unitary transformations mixing the pair of eigenvectors that go with each eigenvalue; one could go even further and make these transformations dependent on the eigenvalue. We resolve this ambiguity by once again working with specific realizations

[1] N. Mukunda, J. Math. Phys. **10**, 2086 (1969), preceding paper. See this paper for further references.

of the UIR's involved. The choice of eigenvectors is motivated by simplicity in these realizations, and the representation matrices we compute may differ from alternate computations by a transformation of the type described above.

We shall actually consider single-valued UIR's of the group $SU(1,1)$, which amounts to considering both single- and double-valued UIR's of $O(2,1)$; these are generally the UIR's relevant in physical applications where the $O(2,1)$ group often appears as a subgroup of a larger group like $O(3,1)$. In Sec. 1, we set up the notation, and then consider the continuous nonexceptional UIR's of integral type. Section 2 deals with the continuous UIR's of half-integral type and the concluding Sec. 3 with the continuous UIR's of the exceptional interval.

1. NONEXCEPTIONAL INTEGRAL UIR'S

The continuous class UIR's of $SU(1,1)$ break up into three types: (i) the integral nonexceptional type, wherein the quadratic Casimir invariant Q lies in the range $\frac{1}{4} \leq Q < \infty$ and the compact $O(2)$ generator J_0 has all integers from minus to plus infinity for eigenvalues; (ii) the integral exceptional type, wherein $0 < Q < \frac{1}{4}$ and the generator J_0 has the same eigenvalues as in type (i); and (iii) the half-integral type, with $\frac{1}{4} < Q < \infty$ and J_0 having all half-odd integers for eigenvalues. Conventionally, types (i) and (ii) are together denoted as C_q^0 with $q > 0$ being the value of Q, and type (iii) is denoted as $C_q^{\frac{1}{2}}$, $q > \frac{1}{4}$. Elsewhere we have shown how one can diagonalize the hyperbolic generator J_2 in UIR's of types (i) and (ii).[2] A straightforward extension of those methods accomplishes the same purpose in UIR's of type (iii).

In every one of these UIR's we can introduce a basis of eigenvectors of the $O(1,1)$ generator J_2. We shall have to use slightly different notations in the nonexceptional UIR's (integral as well as half-integral), on the one hand, and the exceptional UIR's, on the other. For the nonexceptional cases, we shall set

$$Q = \tfrac{1}{4} + s^2, \tag{1.1}$$

with $0 \leq s < \infty$ for the integral case and $0 < s < \infty$ for the half-integral case. We refer to the eigenvalues of J_2 by $p, p', \cdots$, and use the letters $a, b, \cdots$ (taking the values $+$ and $-$) to distinguish the two eigenvectors for each eigenvalue. Thus the basis vectors for a nonexceptional integral UIR will be

$$|s, 0; p, a\rangle, \quad s \geq 0, \quad -\infty < p < \infty, \quad a = +, -, \tag{1.2}$$

and they obey the following equations:

$$J_2 |s, 0; p, a\rangle = p |s, 0; p, a\rangle,$$
$$\langle s, 0; p', b \,|\, s, 0; p, a\rangle = \delta(p' - p)\delta_{ba}. \tag{1.3}$$

(The zero signifies the integral type.) For a nonexceptional half-integral UIR, the basis vectors will be

$$|s, \tfrac{1}{2}; p, a\rangle, \quad s > 0, \quad -\infty < p < \infty, \quad a = +, -, \tag{1.4}$$

and these obey equations analogous to (1.3). We define the representation matrices corresponding to the one-parameter subgroups $e^{i\nu J_1}$ and $e^{i\mu J_0}$ as follows:

$$\langle s, \epsilon; p', b| e^{i\nu J_1} |s, \epsilon; p, a\rangle = \mathfrak{F}_{ba}^{(s,\epsilon)}(p', p; \nu),$$
$$\langle s, \epsilon; p', b| e^{i\mu J_0} |s, \epsilon; p, a\rangle = \mathfrak{G}_{ba}^{(s,\epsilon)}(p', p; \mu), \quad \epsilon = 0, \tfrac{1}{2}. \tag{1.5}$$

J_0 is the "compact" $O(2)$ generator and J_1 the other $O(1,1)$ generator.

Turning to the exceptional UIR's, here we write Q in the form

$$Q = \tfrac{1}{4} - \sigma^2, \quad 0 < \sigma < \tfrac{1}{2}, \tag{1.6}$$

and the basis vectors as

$$|\sigma; p, a\rangle, \quad \langle \sigma; p', b \,|\, \sigma; p, a\rangle = \delta(p' - p)\delta_{ba}, \quad b, a = +, -. \tag{1.7}$$

The representation matrices will be

$$\langle \sigma; p', b| e^{i\nu J_1} |\sigma; p, a\rangle = \mathfrak{F}_{ba}^{(\sigma)}(p', p; \nu),$$
$$\langle \sigma; p', b| e^{i\mu J_0} |\sigma; p, a\rangle = \mathfrak{G}_{ba}^{(\sigma)}(p', p; \mu). \tag{1.8}$$

For the functions defined in (1.5), one can establish second-order differential equations in ν and μ.[3] These are the same for both choices of ϵ, and read

$$\left[\frac{d^2}{d\nu^2} + \coth\nu \frac{d}{d\nu} + \frac{1}{4} + s^2 + \frac{p'^2 + p^2 - 2p'p\cosh\nu}{\sinh^2\nu}\right] \times \mathfrak{F}_{ba}^{(s,\epsilon)}(p', p; \nu) = 0, \tag{1.9a}$$

$$\left[\frac{d^2}{d\mu^2} + \cot\mu \frac{d}{d\mu} - \frac{1}{4} - s^2 + \frac{p'^2 + p^2 - 2p'p\cos\mu}{\sin^2\mu}\right] \times \mathfrak{G}_{ba}^{(s,\epsilon)}(p', p; \mu) = 0. \tag{1.9b}$$

To conform with the conventions set up in I, we shall choose the linearly independent solutions of

[2] (a) N. Mukunda, J. Math. Phys. **8**, 2210 (1967); (b) **9**, 417 (1968).

[3] For the method of derivation, see, for example, J. F. Boyce, R. Delbourgo, A. Salam, and J. Strathdee, "Partial Wave Analysis (Part I)," ICTP Preprint IC/67/9, Trieste, 1967.

these equations to be the following: for (1.9a),

$$\begin{aligned}
&\phi_1(s;\, p',\, p;\, \nu)\\
&\quad = (\cosh^2 \tfrac{1}{2}\nu)^{i(p'+p)/2}(\sinh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p'-p)}\\
&\qquad \times F(\tfrac{1}{2} + is + ip',\, \tfrac{1}{2} - is + ip';\\
&\qquad\qquad 1 + ip' - ip;\, -\sinh^2 \tfrac{1}{2}\nu),\\
&\phi_2(s;\, p',\, p;\, \nu)\\
&\quad = (\cosh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p+p')}(\sinh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p-p')}\\
&\qquad \times F(\tfrac{1}{2} + is + ip,\, \tfrac{1}{2} - is + ip;\\
&\qquad\qquad 1 + ip - ip';\, -\sinh^2 \tfrac{1}{2}\nu),\\
&\qquad \phi_2(s;\, p',\, p;\, \nu) = \phi_1(s;\, p,\, p';\, \nu);
\end{aligned} \tag{1.10}$$

and, for (1.9b).

$$\begin{aligned}
&\psi_1(s;\, p',\, p;\, \mu)\\
&\quad = (\cos^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p'+p)}(\sin^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p'-p)}\\
&\qquad \times F(\tfrac{1}{2} + is + ip',\, \tfrac{1}{2} - is + ip';\\
&\qquad\qquad 1 + ip' - ip;\, \sin^2 \tfrac{1}{2}\mu),\\
&\psi_2(s;\, p',\, p;\, \mu)\\
&\quad = (\cos^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p+p')}(\sin^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p-p')}\\
&\qquad \times F(\tfrac{1}{2} + is + ip,\, \tfrac{1}{2} - is + ip;\\
&\qquad\qquad 1 + ip - ip';\, \sin^2 \tfrac{1}{2}\mu),\\
&\qquad \psi_2(s;\, p',\, p;\, \mu) = \psi_1(s;\, p,\, p';\, \mu).
\end{aligned} \tag{1.11}$$

Our aim is to express each of the $\mathcal{F}$ functions as linear combinations of ϕ_1 and ϕ_2, and each of the $\mathcal{G}$ functions in terms of ψ_1 and ψ_2. In the remainder of this section, we do this for the UIR's of the continuous non-exceptional integral type.

Let us set up an explicit realization of these UIR's following the work in Ref. 2(a). We have a Hilbert space $\mathcal{K}$ whose elements are pairs of functions of a real variable q running from $-\infty$ to $+\infty$:

$$f \to \begin{pmatrix} f_1(q) \\ f_2(q) \end{pmatrix}. \tag{1.12}$$

The scalar product of the vectors f and h is

$$(h, f) = \sum_{r=1}^{2} \int_{-\infty}^{\infty} dq\, h_r(q)^* f_r(q). \tag{1.13}$$

The representation corresponding to a given value of s is specified adequately by stating the way the elements of the one-parameter subgroups act on f. Letting g denote in turn the elements $e^{i\zeta J_2}$, $e^{i\nu J_1}$, and $e^{i\mu J_0}$, $U(g)$ the corresponding unitary operator, and $h = U(g)f$, we get[4] the following for $g = e^{i\zeta J_2}$:

$$h_r(q) = f_r(q + \zeta), \quad r = 1, 2. \tag{1.14}$$

[4] These formulas are obtained by extending the method described in Ref. 2(a) from the Lie algebra to finite group elements.

For $g = e^{i\nu J_1}$, $\nu \geq 0$ we have the following:

For all q, $h_1(q) = (\cosh \nu + \cosh q \sinh \nu)^{-\frac{1}{2}-is} f_1(q')$,

$$e^{q'} = (e^q + \tanh \tfrac{1}{2}\nu)/(1 + e^q \tanh \tfrac{1}{2}\nu); \tag{1.15a}$$

for $q \geq \ln \coth \frac{1}{2}\nu$ and for $q \leq \ln \tanh \frac{1}{2}\nu$:

$$h_2(q) = (\cosh q \sinh \nu - \cosh \nu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (e^q - \tanh \tfrac{1}{2}\nu)/(e^q \tanh \tfrac{1}{2}\nu - 1); \tag{1.15b}$$

for $\ln \tanh \frac{1}{2}\nu \leq q \leq \ln \coth \frac{1}{2}\nu$:

$$h_2(q) = (\cosh \nu - \cosh q \sinh \nu)^{-\frac{1}{2}-is} f_2(q'),$$
$$e^{q'} = (e^q - \tanh \tfrac{1}{2}\nu)/(1 - e^q \tanh \tfrac{1}{2}\nu). \tag{1.15c}$$

For $g = e^{i\mu J_0}$, $0 \leq \mu \leq \pi$ we have the following:

For $-\infty < q \leq \ln \cot \frac{1}{2}\mu$:

$$h_1(q) = (\cos \mu - \sinh q \sin \mu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (e^q + \tan \tfrac{1}{2}\mu)/(1 - e^q \tan \tfrac{1}{2}\mu); \tag{1.16a}$$

for $\ln \cot \frac{1}{2}\mu \leq q < \infty$:

$$h_1(q) = (\sinh q \sin \mu - \cos \mu)^{-\frac{1}{2}-is} f_2(q'),$$
$$e^{q'} = (e^q + \tan \tfrac{1}{2}\mu)/(e^q \tan \tfrac{1}{2}\mu - 1); \tag{1.16b}$$

for $\ln \tan \frac{1}{2}\mu \leq q < \infty$:

$$h_2(q) = (\cos \mu + \sinh q \sin \mu)^{-\frac{1}{2}-is} f_2(q'),$$
$$e^{q'} = (e^q - \tan \tfrac{1}{2}\mu)/(1 + e^q \tan \tfrac{1}{2}\mu); \tag{1.16c}$$

for $-\infty < q \leq \ln \tan \frac{1}{2}\mu$:

$$h_2(q) = (-\cos \mu - \sinh q \sin \mu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (\tan \tfrac{1}{2}\mu - e^q)/(1 + e^q \tan \tfrac{1}{2}\mu). \tag{1.16d}$$

Using these formulas, one can obtain the ones that are valid for the other ranges of ν and μ.

We choose the linearly independent eigenvectors of J_2 in the following way:

$$|s, 0; p, \pm\rangle \to (4\pi)^{-\frac{1}{2}} \begin{pmatrix} e^{ipq} \\ \pm e^{ipq} \end{pmatrix}. \tag{1.17}$$

For the UIR's of the exceptional type, we will see that we have to choose the basis in this way; therefore, to make things as uniform as possible, we adopt the same basis in the present case also. With this, we can use (1.15) and (1.16) to express the $\mathcal{F}$ and $\mathcal{G}$ functions in the form of integrals over q. As an example, we have

$$\begin{aligned}
&\mathcal{F}^{(s,0)}_{++}(p',\, p;\, \nu)\\
&\quad = (4\pi)^{-1} \int_{-\infty}^{\infty} dq e^{-ip'q} (\cosh \nu + \cosh q \sinh \nu)^{-\frac{1}{2}-is}\\
&\qquad \times \left(\frac{e^q + \tanh \frac{1}{2}\nu}{1 + e^q \tanh \frac{1}{2}\nu} \right)^{ip}\\
&\qquad + (4\pi)^{-1} \int_{-\infty}^{\infty} dq e^{-ip'q} |\cosh \nu - \cosh q \sinh \nu|^{-\frac{1}{2}-is}\\
&\qquad \times \left| \frac{e^q - \tanh \frac{1}{2}\nu}{1 - e^q \tanh \frac{1}{2}\nu} \right|^{ip}.
\end{aligned} \tag{1.18}$$

Similar expressions for the other $\mathcal{F}$ functions may be written down. We can now either explicitly evaluate these integrals and express them as linear combinations of ϕ_1 and ϕ_2, or alternatively we may extract the asymptotic behavior, as $\nu \to \infty$, of these integrals and then set up the appropriate linear combinations of ϕ_1 and ϕ_2 to reproduce this behavior.[5] The latter procedure is somewhat easier in the present case, and in this way we obtain the following results:

$\nu \geq 0$:

$$\begin{aligned}
&\mathcal{F}^{(s,0)}_{ba}(p', p; \nu) \\
&= (2\pi)^{-2}\Gamma(\tfrac{1}{2} + ip' + is)\Gamma(\tfrac{1}{2} - ip - is)\Gamma(ip - ip') \\
&\quad \times [\cosh \pi(p + s) + ba \cosh \pi(p' + s) \\
&\quad + ib \sinh \pi(p - p')]\phi_1(s; p', p; \nu) \\
&\quad + (2\pi)^{-2}\Gamma(\tfrac{1}{2} - ip' + is)\Gamma(\tfrac{1}{2} + ip - is)\Gamma(ip' - ip) \\
&\quad \times [\cosh \pi(p - s) + ba \cosh \pi(p' - s) \\
&\quad + ib \sinh \pi(p' - p)]\phi_2(s; p', p; \nu), \quad b, a = +, -;
\end{aligned}$$

$\nu \leq 0$:

$$\mathcal{F}^{(s,0)}_{ba}(p', p; \nu) = [\mathcal{F}^{(s,0)}_{ab}(p, p'; -\nu)]^*. \quad (1.19)$$

One can check that these expressions, which have been obtained using the knowledge of their values for large ν, have the proper boundary values for $\nu = 0$:

$$\mathcal{F}^{(s,0)}_{ba}(p', p; 0) = \delta_{ba}\delta(p' - p). \quad (1.20)$$

Turning next to the $\mathcal{G}$ functions, we give as an example the integral obtained for $\mathcal{G}_{++}$, in the range $0 \leq \mu \leq \pi$:

$$\begin{aligned}
&\mathcal{G}^{(s,0)}_{++}(p', p; \mu) \\
&\quad = (4\pi)^{-1}\int_{-\infty}^{\infty} dq e^{-ip'q} |\cos \mu - \sinh q \sin \mu|^{-\frac{1}{2}-is} \\
&\qquad \times \left| \frac{e^q + \tan \tfrac{1}{2}\mu}{1 - e^q \tan \tfrac{1}{2}\mu} \right|^{ip} \\
&\qquad + (4\pi)^{-1}\int_{-\infty}^{\infty} dq e^{-ip'q} |\cos \mu + \sinh q \sin \mu|^{-\frac{1}{2}-is} \\
&\qquad \times \left| \frac{e^q - \tan \tfrac{1}{2}\mu}{1 + e^q \tan \tfrac{1}{2}\mu} \right|^{ip}. \quad (1.21)
\end{aligned}$$

Similar expressions for the other components of $\mathcal{G}_{ba}$ can be written down. By changes of variable, these can be thrown into standard integral representations of the hypergeometric functions. The final expression of $\mathcal{G}$ in terms of ψ_1 and ψ_2 is

$0 \leq \mu \leq \pi$:

$$\begin{aligned}
&\mathcal{G}^{(s,0)}_{ba}(p', p; \mu) \\
&= (2\pi)^{-2}\Gamma(\tfrac{1}{2} + ip' + is)\Gamma(\tfrac{1}{2} - ip - is)\Gamma(ip - ip') \\
&\quad \times [\cosh \pi(p' + s) + ba \cosh \pi(p + s) \\
&\quad + ia \sinh \pi(p - p')]\psi_1(s; p', p; \mu) \\
&\quad + (2\pi)^{-2}\Gamma(\tfrac{1}{2} - ip' + is)\Gamma(\tfrac{1}{2} + ip - is)\Gamma(ip' - ip) \\
&\quad \times [\cosh \pi(p - s) + ba \cosh \pi(p' - s) \\
&\quad + ib \sinh \pi(p' - p)]\psi_2(s; p', p; \mu), \quad b, a = +, -;
\end{aligned}$$

$\pi \leq \mu \leq 2\pi$:

$$\mathcal{G}^{(s,0)}_{ba}(p', p; \mu) = [\mathcal{G}^{(s,0)}_{ab}(p, p'; 2\pi - \mu)]^*. \quad (1.22)$$

This completes the evaluation of the representation matrices for the UIR's of the integral nonexceptional class. From (1.22), we note that $\mathcal{G}$ assumes the following values for $\mu = 0$, π, and 2π:

$$\begin{aligned}
&\mathcal{G}^{(s,0)}_{ba}(p', p; 0) = \mathcal{G}^{(s,0)}_{ba}(p', p; 2\pi) = \delta_{ba}\delta(p' - p), \\
&\mathcal{G}^{(s,0)}_{ba}(p', p; \pi) = a\delta_{ba}\delta(p' + p), \quad b, a = +, -.
\end{aligned} \quad (1.23)$$

As expected, in these UIR's the element $e^{2\pi i J_0}$ is represented by the identity operator.

2. HALF-INTEGRAL UIR'S

We treat in this section the UIR's of the continuous half-integral class. The corresponding representation matrices are defined in (1.5). They obey the same differential equations, namely (1.9), as do the functions in the case of the continuous integral nonexceptional class, so that we can still use the basic set of solutions written down in (1.10) and (1.11). Since the two kinds of UIR's, integral and half-integral, share the same spectrum properties for J_2 and the same values for the Casimir operator Q, it will be interesting to see just at what point the representation matrices exhibit a difference.

Let us first describe the construction of the UIR's $C^{\frac{1}{2}}_q$ analogous to the construction in the previous section.[6] The Hilbert space $\mathcal{H}$ and the definition of the inner product are the same as before, being given by Eqs. (1.12) and (1.13). Let the real function $\xi(q)$ be defined in the following way:

$$\begin{aligned}
&\xi(q) = i \ln [(e^q + i)/(e^q - i)], \\
&\xi(-\infty) = -\pi, \quad \xi(\infty) = 0.
\end{aligned} \quad (2.1)$$

Then, continuing to denote the vector $U(g)f$ by h for various choices of the element g, Eqs. (1.14) and (1.15)

[5] The necessary properties of the hypergeometric functions are contained in *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., Inc., New York, 1953), Vol. I. Also in N. N. Lebedev, *Special Functions and Their Applications* (Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965).

[6] We have applied the technique given in Ref. 2(a) to the construction of these UIR's as found in V. Bargmann, Ann. Math. **48**, 568 (1947).

are replaced by the following set:

$g = e^{i\zeta J_2}$:

$$h_1(q) = e^{\frac{1}{2}i[\xi(q+\zeta)-\xi(q)]} f_1(q+\zeta),$$
$$h_2(q) = e^{\frac{1}{2}i[\xi(q)-\xi(q+\zeta)]} f_2(q+\zeta); \qquad (2.2)$$

$g = e^{i\nu J_1}$, $\nu \geq 0$.

For all q we have the following:

$$h_1(q) = e^{\frac{1}{2}i[\xi(q')-\xi(q)]}(\cosh\nu + \cosh q \sinh\nu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (e^q + \tanh\tfrac{1}{2}\nu)/(1 + e^q \tanh\tfrac{1}{2}\nu); \qquad (2.3a)$$

for $q \geq \ln\coth\frac{1}{2}\nu$:

$$h_2(q) = e^{\frac{1}{2}i[\xi(q')+\xi(q)]}(\cosh q \sinh\nu - \cosh\nu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (e^q - \tanh\tfrac{1}{2}\nu)/(e^q \tanh\tfrac{1}{2}\nu - 1); \qquad (2.3b)$$

for $\ln\coth\frac{1}{2}\nu \geq q \geq \ln\tanh\frac{1}{2}\nu$:

$$h_2(q) = e^{\frac{1}{2}i[\xi(q)-\xi(q')]}(\cosh\nu - \cosh q \sinh\nu)^{-\frac{1}{2}-is} f_2(q'),$$
$$e^{q'} = (e^q - \tanh\tfrac{1}{2}\nu)/(1 - e^q \tanh\tfrac{1}{2}\nu); \qquad (2.3c)$$

for $\ln\tanh\frac{1}{2}\nu \geq q$:

$$h_2(q) = -e^{\frac{1}{2}i[\xi(q')+\xi(q)]}(\cosh q \sinh\nu - \cosh\nu)^{-\frac{1}{2}-is} f_1(q'),$$
$$e^{q'} = (\tanh\tfrac{1}{2}\nu - e^q)/(1 - e^q \tanh\tfrac{1}{2}\nu). \qquad (2.3d)$$

For the case $g = e^{i\mu J_0}$, it is not necessary to write out the expressions for $h_r(q)$ in detail, since the only change from the previous situation is that Eqs. (1.16a)–(1.16d) all acquire an extra factor $e^{i\mu/2}$ on the right-hand side.

With the help of (2.2), we choose the eigenvectors of J_2 in the following way:

$$|s, \tfrac{1}{2}; p, \pm\rangle \to (4\pi)^{-\frac{1}{2}} \begin{pmatrix} e^{ipq-\frac{1}{2}i\xi(q)} \\ \pm e^{ipq+\frac{1}{2}i\xi(q)} \end{pmatrix}. \qquad (2.4)$$

Using this explicit construction, we can once again write down the $\mathcal{F}$ and $\mathcal{G}$ functions in the form of integrals. These turn out to be very similar to what we had previously, the same integrals appearing in different linear combinations. Omitting details, we quote the final expressions:

$\nu \geq 0$:

$$\mathcal{F}_{ba}^{(s,\frac{1}{2})}(p', p; \nu)$$
$$= (2\pi)^{-2}\Gamma(\tfrac{1}{2} + ip' + is)\Gamma(\tfrac{1}{2} - ip - is)\Gamma(ip - ip')$$
$$\times [\cosh\pi(p + s) + ba\cosh\pi(p' + s)$$
$$+ ib\sinh\pi(p - p')]\phi_1(s; p', p; \nu)$$
$$+ (2\pi)^{-2}\Gamma(\tfrac{1}{2} - ip' + is)\Gamma(\tfrac{1}{2} + ip - is)\Gamma(ip' - ip)$$
$$\times [\cosh\pi(p - s) + ba\cosh\pi(p' - s)$$
$$- ib\sinh\pi(p' - p)]\phi_2(s; p', p; \nu), \quad b, a = +, -;$$

$\nu \leq 0$:

$$\mathcal{F}_{ba}^{(s,\frac{1}{2})}(p', p; \nu) = [\mathcal{F}_{ab}^{(s,\frac{1}{2})}(p, p'; -\nu)]^*. \qquad (2.5)$$

Comparing this with the expression for $\mathcal{F}_{ba}^{(s,0)}(p', p; \nu)$ given in (1.19), we see that the only difference is that the third term in the square bracket multiplying ϕ_2 has changed sign. For $\mathcal{G}$, we have:

$0 \leq \mu \leq \pi$:

$$\mathcal{G}_{ba}^{(s,\frac{1}{2})}(p', p; \mu)$$
$$= (2\pi)^{-2}\Gamma(\tfrac{1}{2} + ip' + is)\Gamma(\tfrac{1}{2} - ip - is)\Gamma(ip - ip')$$
$$\times [\cosh\pi(p' + s) + ba\cosh\pi(p + s)$$
$$+ ia\sinh\pi(p - p')]\psi_1(s; p', p; \mu)$$
$$+ (2\pi)^{-2}\Gamma(\tfrac{1}{2} - ip' + is)\Gamma(\tfrac{1}{2} + ip - is)\Gamma(ip' - ip)$$
$$\times [\cosh\pi(p - s) + ba\cosh\pi(p' - s)$$
$$- ib\sinh\pi(p' - p)]\psi_2(s; p', p; \mu). \qquad (2.6)$$

Here again, on comparison with the form of $\mathcal{G}^{(s,0)}$ given in (1.22), we see that the third term in the square bracket multiplying ψ_2 has changed sign. To extend (2.6) to other values of μ, we first find the limit as $\mu \to \pi$ of the expression given in (2.6). In contrast to (1.23), here we get

$$\mathcal{G}_{++}^{(s,\frac{1}{2})}(p', p; \pi) = \mathcal{G}_{--}^{(s,\frac{1}{2})}(p', p; \pi) = 0,$$
$$\mathcal{G}_{-+}^{(s,\frac{1}{2})}(p', p; \pi) = -\mathcal{G}_{+-}^{(s,\frac{1}{2})}(p', p; \pi) = \delta(p' + p). \qquad (2.7)$$

Thus, as is to be expected, in the $C_q^{\frac{1}{2}}$ representations the element $e^{2\pi i J_0}$ is represented by the negative of the identity operator. So we can extend (2.6) in the following way:

$\pi \leq \mu \leq 2\pi$:

$$\mathcal{G}_{ba}^{(s,\frac{1}{2})}(p', p; \mu) = -[\mathcal{G}_{ab}^{(s,\frac{1}{2})}(p, p'; 2\pi - \mu)]^*, \qquad (2.8)$$

the extension to the range $2\pi \leq \mu \leq 4\pi$ being obvious.

3. EXCEPTIONAL UIR'S

In this final section we consider the case of the continuous exceptional UIR's. These are labeled by the parameter σ lying in the open interval $0 < \sigma < \frac{1}{2}$. The corresponding representation functions are defined in (1.8) and obey

$$\left[\frac{d^2}{d\nu^2} + \coth\nu\frac{d}{d\nu} + \frac{1}{4} - \sigma^2 + \frac{p'^2 + p^2 - 2p'p\cosh\nu}{\sinh^2\nu}\right] \times \mathcal{F}_{ba}^{(\sigma)}(p', p; \nu) = 0, \qquad (3.1a)$$

$$\left[\frac{d^2}{d\mu^2} + \cot\mu\frac{d}{d\mu} - \frac{1}{4} + \sigma^2 + \frac{p'^2 + p^2 - 2p'p\cos\mu}{\sin^2\mu}\right] \times \mathcal{G}_{ba}^{(\sigma)}(p', p; \mu) = 0. \qquad (3.1b)$$

These equations are the same as (1.9) with s replaced by $-i\sigma$. We therefore choose the basic solutions by the same replacement in the functions defined in

(1.10) and (1.11). We define

$$\begin{aligned}
&\phi_1(\sigma; p', p; \nu)\\
&\quad = (\cosh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p'+p)}(\sinh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p'-p)}\\
&\qquad \times F(\tfrac{1}{2} + \sigma + ip', \tfrac{1}{2} - \sigma + ip';\\
&\qquad\qquad 1 + ip' - ip; -\sinh^2 \tfrac{1}{2}\nu),\\
&\phi_2(\sigma; p', p; \nu) \qquad (3.2)\\
&\quad = (\cosh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p+p')}(\sinh^2 \tfrac{1}{2}\nu)^{\frac{1}{2}i(p-p')}\\
&\qquad \times F(\tfrac{1}{2} + \sigma + ip, \tfrac{1}{2} - \sigma + ip;\\
&\qquad\qquad 1 + ip - ip'; -\sinh^2 \tfrac{1}{2}\nu),\\
&\qquad \phi_2(\sigma; p', p; \nu) = \phi_1(\sigma; p, p'; \nu),
\end{aligned}$$

and

$$\begin{aligned}
&\psi_1(\sigma; p', p; \mu)\\
&\quad = (\cos^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p'+p)}(\sin^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p'-p)}\\
&\qquad \times F(\tfrac{1}{2} + \sigma + ip', \tfrac{1}{2} - \sigma + ip';\\
&\qquad\qquad 1 + ip' - ip; \sin^2 \tfrac{1}{2}\mu),\\
&\psi_2(\sigma; p', p; \mu) \qquad (3.3)\\
&\quad = (\cos^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p+p')}(\sin^2 \tfrac{1}{2}\mu)^{\frac{1}{2}i(p-p')}\\
&\qquad \times F(\tfrac{1}{2} + \sigma + ip, \tfrac{1}{2} - \sigma + ip;\\
&\qquad\qquad 1 + ip - ip'; \sin^2 \tfrac{1}{2}\mu),\\
&\qquad \psi_2(\sigma; p', p; \mu) = \psi_1(\sigma; p, p'; \mu).
\end{aligned}$$

In Ref. 2(b) we have shown how one can write the exceptional class UIR's of $O(2, 1)$ in a basis suited to diagonalization of J_2. We have a Hilbert space $\mathcal{K}_\sigma$ whose elements consist of pairs of functions of the real variable q running from $-\infty$ to $+\infty$:

$$f \to \begin{pmatrix} f_1(q) \\ f_2(q) \end{pmatrix}. \qquad (3.4)$$

The inner product of two vectors f and h depends on σ and is defined as follows: Let $f_\pm(q)$ be defined by

$$f_\pm(q) = \tfrac{1}{2}[f_1(q) \pm f_2(q)], \qquad (3.5)$$

and similarly for h. Then,

$$\begin{aligned}
(h, f) &= \sum_{a=+,-} \int_{-\infty}^{\infty} dq' \int_{-\infty}^{\infty} dq\\
&\qquad \times h_a(q')^* K_a(\sigma; q' - q) f_a(q),\\
K_a(\sigma; q) &= (2\pi)^{-1} \int_{-\infty}^{\infty} dp e^{ipq} \Lambda_a(\sigma; p), \qquad (3.6)\\
\Lambda_\pm(\sigma; p) &= \Gamma(\tfrac{1}{2} - \sigma + ip)\Gamma(\tfrac{1}{2} - \sigma - ip)\\
&\qquad \times [\cosh \pi p \pm \sin \pi\sigma].
\end{aligned}$$

Notice that in the q space the scalar product is nonlocal and translation invariant, while in p space it becomes local. Next we must give the effect of the unitary operators $U(g)$ on an arbitrary vector f. If we write as before $h = U(g)f$, then the components $h_r(q)$, $r = 1, 2$, are given in terms of $f_r(q)$ by the same equations as in the case of the integral nonexceptional UIR's C_q^0 with the replacement everywhere of s by $-i\sigma$, namely by (1.14), (1.15), and (1.16) with $s \to -i\sigma$. Thus the effect of $U(g)$ has a uniform appearance if we specify a vector f by its components $f_1(q)$ and $f_2(q)$, whether we are dealing with the nonexceptional or the exceptional set of UIR's.[7] However, in order to "diagonalize" the expression for the inner product, we have to pass to the functions $f_\pm(q)$ in the exceptional case. The orthonormal eigenvectors of J_2 have now to be chosen as follows:

$$|\sigma; p, \pm\rangle \to [2\pi\Lambda_\pm(\sigma; p)]^{-\frac{1}{2}} \begin{pmatrix} e^{ipq} \\ \pm e^{ipq} \end{pmatrix}. \qquad (3.7)$$

This choice is determined by the structure of the scalar product.

Using this explicit construction, we could express $\mathcal{F}^{(\sigma)}$ and $\mathcal{G}^{(\sigma)}$ in the form of integrals. As an example we find, for $\nu \geq 0$,

$$\begin{aligned}
\mathcal{F}^{(\sigma)}_{++}(p', p; \nu) &= \frac{1}{\pi}\left[\frac{\Lambda_+(\sigma; p')}{\Lambda_+(\sigma; p)}\right]^{\frac{1}{2}}\\
&\qquad \times \left\{ \int_{-\infty}^{\infty} dq e^{-ip'q} (\cosh \nu + \cosh q \sinh \nu)^{-\frac{1}{2}-\sigma} \right.\\
&\qquad \times \left(\frac{e^q + \tanh \frac{1}{2}\nu}{1 + e^q \tanh \frac{1}{2}\nu} \right)^{ip}\\
&\qquad + \int_{-\infty}^{\infty} dq e^{-ip'q} |\cosh \nu - \cosh q \sinh \nu|^{-\frac{1}{2}-\sigma}\\
&\qquad \left. \times \left| \frac{e^q - \tanh \frac{1}{2}\nu}{e^q \tanh \frac{1}{2}\nu - 1} \right|^{ip} \right\}. \qquad (3.8)
\end{aligned}$$

On examination [see (1.18)] these integrals are the same as those encountered in the continuous nonexceptional classes, with $s \to -i\sigma$, and *for σ in the open region* $0 < \sigma < \frac{1}{2}$, it turns out that their values too are obtained from the previous expressions by just replacing s by $-i\sigma$. This is true for all the integrals we need to evaluate, both for $\mathcal{F}$ and for $\mathcal{G}$. After some algebra, we finally get:

$\nu \geq 0$:

$$\begin{aligned}
&\mathcal{F}^{(\sigma)}_{ba}(p', p; \nu)\\
&\quad = (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}}\\
&\qquad \times \Gamma(\tfrac{1}{2} + \sigma + ip')\Gamma(\tfrac{1}{2} - \sigma - ip)\Gamma(ip - ip')\\
&\qquad \times [\cosh \pi(p - i\sigma) + ba \cosh \pi(p' - i\sigma)\\
&\qquad + ib \sinh \pi(p - p')]\phi_1(\sigma; p', p; \nu)\\
&\qquad + (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}}\\
&\qquad \times \Gamma(\tfrac{1}{2} + \sigma - ip')\Gamma(\tfrac{1}{2} - \sigma + ip)\Gamma(ip' - ip)\\
&\qquad \times [\cosh \pi(p + i\sigma) + ba \cosh \pi(p' + i\sigma)\\
&\qquad + ib \sinh \pi(p' - p)]\phi_2(\sigma; p', p; \nu);
\end{aligned}$$

[7] We could, of course, have worked with the combinations $f_\pm(q)$, instead of $f_1(q)$ and $f_2(q)$, even in the continuous nonexceptional UIR's without spoiling the structure of the scalar product.

$\nu \leq 0$:

$$\mathcal{F}^{(\sigma)}_{ba}(p', p; \nu) = [\mathcal{F}^{(\sigma)}_{ab}(p, p'; -\nu)]^*, \quad b, a = +, -; \tag{3.9}$$

$0 \leq \mu \leq \pi$:

$$\begin{aligned}\mathcal{G}^{(\sigma)}_{ba}(p', p; \mu) &= (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}} \\ &\times \Gamma(\tfrac{1}{2} + \sigma + ip')\Gamma(\tfrac{1}{2} - \sigma - ip)\Gamma(ip - ip') \\ &\times [\cosh \pi(p' - i\sigma) + ba \cosh \pi(p - i\sigma) \\ &+ ia \sinh \pi(p - p')]\psi_1(\sigma; p', p; \mu) \\ &+ (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}} \\ &\times \Gamma(\tfrac{1}{2} + \sigma - ip')\Gamma(\tfrac{1}{2} - \sigma + ip)\Gamma(ip' - ip) \\ &\times [\cosh \pi(p + i\sigma) + ba \cosh \pi(p' + i\sigma) \\ &+ ib \sinh \pi(p' - p)]\psi_2(\sigma; p', p; \mu);\end{aligned}$$

$\pi \leq \mu \leq 2\pi$:

$$\mathcal{G}^{(\sigma)}_{ba}(p', p; \mu) = [\mathcal{G}^{(\sigma)}_{ab}(p, p'; 2\pi - \mu)]^*, \quad b, a = +, -. \tag{3.10}$$

This completes the evaluation of the representation matrices for the exceptional continuous class of UIR's of $SU(1, 1)$. This class is defined for values of the parameter σ in the *open* interval $0 < \sigma < \frac{1}{2}$. However, if we extrapolate the expressions given in (3.9) and (3.10) for $\mathcal{F}^{(\sigma)}$ and $\mathcal{G}^{(\sigma)}$ to the point $\sigma = 0$, we see that they coincide exactly with the representation matrices $\mathcal{F}^{(s,0)}$ and $\mathcal{G}^{(s,0)}$ of the continuous integral nonexceptional series evaluated at $s = 0$. [See (1.19) and (1.22).] Further, for $\sigma = 0$, the two weight functions $\Lambda_\pm(\sigma; p)$ become equal and independent of p, so that the kernels $K_a(\sigma; q)$ appearing in (3.6) reduce to delta functions. All these properties are in agreement with the statement that the continuous nonexceptional integral UIR's and the continuous exceptional UIR's "meet" at the point $s = \sigma = 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Further Note on Two Binomial Coefficient Identities of Rosenbaum

DAVID CHIANG
Université Laval, Québec, P.Q., Canada

(Received 25 April 1969)

This note gives additional condition to the finding of Gould on two binomial coefficient identities of Rosenbaum.

Using the Γ function to examine the binomial coefficient,[1] when x is an integer and n is a nonnegative integer, it is found that

$$\begin{aligned}\binom{x}{n} &= 0, && 0 \leq x < n, \\ \binom{x}{n} &= 1, && x = n, \\ \binom{x}{n} &= 1, && n = 0, \quad x \geq 0, \\ \binom{x}{n} &= \text{indeterminate}, && x < 0, \quad x \neq n.\end{aligned}$$

Therefore, Eqs. (4) and (5) of Gould[2] will not hold true if x or y or $(x + y)$ is a negative integer, in Eq. (4),

$$\sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n}$$

and if x is an integer, in Eq. (5),

$$\binom{-x}{n} = (-1)^n\binom{x+n-1}{n}.$$

Considering the expression

$$A = \sum_{n=0}^{\alpha}\binom{-\epsilon}{n}\binom{\epsilon}{\alpha-n} = \binom{0}{\alpha},$$

it is evident that it will be null when ϵ is a noninteger, while in addition α and n are integers and at the same time $\alpha > 0$, $n \geq 0$. This differs from the conditions mentioned by Gould.[2]

[1] D. M. Rosenbaum, J. Math. Phys. **8**, 1973 (1967).
[2] H. W. Gould, J. Math. Phys. **10**, 49 (1969).

$\nu \leq 0$:

$$\mathcal{F}^{(\sigma)}_{ba}(p', p; \nu) = [\mathcal{F}^{(\sigma)}_{ab}(p, p'; -\nu)]^*, \quad b, a = +, -; \tag{3.9}$$

$0 \leq \mu \leq \pi$:

$$\begin{aligned}\mathcal{G}^{(\sigma)}_{ba}(p', p; \mu) &= (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}} \\ &\times \Gamma(\tfrac{1}{2} + \sigma + ip')\Gamma(\tfrac{1}{2} - \sigma - ip)\Gamma(ip - ip') \\ &\times [\cosh \pi(p' - i\sigma) + ba \cosh \pi(p - i\sigma) \\ &+ ia \sinh \pi(p - p')]\psi_1(\sigma; p', p; \mu) \\ &+ (2\pi)^{-2}[\Lambda_b(\sigma; p')/\Lambda_a(\sigma; p)]^{\frac{1}{2}} \\ &\times \Gamma(\tfrac{1}{2} + \sigma - ip')\Gamma(\tfrac{1}{2} - \sigma + ip)\Gamma(ip' - ip) \\ &\times [\cosh \pi(p + i\sigma) + ba \cosh \pi(p' + i\sigma) \\ &+ ib \sinh \pi(p' - p)]\psi_2(\sigma; p', p; \mu);\end{aligned}$$

$\pi \leq \mu \leq 2\pi$:

$$\mathcal{G}^{(\sigma)}_{ba}(p', p; \mu) = [\mathcal{G}^{(\sigma)}_{ab}(p, p'; 2\pi - \mu)]^*, \quad b, a = +, -. \tag{3.10}$$

This completes the evaluation of the representation matrices for the exceptional continuous class of UIR's of $SU(1, 1)$. This class is defined for values of the parameter σ in the *open* interval $0 < \sigma < \frac{1}{2}$. However, if we extrapolate the expressions given in (3.9) and (3.10) for $\mathcal{F}^{(\sigma)}$ and $\mathcal{G}^{(\sigma)}$ to the point $\sigma = 0$, we see that they coincide exactly with the representation matrices $\mathcal{F}^{(s,0)}$ and $\mathcal{G}^{(s,0)}$ of the continuous integral nonexceptional series evaluated at $s = 0$. [See (1.19) and (1.22).] Further, for $\sigma = 0$, the two weight functions $\Lambda_\pm(\sigma; p)$ become equal and independent of p, so that the kernels $K_a(\sigma; q)$ appearing in (3.6) reduce to delta functions. All these properties are in agreement with the statement that the continuous nonexceptional integral UIR's and the continuous exceptional UIR's "meet" at the point $s = \sigma = 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Further Note on Two Binomial Coefficient Identities of Rosenbaum

DAVID CHIANG
Université Laval, Québec, P.Q., Canada

(Received 25 April 1969)

This note gives additional condition to the finding of Gould on two binomial coefficient identities of Rosenbaum.

Using the Γ function to examine the binomial coefficient,[1] when x is an integer and n is a nonnegative integer, it is found that

$$\begin{aligned}\binom{x}{n} &= 0, && 0 \leq x < n, \\ \binom{x}{n} &= 1, && x = n, \\ \binom{x}{n} &= 1, && n = 0, \quad x \geq 0, \\ \binom{x}{n} &= \text{indeterminate}, && x < 0, \quad x \neq n.\end{aligned}$$

Therefore, Eqs. (4) and (5) of Gould[2] will not hold true if x or y or $(x + y)$ is a negative integer, in Eq. (4),

$$\sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n}$$

and if x is an integer, in Eq. (5),

$$\binom{-x}{n} = (-1)^n\binom{x+n-1}{n}.$$

Considering the expression

$$A = \sum_{n=0}^{\alpha}\binom{-\epsilon}{n}\binom{\epsilon}{\alpha-n} = \binom{0}{\alpha},$$

it is evident that it will be null when ϵ is a noninteger, while in addition α and n are integers and at the same time $\alpha > 0$, $n \geq 0$. This differs from the conditions mentioned by Gould.[2]

[1] D. M. Rosenbaum, J. Math. Phys. **8**, 1973 (1967).
[2] H. W. Gould, J. Math. Phys. **10**, 49 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

On a Generalized Distribution of the Poles of the Unitary Collision Matrix

NAZAKAT ULLAH*
Tata Institute of Fundamental Research, Bombay, India

(Received 23 October 1967; Revised Manuscript Received 10 February 1969)

An expression for the joint distribution of the complex poles of the unitary collision matrix is derived for the single-channel case, which is valid for all values of the ratio of the width to the spacing. The derivation uses the statistical distribution of the parameters of the real R-matrix theory. We find that unitarity gives rise to the statistical correlations between the width and the spacing of the collision matrix. It is shown that the distribution of the poles of the unitary collision matrix using Feshbach's unified theory of nuclear reactions is the same as the one obtained using R-matrix theory, provided we make a particular choice of the arbitrary boundary condition in the latter theory. A remark is made about the use of the random complex orthogonal matrix in the study of the parameters of the statistical collision matrix.

I. INTRODUCTION

Recently it was shown[1] that, for purely elastic scattering, the unitary pole resonance form of the low-energy collision matrix U can always be written as

$$U = [\exp(-2i\phi)]\left(1 - i\sum_{\mu=1}^{N}\frac{G_\mu}{E - Z_\mu}\right), \qquad (1)$$

where N is the number of compound nucleus resonances which may be interfering with each other. The first term containing ϕ gives rise to potential scattering and $Z_\mu = \epsilon_\mu - \frac{1}{2}i\Gamma_\mu$ are the complex poles of U. The amplitudes G_μ are complex and are given by the expression

$$G_\mu = (2\,\mathrm{Im}\,Z_\mu^*)\prod_{\nu\neq\mu}^{N}(Z_\mu - Z_\nu^*)(Z_\mu - Z_\nu)^{-1}. \qquad (2)$$

In the study of the cross-section fluctuations[2] and intermediate structure,[3] a knowledge of the statistical properties of the parameters of U is needed. For the case of isolated resonances, when the average width is much smaller than the average spacing, the parameters of U become the same as the parameters of the real R-matrix theory,[4] which have been very well studied in the past.[5] The purpose of the present paper is to derive a joint-distribution function of the complex poles Z_μ, starting from the known distribution of the parameters of the real R-matrix theory. Since Eq. (2) gives the complex amplitudes G_μ in terms of Z_μ, all the statistical properties of G_μ for any value of the ratio of the average width to the average spacing, therefore, will also be known once the joint distribution of Z_μ is given. We give this distribution in Sec. II.

In Sec. III we shall discuss the distribution of the parameters of U using Feshbach's unified theory of nuclear reactions.[6,7] A few remarks will be made about the complex boundary-value problem[8] and the use of the random complex orthogonal matrix[9] in the statistical study of the complex amplitudes of the statistical collision matrix defined by Moldauer.[8]

II. DISTRIBUTION OF THE POLES OF THE UNITARY COLLISION MATRIX

The resonances of the real R-matrix theory[4] are obtained by solving the eigenvalue equation

$$H\Phi_\mu = E_\mu\Phi_\mu, \qquad (3)$$

where H is the compound nucleus Hamiltonian, and Φ_μ and E_μ are its eigenfunctions and eigenvalues, respectively. The eigenvalue equation (3) is solved in the internal region by specifying a certain real boundary condition at the surface which divides the whole of configuration space into an external and an internal region. Apart from a constant, the overlap integral of the wavefunction Φ_μ with the smooth-channel wavefunction defines the reduced width amplitude $\gamma_{\mu c}$, where c denotes a particular channel. In this paper we shall restrict ourselves to the problem of a single channel and so the subscript c will be omitted.

Let us assume that the Hamiltonian H is invariant under rotations and under time reversal, so that the joint distribution of the amplitudes γ_μ is given by[10]

$$P(\{\gamma_\mu\}) = f\left(\sum_{\mu=1}^{N}\gamma_\mu^2\right). \qquad (4)$$

* Present address: Department of Physics, University of Toronto, Toronto, Ontario, Canada.

[1] N. Ullah and C. S. Warke, Phys. Rev. **170**, 857 (1968); see also C. Mahaux and H. A. Weidenmüller, N. P. **A91**, 241 (1967).
[2] T. Ericson, Ann. Phys. (N.Y.) **23**, 390 (1963).
[3] H. Feshbach, A. K. Kerman, and R. H. Lemmer, Ann. Phys. (N.Y.) **41**, 230 (1967).
[4] A. M. Lane and R. G. Thomas, Rev. Mod. Phys. **30**, 257 (1958).
[5] C. E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic Press Inc., New York, 1965).
[6] H. Feshbach, Ann. Phys. (N.Y.) **5**, 357 (1958); **19**, 287 (1962).
[7] H. Feshbach, Ann. Phys. (N.Y.) **43**, 410 (1967).
[8] P. A. Moldauer, Phys. Rev. **135**, B642 (1964).
[9] Nazakat Ullah, Phys. Rev. **154**, 893 (1967).
[10] Nazakat Ullah, J. Math. Phys. **8**, 1095 (1967).

The joint-distribution function of the eigenvalues[5] E_μ, apart from the factor $\prod_{\mu<\nu}^{N}|E_\mu - E_\nu|$, will be a function of the quantities of the type $\sum_{\mu=1}^{N} E_\mu^n$, $1 \leq n < N$. For the derivation in this section we restrict ourselves to the distribution of H,[5,11] which is a function of Tr H^2. The joint distribution of the eigenvalues E_μ can now be written as

$$P(\{E_\mu\}) = g\left(\sum_{\mu=1}^{N} E_\mu^2\right) \prod_{\mu<\nu}^{N} |E_\mu - E_\nu|. \tag{5}$$

We shall see later that the general case in which the function g contains other $\sum_{\mu=1}^{N} E_\mu^n$ can also be worked out without much difficulty.

From expressions (4) and (5) we see that the correlation between γ_μ and E_μ is strictly zero. Therefore any correlation between the parameters ϵ_μ and Γ_μ of U will be due to unitarity only.

According to R-matrix theory[4] the function R is given by

$$R = R^0 + \sum_{\mu=1}^{N} \frac{\gamma_\mu^2}{E - E_\mu}, \tag{6}$$

where R^0 gives rise to background scattering. Using the connection between the collision function U and R, we find that the relation between the complex poles Z_μ and the quantities E_μ, γ_μ can be expressed as

$$\begin{aligned}
&\sum_{\mu=1}^{N} Z_\mu = \sum_{\mu=1}^{N} E_\mu - \eta \sum_{\mu=1}^{N} \gamma_\mu^2, \\
&\sum_{\mu<\nu}^{N} Z_\mu Z_\nu = \sum_{\mu<\nu}^{N} E_\mu E_\nu - \eta \sum_{\mu=1}^{N} \gamma_\mu^2 \left(\sum_{\alpha\neq\mu}^{N} E_\alpha\right), \\
&\quad\cdot \qquad\quad \cdot \qquad\quad \cdot \\
&\quad\cdot \qquad\quad \cdot \qquad\quad \cdot \\
&\quad\cdot \qquad\quad \cdot \qquad\quad \cdot \\
&\prod_{\mu=1}^{N} Z_\mu = \prod_{\mu=1}^{E} E_\mu - \eta \sum_{\mu=1}^{N} \gamma_\mu^2 \left(\prod_{\alpha\neq\mu}^{N} E_\alpha\right),
\end{aligned} \tag{7}$$

where $\eta = L^0(1 - R^0L^0)^{-1}$. The quantity $L^0 = L - B$, where the real and imaginary parts of L define the shift and the penetration factors, respectively, and B is a real number which specifies the boundary condition.

The joint distribution of the real and imaginary parts of Z_μ, which are denoted by Z_μ^r, Z_μ^i, will be obtained using expressions (4), (5), (7), and the usual method of the transformation[12] of the volume element from the space of the variables γ_μ, E_μ to the new variables Z_μ^r, Z_μ^i. For convenience we introduce the quantities x_μ, which are defined by

$$x_\mu = (2 \operatorname{Im} \eta)^{\frac{1}{2}} \gamma_\mu.$$

Instead of expression (4), we now write the distribution of the quantities x_μ as

$$P(\{x_\mu\}) = f(\textstyle\sum x_\mu^2). \tag{8a}$$

The first two relations in (7) enable us to write

$$f\left(\sum_{\mu=1}^{N} x_\mu^2\right) = f\left(-2\sum_{\mu=1}^{N} Z_\mu^i\right), \tag{8b}$$

$$g\left(\sum_{\mu=1}^{N} E_\mu^2\right) = g\left(\sum_{\mu=1}^{N} (Z_\mu^r - \omega Z_\mu^i)^2 + 2(1+\omega^2)\sum_{\mu<\nu}^{N} Z_\mu^i Z_\nu^i\right), \tag{9}$$

where $\omega = (\operatorname{Re}\eta)/(\operatorname{Im}\eta)$. In the general case when the function g contains other $\sum_{\mu=1}^{N} E_\mu^n$, we can again express them in terms of the quantities Z_μ using the relations (7).

Next, let us consider the transformation of the volume element. We first introduce a new set of variables defined by the relations

$$\begin{aligned}
&u_1 = \sum_{\mu=1}^{N} E_\mu, \quad u_2 = \sum_{\mu<\nu}^{N} E_\mu E_\nu, \cdots, \quad u_N = \prod_{\mu=1}^{N} E_\mu, \\
&v_1 = \textstyle\sum x_\mu^2, \quad v_2 = \sum_{\mu=1}^{N} x_\mu^2 \left(\sum_{\alpha\neq\mu} E_\alpha\right), \cdots, \\
&\qquad\qquad v_N = \sum_{\mu=1}^{N} x_\mu^2 \left(\prod_{\alpha\neq\mu} E_\alpha\right).
\end{aligned} \tag{10}$$

The volume element in the space of the variables u_μ, v_μ is related to the volume element in the space of the variables E_μ, x_μ in the following way:

$$\prod_{\mu=1}^{N} du_\mu\, dv_\mu = \begin{vmatrix} d_1 & 0 \\ d_3 & d_2 \end{vmatrix} \prod_{\mu=1}^{N} dE_\mu\, dx_\mu, \tag{11}$$

where d_1, d_2, d_3 are $N \times N$ determinants and 0 denotes an $N \times N$ determinant which has all elements 0. Because of the particular form of the $2N \times 2N$ determinant in expression (11), it can be written as a product of the determinants d_1, d_2. Therefore,

$$\prod_{\mu=1}^{N} du_\mu\, dv_\mu = d_1\, d_2 \prod_{\mu=1}^{N} dE_\mu\, dx_\mu. \tag{12}$$

The $N \times N$ determinant d_1 can be written as

$$d_1 = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \sum_{\mu\neq1} E_\mu & \sum_{\mu\neq2} E_\mu & \cdots & \sum_{\mu\neq N} E_\mu \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \prod_{\mu\neq1} E_\mu & \prod_{\mu\neq2} E_\mu & \cdots & \prod_{\mu\neq N} E_\mu \end{vmatrix}$$

[11] N. Rosenzweig, *Brandeis University Summer Institute Lectures in Theoretical Physics, 1962 Lectures, Vol. 3*, K. W. Ford, Ed. (W. A. Benjamin, Inc., New York, 1963), p. 91.

[12] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, Inc., New York, 1956), p. 11.

It can be easily shown that the determinant d_1 reduces to the following simple form:

$$d_1 = \prod_{\mu<\nu}^{N} (E_\mu - E_\nu). \tag{13}$$

The determinant d_2 is the same as determinant d_1 except that it is multiplied by an extra factor $2^N(\prod_{\mu=1}^{N} x_\mu)$. The constant factors, like the factor 2^N, will be absorbed in the normalization constant and, since the absolute value of the Jacobian enters in the transformation of the volume element, we shall write the determinants as their absolute values. Expressions (12) and (13) then give us

$$\prod_{\mu=1}^{N} du_\mu\, dv_\mu = \left(\prod_{\mu=1}^{N} |x_\mu|\right)\left(\prod_{\mu<\nu} |E_\mu - E_\nu|^2\right). \tag{14}$$

The relations (10) which define the quantities v_μ can be looked upon as N simultaneous linear equations for the N unknowns x_μ^2. Using the well-known methods for the solution of the linear equations, after a few simplifying steps we obtain the following relation, using expression (14):

$$\prod_{\mu<\nu} |E_\mu - E_\nu| \prod_{\mu=1}^{N} dE_\mu\, dx_\mu = \left[\prod_{\mu=1}^{N} D_\mu^{-\frac{1}{2}}\right] \prod_{\mu=1}^{N} du_\mu\, dv_\mu, \tag{15}$$

where D_μ is the $N \times N$ determinant

$$D_\mu = \begin{vmatrix} v_1 & 0 & 0 & \cdots & 0 & 1 \\ v_2 & 0 & 0 & \cdots & 1 & E_\mu \\ v_3 & 0 & 0 & \cdots & E_\mu & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ v_{N-1} & 1 & E_\mu & \cdots & 0 & 0 \\ v_N & E_\mu & 0 & \cdots & 0 & 0 \end{vmatrix}. \tag{16}$$

Before we make the final transformation to the set of variables Z_μ^r, Z_μ^i, we replace the set of variables u_μ by the new set of variables y_μ, which are defined by the relations

$$y_\mu = u_\mu - \tfrac{1}{2}\omega v_\mu. \tag{17a}$$

As can easily be checked, the Jacobian of the transformation from the set of variables (u, v) to the new set (y, v) is unity.

The variables y_μ, v_μ are nothing but the real and imaginary parts of the complex quantities on the left-hand side of the relations (7). The transformation of the volume element from the set of the variables y_μ, v_μ to the set of variables Z_μ^r, Z_μ^i is given by

$$\prod_{\mu=1}^{N} dy_\mu\, dv_\mu = \begin{vmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \sum_{\mu\neq 1} Z_\mu^r & \sum_{\mu\neq 2} Z_\mu^r & \cdots & \sum_{\mu\neq N} Z_\mu^r & -\sum_{\mu\neq 1} Z_\mu^i & \cdots & -\sum_{\mu\neq N} Z_\mu^i \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \mathrm{Re}\prod_{\mu\neq 1} Z_\mu & \mathrm{Re}\prod_{\mu\neq 2} Z_\mu & \cdots & \mathrm{Re}\prod_{\mu\neq N} Z_\mu & -\mathrm{Im}\prod_{\mu\neq 1} Z_\mu & \cdots & -\mathrm{Im}\prod_{\mu\neq N} Z_\mu^i \\ 0 & 0 & \cdots & 0 & -2 & \cdots & -2 \\ -2\sum_{\mu\neq 1} Z_\mu^i & -2\sum_{\mu\neq 2} Z_\mu^i & \cdots & -2\sum_{\mu\neq N} Z_\mu^i & -2\sum_{\mu\neq 1} Z_\mu^r & \cdots & -2\sum_{\mu\neq N} Z_\mu^r \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ -2\,\mathrm{Im}\prod_{\mu\neq 1} Z_\mu & -2\,\mathrm{Im}\prod_{\mu\neq 2} Z_\mu & \cdots & -2\,\mathrm{Im}\prod_{\mu\neq N} Z_\mu & -2\,\mathrm{Re}\prod_{\mu\neq 1} Z_\mu & \cdots & -2\,\mathrm{Re}\prod_{\mu\neq N} Z_\mu \end{vmatrix} \prod_{\mu=1}^{N} dZ_\mu^r\, dZ_\mu^i. \tag{17b}$$

The Jacobian of the transformation is a $2N \times 2N$ determinant. By subtracting out the first column from the second, third, $\cdots$, Nth column and the $(N+1)$th column from the $(N+2)$th, $\cdots$, $2N$th column, we find that the first row contains zero in all columns except the first one. We next multiply the second column by $(Z_1^r - Z_2^r)$, $\cdots$, the Nth column by $(Z_1^r - Z_N^r)$, the $(N+1)$th column by $(Z_1^i - Z_2^i)$, $\cdots$, the $2N$th column by $(Z_1^i - Z_N^i)$ and subtract the $(N+1)$th column from the second column, $\cdots$, the

Nth column from the $2N$th column. This gives us a factor

$$\prod_{\mu=2}^{N}[(Z_1^r - Z_\mu^r)^2 + (Z_1^i - Z_\mu^i)^2]$$

and a $(2N-2)\times(2N-2)$ determinant, which is similar in form to the original $2N\times 2N$ determinant. Continuing this process, we find that

$$\prod_{\mu=1}^{N} dy_\mu\, dv_\mu = \left(\prod_{\mu<\nu}^{N}[(Z_\mu^r - Z_\nu^r)^2 + (Z_\mu^i - Z_\nu^i)^2]\right)\prod_{\mu=1}^{N} dZ_\mu^r\, dZ_\mu^i. \tag{18}$$

The last step in the derivation is to express the products of the determinants D_μ in terms of the quantities Z_μ. This product can be put in a form which is called the resultant of two polynomials.[13] We write the two polynomials $h(x)$ and $k(x)$ as

$$h(x) = v_1x^{N-1} - v_2x^{n-2} + \cdots + (-1)^{N-1}v_N,$$
$$k(x) = x^N - y_1x^{N-1} + y_2x^{N-2} + \cdots + (-1)^N y_N. \tag{19}$$

Then,

$$\prod_{\mu=1}^{N} D_\mu = R[h(x), k(x)], \tag{20}$$

where R is the resultant of $h(x)$, $k(x)$. The resultant R is a polynomial[13] in the coefficients of $h(x)$, $k(x)$. According to Sylvester's method,[14] R can be written as a $(2N-1)\times(2N-1)$ determinant:

$$R = \begin{vmatrix} 1 & -y_1 & \cdots & (-1)^N y_N & 0 & \cdots & 0 \\ 0 & 1 & \cdots & (-1)^{N-1}y_{N-1} & (-1)^N y_N & \cdots & 0 \\ 0 & 0 & \cdots & (-1)^{N-2}y_{N-2} & (-1)^{N-1}y_{N-1} & \cdots & 0 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & y_2 & y_3 & \cdots & (-1)^N y_N \\ v_1 & -v_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & v_1 & \cdots & (-1)^{N-1}v_N & 0 & \cdots & 0 \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & -v_2 & v_3 & \cdots & (-1)^{N-1}v_N \end{vmatrix}.$$

Manipulating the columns of the above determinant, we can show that

$$R = \left(\prod_{\mu=1}^{N} Z_\mu^i\right)\left(\prod_{\mu<\nu}^{N}[(Z_\mu^r - Z_\nu^r)^2 + (Z_\mu^i + Z_\nu^i)^2]\right). \tag{21}$$

Using expressions (5), (7), (8a), (8b), (9), (15), (18), and (21), we finally get an expression for the distribution of the complex poles. It is given by

$$\begin{aligned} P(\{Z_\mu^r, Z_\mu^i\})\prod_{\mu=1}^{N} dZ_\mu^r\, dZ_\mu^i &= K^{-1} f\left(-2\sum_{\mu=1}^{N} Z_\mu^i\right) \\ &\times g\left(\sum_{\mu=1}^{N}(Z_\mu^r - \omega Z_\mu^i)^2 + 2(1+\omega^2)\sum_{\mu<\nu}^{N} Z_\mu^i Z_\nu^i\right) \\ &\times \prod_{\mu<\nu}^{N}[(Z_\mu^r - Z_\nu^r)^2 + (Z_\mu^i - Z_\nu^i)^2] \\ &\times \left[\left(\prod_{\mu=1}^{N} Z_\mu^i\right)\left(\prod_{\mu<\nu}^{N}[(Z_\mu^r - Z_\nu^r)^2 + (Z_\mu^i + Z_\nu^i)^2]\right)\right]^{-\frac{1}{2}} \\ &\times \prod_{\mu=1}^{N} dZ_\mu^r\, dZ_\mu^i, \end{aligned} \tag{22}$$

where K is the normalization constant.

III. CONCLUDING REMARKS

Expression (22) gives the joint probability distribution of the parameters of the collision function U. We see from this expression that the parameters $Z_\mu^r = \epsilon_\mu$, $Z_\mu^i = -\frac{1}{2}\Gamma_\mu$ are statistically correlated. Since the statistical correlation of the parameters of the real R-matrix was zero, this correlation is a consequence of unitarity for all values of the ratio of average width to the average spacing.

We remark here that even though we have not used a specific form for the functions f and g, they can be taken to be exponential[5,10,11] for many kinds of the random Hamiltonian matrix ensembles of large dimension.

The derivation of the joint-distribution function of the complex poles of U, given in Sec. II, was based on the real R-matrix theory of nuclear reactions. We can now ask whether this distribution will change

[13] C. C. MacDuffee, *Theory of Equations* (John Wiley & Sons, Inc., New York, 1954), p. 111.

[14] W. S. Burnside and A. W. Panton, *The Theory of Equations* (Dover Publications, Inc., New York, 1960), Vol. 2, p. 75.

if we start from some other theory, e.g., Feshbach's unified theory.[6,7] It should be noted that the form of the unitary collision matrix given by (1) is independent of any formal reaction theory and is a consequence of unitarity. When one studies the statistical properties of the quantities Z_μ, then a particular reaction theory has to be used to connect Z_μ with the eigenvalue and eigenvector components of the compound nucleus Hamiltonian; and if these relations turn out to be different for different reaction theories, then the joint-distribution function of the quantities Z_μ will also be different.

Recently, Feshbach[7] has expressed his scattering matrix T in the form

$$T_{fi} = T_{fi}^{\text{pot}} + \frac{1}{2\pi} e^{2i\delta} \sum \frac{\Gamma_\alpha}{E - \epsilon_\alpha} \left(1 + \frac{i}{2} \sum \frac{\Gamma_\alpha}{E - \epsilon_\alpha}\right)^{-1}, \tag{23}$$

where

$$\Gamma_\alpha = 2\pi |\langle \chi_\alpha, H_{QP} \psi_i^{(+)} \rangle|^2. \tag{24}$$

χ_α, ϵ_α are the eigenfunctions and the eigenvalues of a Hamiltonian H_{RR}. [See Ref. 7 for the definition of H_{RR} and other quantities in expressions (23) and (24).] Expression (23), which has the advantage that no arbitrary boundaries have to be used, is the same as the R-matrix expression if we choose the arbitrary boundary condition such that $\omega = 0$. The usual statistical assumptions[5] should now be made on the Hamiltonian H_{RR}. The randomness of the Hamiltonian H_{RR} will imply that its eigenfunctions χ_α and eigenvalues ϵ_α are also random. Assuming that the Hamiltonian H_{RR} is invariant under rotations and under time reversal, then a possible distribution of the eigenvalues ϵ_α will be of the form (5) and the distribution of the eigenvector components χ_α will be given by a random orthogonal matrix.[10] It has been shown by Feshbach[7] that the width Γ_α defined by expression (24) can be written as the square of a real amplitude. It is easy to see, then, that the joint distribution of the poles Z_μ, using Feshbach's theory,[7] will also be given by expression (22) with $\omega = 0$. Therefore, the distribution of the poles Z_μ using Feshbach's theory will be the same as the one obtained using R-matrix theory, provided we choose the arbitrary boundary condition in the later theory such that the parameter ω vanishes.

We now pass a remark about the use of the random complex orthogonal matrix[9] in the study of the parameters of the statistical collision matrix.[8] The joint distribution of the elements of a random complex orthogonal matrix is given by[9] the invariant volume element in the space of random complex orthogonal matrices multiplied by a statistical weight function that ensures the convergence of the probability integral. This weight function can now be constructed. A detailed account of this will be given later. The statistical properties of the parameters of the collision matrix which depend only on the invariant volume can always be worked out[9] without a knowledge of the weight function.

In these calculations care has to be taken that the relations obtained between the resonance parameters are consistent with unitarity, since the approximate forms of the collision matrix[8] may not be unitary.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Accidental Degeneracy in the Bethe–Salpeter Equation

DEBABRATA BASU AND S. N. BISWAS
Centre for Advanced Study in Physics, Department of Physics and Astrophysics
University of Delhi, Delhi-7, India

(Received 7 February 1969)

The Bethe–Salpeter equation for the bound state of the pion–nucleon system has been studied in the ladder approximation; the propagation time of the exchanged nucleon is neglected. By using the two-component formalism, the spinor equation is first reduced to a pair of simultaneous integral equations in momentum space. Following Fock, we transform these equations into ones in a four-dimensional hyperspace and the solutions are obtained in terms of series of $O(4)$ harmonics. As a simple illustration of our method, we have also considered the Bethe–Salpeter equation for the scalar-meson system. We find that the pion–nucleon Bethe–Salpeter equation shows an accidental degeneracy in the discrete-energy spectra similar to that in the solutions of the Dirac equation for the hydrogen atom, provided the coupling constant does not exceed a certain critical limit. The scalar problem exhibits at small binding energies a Schrödinger-type degeneracy. Convergence criteria for the pion–nucleon Bethe–Salpeter eigensolutions as series in $O(4)$ harmonics have been discussed; it is found that no solution exists when the coupling constant exceeds a certain critical value. In our approximation scheme, there are no abnormal solutions as are encountered in the fully covariant treatment of the equation.

1. INTRODUCTION

There has been a great deal of interest in the study of methods of solving the Bethe–Salpeter equation[1] for the relativistic two-body problem, particularly as a means of investigating the origin of dynamical symmetries. The Bethe–Salpeter equation is a fully covariant equation describing the interaction of a pair of elementary particles. Attempts to solve this equation in a fully covariant manner[2] led to many difficulties. The bound-state solutions of the homogeneous Bethe–Salpeter equation describing the interaction of a pair of nucleons was originally investigated in an approximation scheme in which (i) the interaction kernel was given in the ladder approximation, (ii) the mass of the exchanged meson was taken to be zero, and (iii) solutions were sought for the unphysical case where the total c.m. energy of the two-nucleon system was assumed to vanish. Goldstein[3] found a solution for this special case; however, the solution led to a continuous spectrum rather than a discrete one. The major difficulty is largely due to the existence of the extra degree of freedom of the relative-time variable which has no analog in nonrelativistic quantum mechanics, as a consequence of which the boundary conditions to be imposed on the solutions of the Bethe–Salpeter equation cannot be clearly indicated.[4] Wick[5] and Cutkosky[6] shed some light on these questions by considering a modification of the Bethe–Salpeter equation for the bound states of two spinless particles interacting via a massless scalar meson. An exact solution with arbitrary binding energy is possible for this case if one adopts a physically reasonable condition on the wavefunction which is also necessary for the existence of its Fourier transform.

In addition to the studies of the Bethe–Salpeter equation for bound-state solutions, there are a number of discussions on the analytic properties of the scattering amplitude within the formalism of the inhomogeneous Bethe–Salpeter equation.[7] It is known that scattering at high energies is determined not only by the fundamental Regge poles, but by a number of daughter poles taken in conjunction with the fundamental ones. It is generally believed that the existence of these daughter poles is intimately related to the higher symmetry inherent in the structure of the Bethe–Salpeter equation at zero energy.[8] The existence

[1] M. Gell-Mann and F. E. Low, Phys. Rev. **84**, 350 (1951); E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).

[2] J. S. Goldstein, Phys. Rev. **91**, 1516 (1953); H. S. Green, Phys. Rev. **97**, 540 (1955); G.-C. Wick, Phys. Rev. **96**, 1124 (1954); R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954); F. L. Scarf, Phys. Rev. **100**, 912 (1954); H. S. Green and S. N. Biswas, Progr. Theoret. Phys. (Kyoto), **18**, 121 (1954); S. N. Biswas and H. S. Green, Nucl. Phys. **2**, 177 (1956); H. S. Green, Nuovo Cimento **5**, 866 (1957); S. N. Biswas, Nuovo Cimento **7**, 577 (1958); S. Okubo and A. Feldman, Phys. Rev. **117**, 279, 292 (1960); L. H. D. Reeves, Ph.D. thesis, Adelaide University, 1962.

[3] See Ref. 2.

[4] This has also been noted by K. Rothe, Phys. Rev. **170**, 1550 (1968). Some radially symmetric solutions were obtained (see Biswas and Green, Ref. 2) by requiring that the solution of the Bethe–Salpeter equation together with its spatio-temporal derivatives should be finite and continuous, and particularly on the light cone.

[5] See Ref. 2.

[6] See Ref. 2.

[7] D. Z. Freedman and J. M. Wang, Phys. Rev. **153**, 1596 (1967); D. Z. Freedman, C. E. Jones, and J. M. Wang, Phys. Rev. **155**, 1645 (1967); A. R. Swift (report of work prior to publication); B. B. Deo and R. E. Cutkosky, Phys. Rev. **174**, 1859 (1968); B. W. Lee and R. F. Sowyer, Phys. Rev. **127**, 2266 (1962); H. S. Green and S. N. Biswas, Phys. Rev. **171**, 1511 (1968); N. P. Chang and R. P. Saxena (report of work prior to publication).

[8] G. Domokos and P. Suranyi, Nucl. Phys. **54**, 529 (1964).

of a hidden or dynamical symmetry was first noted by Cutkosky,[9] who transformed the equation by applying a stereographic projection to a five-dimensional pseudosphere. This has later been noted by several other authors.[10]

Although the covariant treatment of the Bethe–Salpeter equation has thus been the subject of several investigations, some hope has also been fostered that some aspects of the equation might be understood using a noncovariant approximation. For instance, Klein[11] examined the two-nucleon system rather exhaustively in the ladder approximation; however, he neglected the time of propagation of the exchanged virtual meson. This second approximation is commonly known as the "instantaneous interaction approximation." Recently, a similar type of approximation[12] has frequently been used in the study of the two-body scattering problem in the ladder approximation. We would like to consider in this paper the bound state solutions of pion–nucleon Bethe–Salpeter equation in the ladder approximation, neglecting the time of propagation, of the exchanged nucleon. A relativistic treatment of this problem has been considered by Rothe,[13] using the variational techniques of Schwartz,[14] which have led to renewed interest in the Bethe–Salpeter equation. We would like to solve the pion–nucleon Bethe–Salpeter equation in the momentum space by adopting the method of Fock,[15] who originally solved the nonrelativistic Schrödinger integral equation in momentum space and showed that the accidental degeneracy of the hydrogen atom is related to the invariance of Schrödinger equation under transformations of the $O(4)$ group. Levy[16] later showed how the $O(4)$ harmonics in momentum space could also be used to solve the Dirac equation.

In Sec. 2 we have considered the pion–nucleon bound-state equation. We show that, by adopting a two-component decomposition for the four-component spinor pion–nucleon wavefunction, this bound-state equation can be reduced to a pair of simultaneous integral equations in two momentum-space radial functions. To solve these equations, we go over to the four-dimensional hyperspace by means of a suitable substitution. To illustrate this method, we discuss the solution of the simpler scalar problem with various suitable approximations. An important consequence of the use of this method in this particular case is that the eigenvalue problem is completely determinate and an accidental degeneracy of the Schrödinger-hydrogen type emerges for small binding energies. This has been done in Sec. 3. The general method of the determination of the eigenvalue problem for arbitrary energies for the scalar problem has also been discussed here. In Sec. 4 we extend this method to solve the coupled equations for the pion–nucleon system [Eqs. (16a) and (16b)]. We find that the equations can be solved by a pair of infinite sums of $O(4)$ harmonics with arbitrary coefficients. The problem of determination of the energy eigenvalue reduces to that of finding the necessary compatibility condition for the existence of solutions for the coefficients of $O(4)$ harmonics which satisfy a pair of difference equations. We again find that, for binding energies, the energy-eigenvalues are discrete and this leads to the so-called accidental degeneracy of the Dirac-hydrogen type. We must remark that in solving these equations we retain only the most singular parts of the respective potentials which appear in the scalar-meson and the pion–nucleon problems. The modifications in the solution due to the inclusion of the rest of the potential are also discussed. We may remark that the various features of the results we obtain may be closely related to the nature of the various approximations we have used. The existence of bound-state solutions in various channels has been pointed out; in particular we obtain, as expected, the $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{3}{2}, \frac{3}{2})$ states. Eigensolutions for these states correspond to normal solutions so long as coupling strength has a certain upper bound. Solutions disappear when the coupling exceeds this critical value. These features have been discussed in the last section.

2. THE PION–NUCLEON BETHE–SALPETER EQUATION AND REDUCTION INTO TWO COMPONENTS

The relativistic two-body wavefunction for the pion and nucleon[13] in a state β is defined by

$$\Psi^i(x_1, x_2) = \langle 0| T\{\Psi(x_1)\varphi^i(x_2)\} |\beta\rangle, \qquad (1)$$

where $\varphi^i(x)$ and $\Psi(x)$ are Heisenberg fields of the pion and nucleon, respectively, $|0\rangle$ is the physical vacuum, and T is the time-ordering operator.

[9] See Ref. 2.

[10] R. Delbourgo, A. Salam, and J. Strathdee, ICTP Preprint No. IC/66/60; A. O. Barut, P. Budini, and C. Fronsdal, ICTP preprint No. IC/65/34; S. N. Biswas, J. Math. Phys. **8**, 1109 (1967); E. Kyriakopoulos, Phys. Rev. **174**, 1846 (1968). For a general discussion on dynamical symmetry, see N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Letters **15**, 1041 (1965).

[11] A. Klein, Phys. Rev. **90**, 1101 (1953); **91**, 740 (1953); **92**, 1017 (1953); **94**, 1052 (1954). See also J. S. Goldstein (Ref. 2).

[12] This type of approximation is also known in literature; see, for example, A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **39**, 380 (1963). See also R. Blankenbecler and R. Sugar, Phys. Rev. **142**, 1031 (1960).

[13] K. Rothe, Phys. Rev. **170**, 1548 (1968).

[14] C. Schwartz, Phys. Rev. **137**, B717 (1965); C. Schwartz and C. Zemach, Phys. Rev. **141**, 1454 (1966).

[15] V. Fock, Z. Physik **98**, 145 (1935).

[16] M. Levy, Proc. Roy. Soc. (London) **204A**, 145 (1950).

If we restrict ourselves to the ladder approximation, we arrive at the following homogeneous integral equation for the two-body bound-state wavefunction in a specific isospin channel I:

$$\Psi^I(x_1, x_2) = G^I \int d^4x_1' \, d^4x_2' G_1(x_1 - x_1') \times G_2(x_2 - x_2') I(x_1' - x_2') \Psi^I(x_2', x_1'), \quad (2)$$

where

$$G^I = \eta(I) g^2/(4\pi)^2,$$
$$\eta(I) = -1, \quad I = \tfrac{1}{2},$$
$$= 2, \quad I = \tfrac{3}{2}.$$

$G_1(x)$ and $G_2(x)$ are the usual nucleon and pion Green's functions. $I(x)$ is the interaction function and, if we restrict ourselves to the nucleon-exchange diagram only, it is given by

$$I(x) = \frac{1}{i\pi^2} \int \frac{d^4k e^{-ikx}}{\gamma \cdot k + M}, \quad (3)$$

where M is the mass of the exchanged nucleon.

Henceforth we shall suppress the isospin labels and the subscripts 1 and 2 will always refer to the nucleon and the pion, respectively.

Now, by writing the wavefunction as the product of two terms describing the center-of-mass motions and the relative motion of the two particles, Eq. (2) reduces in momentum space to

$$\psi(k) = \frac{if(m_1 + \gamma^0 E/2 + \gamma \cdot k)}{D(k)} \times \int d^4k' \frac{1}{\gamma \cdot (k + k') + M} \psi(k'), \quad (4)$$

where

$$f = G^I/\pi^2$$

and

$$D(k) = [(k^0 + E/2)^2 - (\mathbf{k}^2 + m_1^2)] \times [(k^0 - E/2)^2 - (\mathbf{k}^2 + m_2^2)].$$

We would like to study this Eq. (4) for the bound-state solutions. As mentioned in the Introduction, the boundary conditions to be imposed to extract the discrete solutions are not exactly known. We therefore make the following noncovariant approximation, namely, that the propagation time for the exchanged nucleon is neglected. Thus we write

$$I(x) = V(x)\delta(t). \quad (5)$$

With this approximation, Eq. (4) reduces to the following:

$$(m_1 + \gamma^0 E/2 + \gamma \cdot k)^{-1} \psi(k) D(k) = if \int \frac{d^4k'}{M - \boldsymbol{\gamma} \cdot (\mathbf{k} + \mathbf{k}')} \psi(k'). \quad (6)$$

The compatibility requirement for Eq. (6) suggests that we put

$$(m_1 + \gamma^0 E/2 + \gamma \cdot k)^{-1} \psi(k) D(k) = S(\mathbf{k}). \quad (7)$$

Then $S(\mathbf{k})$ satisfies

$$S(\mathbf{k}) = if \int \frac{d^3k' \, dk^{0'} (m_1 + \gamma^0 E/2 + \gamma \cdot k')}{\{M - \boldsymbol{\gamma} \cdot (\mathbf{k} + \mathbf{k}')\} D(k')} S(\mathbf{k}'). \quad (8)$$

The $k^{0'}$ integration in (8) can be performed at once to give a three-dimensional integral equation. We define a new function φ through

$$\varphi(\mathbf{k}) = N(\mathbf{k}) S(\mathbf{k}), \quad (9)$$

where

$$N(\mathbf{k}) = [(m_1 - \boldsymbol{\gamma} \cdot \mathbf{k}) J(\mathbf{k}) + \gamma^0 E L(\mathbf{k})]/T(\mathbf{k}),$$

with

$$J(\mathbf{k}) = (\mathbf{k}^2 + m_1^2)^{-\frac{1}{2}} + (\mathbf{k}^2 + m_2^2)^{-\frac{1}{2}},$$
$$L(\mathbf{k}) = (\mathbf{k}^2 + m_2^2)^{-\frac{1}{2}},$$
$$T(\mathbf{k}) = E^2 - \{(\mathbf{k}^2 + m_1^2)^{\frac{1}{2}} + (\mathbf{k}^2 + m_2^2)^{\frac{1}{2}}\}^2.$$

The integral equation for φ can be easily obtained from Eq. (8). To avoid inessential complications in our study, we assume that the pion–nucleon mass difference is not very large. Under this approximation we have the following equation for $\varphi(\mathbf{k})$:

$$R(\mathbf{k})[(m_1 - \gamma^0 E_m) + \boldsymbol{\gamma} \cdot \mathbf{k}] \varphi(\mathbf{k}) = \frac{\lambda}{2\pi^2} \int \frac{d^3k'}{M - \boldsymbol{\gamma} \cdot (\mathbf{k} + \mathbf{k}')} \varphi(\mathbf{k}'), \quad (10)$$

where

$$\lambda/2\pi^2 = f\pi,$$
$$E_m = E/[1 + (m_2/m_1)],$$

and

$$R(\mathbf{k}) = T(\mathbf{k}) J(\mathbf{k}) / \{(\mathbf{k}^2 + m_1^2) J^2(\mathbf{k}) - E^2 L^2(\mathbf{k})\}. \quad (11)$$

It is evident from (10) that $\varphi(\mathbf{k})$ is a four-component object; we then write φ as

$$\varphi(\mathbf{k}) = \begin{pmatrix} \varphi^+(\mathbf{k}) \\ \varphi^-(\mathbf{k}) \end{pmatrix}, \quad (12)$$

where φ^+ and φ^- play roles analogous to the large and small components of the usual Dirac spinor.

To separate the angular variables, we write the spinor φ as proportional to three-dimensional spherical harmonics. We note that for the pion–nucleon system, we have the total angular momentum $j = l + \frac{1}{2}$ and $j = l - \frac{1}{2}$, and further, because of parity conservation, a transition from $j = l + \frac{1}{2}$ to $l - \frac{1}{2}$ is forbidden. In particular, there are two types of solutions corresponding, respectively, to $j = l + \frac{1}{2}$ and $j = l - \frac{1}{2}$. For $j = l + \frac{1}{2}$, we have the following

angular dependence [see Appendix, Sec. A1]:

$$\varphi^+(\mathbf{k}) = g(k)\begin{pmatrix} \left(\frac{l+m+1}{2l+1}\right)^{\frac{1}{2}} & Y_l^m(\theta,\varphi) \\ -\left(\frac{l-m}{2l+1}\right)^{\frac{1}{2}} & Y_l^{m+1}(\theta,\varphi) \end{pmatrix},$$

$$\varphi^-(\mathbf{k}) = f(k)\begin{pmatrix} \left(\frac{l-m+1}{2l+3}\right)^{\frac{1}{2}} & Y_{l+1}^m(\theta,\varphi) \\ \left(\frac{l+m+2}{2l+3}\right)^{\frac{1}{2}} & Y_{l+1}^{m+1}(\theta,\varphi) \end{pmatrix}, \quad (13)$$

where $g(k)$ and $f(k)$ are two radial, momentum-space functions to be determined through Eq. (10). Substituting Eq. (12) in Eq. (10), we obtain the following set of equations:

$$R(\mathbf{k})[(m_1 - E_m)\varphi^+(\mathbf{k}) + \boldsymbol{\sigma}\cdot\mathbf{k}\varphi^-(\mathbf{k})] = \frac{\lambda}{2\pi^2}\int \frac{d^3k'}{(\mathbf{k}+\mathbf{k}')^2 + M^2}\{M\varphi^+(\mathbf{k}') + \boldsymbol{\sigma}\cdot\mathbf{k}'\varphi^-(\mathbf{k}') + \boldsymbol{\sigma}\cdot\mathbf{k}\varphi^-(\mathbf{k}')\}, \quad (14a)$$

$$R(\mathbf{k})[(m_1 + E_m)\varphi^-(\mathbf{k}) - \boldsymbol{\sigma}\cdot\mathbf{k}\varphi^+(\mathbf{k})] = \frac{\lambda}{2\pi^2}\int \frac{d^3k'}{(\mathbf{k}+\mathbf{k}')^2 + M^2}\{M\varphi^-(\mathbf{k}') - \boldsymbol{\sigma}\cdot\mathbf{k}'\varphi^+(\mathbf{k}') - \boldsymbol{\sigma}\cdot\mathbf{k}\varphi^+(\mathbf{k}')\}. \quad (14b)$$

If we note that

$$\boldsymbol{\sigma}\cdot\hat{\mathbf{k}}\begin{pmatrix} \left(\frac{l-m+1}{2l+3}\right)^{\frac{1}{2}} & Y_{l+1}^m(\theta,\varphi) \\ \left(\frac{l+m+2}{2l+3}\right)^{\frac{1}{2}} & Y_{l+1}^{m+1}(\theta,\varphi) \end{pmatrix} = \begin{pmatrix} \left(\frac{l+m+1}{2l+1}\right)^{\frac{1}{2}} & Y_l^m(\theta,\varphi) \\ -\left(\frac{l-m}{2l+1}\right)^{\frac{1}{2}} & Y_l^{m+1}(\theta,\varphi) \end{pmatrix} \quad (15)$$

[which follows easily through the use of the recursion relations, Eqs. (A3)], we finally obtain the following set of simultaneous integral equations for g and f:

$$X(\rho^2)[g(\rho) + \Gamma\rho f(\rho)] = \Gamma 4\pi\frac{\lambda_l}{2\pi^2}\left[\mu\int d\rho'\,\frac{\rho'^2 g(\rho')Q_l(\xi)}{2\rho\rho'} + \frac{2\Gamma}{(\Gamma^2+1)}\int d\rho'\,\frac{\rho'^2}{2\rho\rho'}\{\rho' Q_l(\xi) - \rho Q_{l+1}(\xi)\}f(\rho')\right], \quad (16a)$$

$$X(\rho^2)[\Gamma f(\rho) - \rho g(\rho)] = 4\pi\frac{\lambda_l}{2\pi^2}\left[-\mu\int d\rho'\,\frac{\rho'^2 f(\rho')Q_{l+1}(\xi)}{2\rho\rho'} + \frac{2\Gamma}{(\Gamma^2+1)}\int d\rho'\,\frac{\rho'^2}{2\rho\rho'}\{\rho' Q_{l+1}(\xi) - \rho Q_l(\xi)\}g(\rho')\right], \quad (16b)$$

where

$$E_m = m_1\epsilon, \quad |\mathbf{k}| = k = m_1(1-\epsilon^2)^{\frac{1}{2}}\rho, \quad \Gamma = [(1+\epsilon)/(1-\epsilon)]^{\frac{1}{2}}, \quad \mu = M/m_1, \quad \xi = (\rho^2 + \rho'^2 + \mu_\epsilon^2)/2\rho\rho', \quad (17)$$

$$\mu_\epsilon = [\mu^2/(1-\epsilon^2)]^{\frac{1}{2}}, \quad X(\rho^2) = \frac{1}{m_1}[R(\mathbf{k})]_{k^2 = m_1^2(1-\epsilon_2)\rho^2}, \quad \lambda_l = -\lambda(-)^{l-1}.$$

We now wish to solve this pair of equations for the energy eigenvalue problem using $O(4)$ spherical harmonics. To illustrate the method of solution which we will adopt, we consider the Bethe–Salpeter equation for the scalar-meson system under approximations similar to those we have just used for the pion–nucleon system.

3. SOLUTION OF SCALAR BETHE–SALPETER EQUATION USING $O(4)$ HARMONICS

The Bethe–Salpeter equation for the interaction of a pair of scalar mesons via scalar-meson exchange[17] is given by

$$[(p_4 + E)^2 + \mathbf{p}^2 + m^2][(p_4 - E)^2 + \mathbf{p}^2 + m^2]\varphi(p) = \frac{\lambda}{\pi^2}\int \varphi(q)\{(p-q)^2 - \mu^2\}^{-1}. \quad (18)$$

In this section we illustrate our general method for solving Eqs. (17) and (18) using hyperspherical harmonics. First we reduce Eq. (18) to the following three-dimensional equation, neglecting the propagation time of the exchanged scalar meson. Equation (18), thus approximated, reads

$$S(\mathbf{p}) = \frac{\lambda}{2\pi}\int \frac{(1+\mathbf{k}^2)^{-\frac{1}{2}}S(\mathbf{k})\,d^3k}{(c^2+\mathbf{k}^2)\{(\mathbf{k}-\mathbf{p})^2 + \mu^2\}}, \quad (19)$$

where

$$\varphi(p) = \{(c^2 - p^2)^2 - 4\epsilon^2 p_4^2\}S(\mathbf{p}), \quad c = (1-\epsilon^2)^{\frac{1}{2}}.$$

We have also put $\epsilon = E/m$, with $m = 1$.

Separation of the angular variables in Eq. (19) gives

$$X(\rho^2)(1+\rho^2)u(\rho) = \frac{\lambda}{2\pi c}\cdot 4\pi\int d\rho'\,\frac{\rho'^2 u(\rho')Q_l(\xi)}{2\rho\rho'}, \quad (20)$$

[17] For the notations and conventions used in this section, see S. N. Biswas (Ref. 10).

where $u(\rho)$ is a suitably defined wavefunction and we have put

$$|\mathbf{k}| = c\rho, \quad \mu_\epsilon = [\mu^2/(1-\epsilon^2)]^{\frac{1}{2}},$$
$$X(\rho^2) = \{1 + (1-\epsilon^2)\rho^2\}^{\frac{1}{2}}.$$

In order to solve the above integral equation, we shall expand the kernel in a series in powers of μ_ϵ. The requirement for the convergence of this series will certainly impose some restrictions as to the domain of validity of the solution subsequently obtained. However, it is easily seen that for sufficiently small μ Eq. (20) reduces to

$$X(\rho^2)(1+\rho^2)u(\rho) = \frac{\lambda}{2\pi c}\sum_{p=0}^{\infty}(-1)^p(\mu_\epsilon)^{2p}\int\frac{P_l(\cos\theta')u(\rho')\,d^3\rho'}{(\rho^2+\rho'^2-2\rho\rho'\cos\theta')^{p+1}} + t\mu_\epsilon u(\rho), \quad (21)$$

where Ω' is the entire three-dimensional space spanned by $\boldsymbol{\rho}'$ and Ω_0 is a small sphere centered at $(0, 0, \rho)$ and $t = (2\lambda/c)(1 - \frac{1}{4}\pi)$.

First we solve Eq. (20), assuming that the exchanged scalar meson is of vanishing mass (thereby retaining only the most singular part of the potential) and that the binding energy is small. We can facilitate the solution by transforming the variables according to a general transformation initiated by Fock.

We go over into Fock's hyperspace through the substitutions

$$\rho = \tan(\psi/2), \quad \rho' = \tan(\psi'/2). \quad (22)$$

Equation (20) is thus transformed to

$$(\alpha + \beta\cos\psi)v(\psi) = \frac{\nu}{2\pi^2}(1+\cos\psi)\int d\Omega'\,\frac{P_l(\cos\theta')v(\psi')}{4\sin^2(\Theta/2)}, \quad (23)$$

where

$$v(\psi) = u[\tan(\psi/2)]/\cos^4(\psi/2),$$
$$\alpha = (3-\epsilon^2)/2, \quad \beta = (1+\epsilon^2)/2, \quad \nu/2\pi^2 = \lambda/4\pi c,$$

and Θ is the angle between two unit vectors (in the four-dimensional space) of polar angles $(\psi, 0, 0)$ and $(\psi', \theta', \varphi')$; $d\Omega'$ is now the four-dimensional solid angle.

We now attempt a solution of (26) in the following form:

$$v(\psi) = \sum_{k=0}^{\infty} a_k P^{(2)}_{n+k,l}(\cos\psi). \quad (24)$$

The functions $P^{(2)}_{n+k,l}(\cos\psi)$ are related to the Gegenbauer polynomials C_n^ν and to the associated Legendre polynomials in a simple manner [see Eqs. (A4), (A5)].

We now substitute (24) in (23) and note that

$$\int d\Omega'\,\frac{P^{(2)}_{n,l}(\cos\psi')P_l(\cos\theta')}{4\sin^2(\Theta/2)} = \lambda_n P^{(2)}_{n,l}(\cos\psi), \quad (24')$$

with

$$\lambda_n = \frac{2\pi^{\frac{3}{2}}}{\Gamma(\frac{3}{2})}\int_{-1}^{1}dx\,\frac{P^{(2)}_{n,0}(x)(1-x^2)^{\frac{1}{2}}}{2(1-x)} = \frac{2\pi^2}{n+1}. \quad (24'')$$

The particular result given in Eq. (24') is a consequence of the general theorem due to Hecke on integral equations satisfied by hyperspherical harmonics. For completeness we have stated the theorem in the Appendix [see Eq. (A6)].

Using (24) and noting the recurrence relation

$$\cos\psi P^{(2)}_{n,l}(\cos\psi) = \frac{(n+2)(n-l+1)}{2(n+1)^2}P^{(2)}_{n+1,l} + \frac{n(n+l+1)}{2(n+1)^2}P^{(2)}_{n-1,l}, \quad (25)$$

we obtain (24) as a solution of (23), the coefficients a_k being determined by

$$a_k\left(\alpha - \frac{\nu}{n+k+1}\right) + a_{k-1}\frac{(n+k+1)(n+k-l)}{2(n+k)^2}\left(\beta - \frac{\nu}{n+k}\right) + a_{k+1}\frac{(n+k+1)(n+k+l+2)}{2(n+k+2)^2} \times\left(\beta - \frac{\nu}{n+k+2}\right) = 0. \quad (26)$$

This is a second-order difference equation; the criteria for the existence of solutions to this difference equation determine the energy-eigenvalue problem. The condition is given in terms of the following infinite-continued fraction[18]:

$$\frac{L_0}{M_0} = \cfrac{K_1/M_1}{\cfrac{L_1}{M_1} - \cfrac{K_2/M_2}{\cdots}}, \quad (27)$$

where M_k, L_k, and K_k are the coefficients occurring in Eq. (26),

$$M_k a_{k+1} + L_k a_k + K_k a_{k-1} = 0. \quad (28)$$

Without elaborating on this condition any further, we simply note that if we neglect terms of order $|\mathbf{k}|^2/m$ in α and β, then the equation is satisfied by

$$v(\psi) = P^{(2)}_{n,l}(\cos\psi),$$

[18] L. M. Milne-Thomson, *The Calculus of Finite Differences* (Macmillan and Co., Ltd., London, 1951), Chap. 17.

with

$$\lambda/4\pi c = (n+1)/2\pi^2, \tag{29}$$

whence

$$c^2 = (1-\epsilon^2) = \lambda^2\pi^2/4(n+1)^2. \tag{30}$$

Thus in this particular case we arrive at the conclusion that the scalar Bethe–Salpeter equation admits an $O(4)$ symmetry, and energy values are given by (30), leading to the accidental degeneracy of the Schrödinger-hydrogen-atom type.

We now consider the case when the mass of the exchanged scalar meson is nonvanishing. Thus we consider Eq. (21) and treat it in the zeroth-order approximation. To generate the higher terms, we have first to solve the following equation:

$$(\alpha+\beta\cos\psi)v(\psi) = \frac{\nu}{2\pi^2}(1+\cos\psi)\int_{S'-S_0} d\Omega' \frac{P_l(\cos\theta')v(\psi')}{4\sin^2(\Theta/2)} + \tfrac{1}{4}t\mu_\epsilon(1+\cos\psi)^2 v(\psi), \tag{31}$$

where S' and S_0 are portions of the hypersphere into which Ω' and Ω_0 are transformed. Substituting (24) in (31) and noting that

$$\int_{S'-S_0} d\Omega' \frac{P^{(2)}_{n,l}(\cos\psi')P_l(\cos\theta')}{4\sin^2(\Theta/2)} = \lambda_n P^{(2)}_{n,l}(\cos\psi) \tag{32a}$$

with

$$\lambda_n = \frac{(2\pi)^{\frac{3}{2}}}{\Gamma(\frac{3}{2})}\int_{-1}^{1-\delta_0} dx \frac{P^{(2)}_{n,0}(x)(1-x^2)^{\frac{1}{2}}}{2(1-x)} = [2\pi^2/(n+1)] - 4\pi\mu_\epsilon(1+\cos\psi) + O(\mu_\epsilon^3),$$
$$\delta_0 = 2\mu_\epsilon^2\cos^4(\psi/2), \tag{32b}$$

we get the following recurrence relation to determine the coefficients a_k:

$$L_k a_{k+2} + M_k a_{k+1} + N_k a_k + Q_k a_{k-1} + R_k a_{k-2} = 0,$$

where L_k, M_k, N_k, Q_k, and R_k are given in the Appendix [see Appendix, Sec. A5].

This leads to a fourth-order difference equation; the general solution now becomes extremely involved. We refrain from studying this equation any further here. In the next section we discuss and obtain the solutions of pion–nucleon problem, using the methods developed in this section.

4. SOLUTION OF THE PION-NUCLEON BETHE–SALPETER EQUATION

We shall now study the solution of pion–nucleon Bethe–Salpeter equations (16) by employing the techniques discussed in the previous section. As before, we perform a power-series expansion of the function in the form

$$X(\rho^2) = A(\epsilon) + B(\epsilon)\rho^2,$$

where A and B are functions of energy and go over into Fock's hyperspace by means of the substitutions (22).

If we further define

$$G(\psi) = g[\tan(\psi/2)]/\cos^6(\psi/2),$$
$$F(\psi) = f[\tan(\psi/2)]/\cos^6(\psi/2), \tag{33}$$

the transformed equations, on retaining only the most singular part of the potential as before, take the following form:

$$(\sigma+\alpha\cos\psi+\beta\cos^2\psi)[(1+\cos\psi)G(\psi) + \Gamma\sin\psi F(\psi)] = \frac{\lambda_l\Gamma}{2\pi^2}\Big[\mu(1+\cos\psi)\int d\Omega' \frac{(1+\cos\psi')G(\psi')P_l(\cos\theta')}{4\sin^2(\Theta/2)} + \frac{2\Gamma}{(\Gamma^2+1)}\int d\Omega' \frac{\{(1+\cos\psi)\sin\psi' P_l(\cos\theta') - \sin\psi(1+\cos\psi')P_{l+1}(\cos\theta')\}F(\psi')}{4\sin^2(\Theta/2)}\Big], \tag{34a}$$

$$(\sigma+\alpha\cos\psi+\beta\cos^2\psi)[\Gamma(1+\cos\psi)F(\psi) - \sin\psi G(\psi)] = \frac{\lambda_l}{2\pi^2}\Big[-\mu(1+\cos\psi)\int d\Omega' \frac{(1+\cos\psi')F(\psi')P_{l+1}(\cos\theta')}{4\sin^2(\Theta/2)} + \frac{2\Gamma}{(\Gamma^2+1)}\int d\Omega' \frac{\{(1+\cos\psi)\sin\psi' P_{l+1}(\cos\theta') - \sin\psi(1+\cos\psi')P_l(\cos\theta')\}G(\psi')}{4\sin^2(\Theta/2)}\Big], \tag{34b}$$

where

$$\lambda_l = \eta(I)(-)^l g^2/8\pi, \tag{35a}$$

$$\sigma = \frac{(\delta^2-1)(\delta+1)}{\delta^2} - \frac{(1+\delta^3)(1-\epsilon^2)}{2\delta^2} - \frac{(1+\delta)}{\delta},$$
$$\beta = \frac{(1+\delta)}{\delta} + \frac{(1+\delta^3)(1-\epsilon^2)}{2\delta^2} - \frac{(1+\delta)(1+\delta^2)}{\delta^2}, \tag{35b}$$
$$\alpha = \sigma+\beta, \quad \delta = m_1/m_2.$$

This pair of equations is very similar to the corresponding equations one obtains in treating, in momentum space, the Dirac electron in an external potential. We are thus tempted to assume the following types of solutions, given by infinite series of $O(4)$ spherical harmonics:

$$F(\psi) = \sum_{k=0}^{\infty} a_k \sin^{-\frac{1}{2}} \psi P^{-(l+\frac{3}{2})}_{\nu+k+\frac{1}{2}}(\cos\psi),$$
$$G(\psi) = \sum_{k=0}^{\infty} b_k \sin^{-\frac{1}{2}} \psi P^{-(l+\frac{1}{2})}_{\nu+k+\frac{1}{2}}(\cos\psi). \tag{36}$$

Since each of $P_{n,l}$'s in Eqs. (36) should be non-vanishing, we then have $\nu = n + l + 1$, where $n \geq 0$ [see equation connecting C^{ν}_{n}'s and P^{ν}_{λ}'s in the Appendix, (A5)].

To obtain the solution of our Bethe–Salpeter equation for the case $j = l - \frac{1}{2}$, we replace l by $l - 1$ everywhere in the above expressions. Hence the rhs of (35a) should read $\eta(I)(-)^{l-1}g^2/8\pi$. Combining these two cases, we then see that the effective coupling strength is

$$\lambda_l \equiv \lambda_j = (-)^{j-\frac{1}{2}}\eta(I)g^2/8\pi \tag{37}$$

for a channel specified by (j, I), and further that

$$\nu = n + j + \tfrac{1}{2}; \quad n \geq 0. \tag{38}$$

We now substitute the expressions for F and G given by Eqs. (34a) and (34b). With proper use of Hecke's theorem [see Eq. (24)] and the recursion relations [see Eqs. (A3)], we obtain the following pair of recurrence relations between the coefficients a_k and b_k:

$$\begin{aligned}
&\sigma\Lambda^{(1)}_k + \frac{(\nu+k-l+1)}{2(\nu+k+2)}\left[\alpha\Lambda^{(1)}_{k+1} + \beta\frac{(\nu+k-l+2)}{2(\nu+k+3)}\Lambda^{(1)}_{k+2} + \beta\frac{(\nu+k+l+2)}{2(\nu+k+1)}\Lambda^{(1)}_{k}\right] \\
&\qquad + \frac{(\nu+k+l+1)}{2(\nu+k)}\left[\alpha\Lambda^{(1)}_{k-1} + \beta\frac{(\nu+k-l)}{2(\nu+k+1)}\Lambda^{(1)}_{k} + \beta\frac{(\nu+k+l)}{2(\nu+k-1)}\Lambda^{(1)}_{k-2}\right] \\
&\qquad = \Gamma\lambda_j\left[\mu\left\{\frac{\Delta^{(b)}_k}{(\nu+k+1)} + \frac{(\nu+k+l+1)}{2(\nu+k)^2}\Delta^{(b)}_{k-1} + \frac{(\nu+k-l+1)}{2(\nu+k+2)^2}\Delta^{(b)}_{k+1}\right\}\right. \\
&\qquad\quad \left. + \frac{\Gamma}{(\Gamma^2+1)(\nu+k+1)}\left\{\frac{a_{k-1}}{(\nu+k)^2} + \frac{2a_k}{(\nu+k)(\nu+k+2)} + \frac{a_{k+1}}{(\nu+k+2)^2}\right\}\right],
\end{aligned} \tag{39a}$$

$$\begin{aligned}
&\sigma\Lambda^{(2)}_k + \frac{(\nu+k+l+2)}{2(\nu+k)}\left[\alpha\Lambda^{(2)}_{k-1} + \beta\frac{(\nu+k+l+1)}{2(\nu+k-1)}\Lambda^{(2)}_{k-2} + \beta\frac{(\nu+k-l-1)}{2(\nu+k+1)}\Lambda^{(2)}_{k}\right] \\
&\qquad + \frac{(\nu+k-l)}{2(\nu+k+2)}\left[\alpha\Lambda^{(2)}_{k+1} + \beta\frac{(\nu+k+l+3)}{2(\nu+k+1)}\Lambda^{(2)}_{k} + \beta\frac{(\nu+k-l+1)}{2(\nu+k+3)}\Lambda^{(2)}_{k+2}\right] \\
&\qquad = \lambda_j\left[-\mu\left\{\frac{\Delta^{(a)}_k}{(\nu+k+1)} + \frac{(\nu+k+l+2)}{2(\nu+k)^2}\Delta^{(a)}_{k-1} + \frac{(\nu+k-l)}{2(\nu+k+2)^2}\Delta^{(a)}_{k+1}\right\}\right. \\
&\qquad\quad + \frac{\Gamma}{(\Gamma^2+1)(\nu+k+1)}\left\{\frac{(\nu+k+l+1)(\nu+k+l+2)}{(\nu+k)^2}b_{k-1}\right. \\
&\qquad\quad \left.\left. + \frac{2(\nu+k-l)(\nu+k+l+1)}{(\nu+k)(\nu+k+2)}b_k + \frac{(\nu+k-l)(\nu+k-l+1)}{(\nu+k+2)^2}b_{k+1}\right\}\right].
\end{aligned} \tag{39b}$$

The various symbols Λ_k and Δ_k which occur above are functions of the coefficients a_k and b_k. These are given by the following relations:

$$\begin{aligned}
\Lambda^{(1)}_k &= b_k + \frac{1}{2(\nu+k+2)}[\Gamma a_{k+1} + (\nu+k-l+1)b_{k+1}] - \frac{1}{2(\nu+k)}[\Gamma a_{k-1} - (\nu+k+l+1)b_{k-1}], \\
\Lambda^{(2)}_k &= \Gamma a_k + \frac{(\nu+k-l)}{2(\nu+k+2)}[\Gamma a_{k+1} + (\nu+k-l+1)b_{k+1}] \\
&\quad + \frac{(\nu+k+l+2)}{2(\nu+k)}[\Gamma a_{k-1} - (\nu+k+l+1)b_{k-1}], \\
\Delta^{(a)}_k &= \left[a_k + \frac{(\nu+k+l+2)}{2(\nu+k)}a_{k-1} + \frac{(\nu+k-l)}{2(\nu+k+2)}a_{k+1}\right], \\
\Delta^{(b)}_k &= \left[b_k + \frac{(\nu+k+l+1)}{2(\nu+k)}b_{k-1} + \frac{(\nu+k-l+1)}{2(\nu+k+2)}b_{k+1}\right].
\end{aligned} \tag{40}$$

5. EIGENVALUE PROBLEM AND ACCIDENTAL DEGENERACY

Equations (39a) and (39b) can now be used to determine all the coefficients a_k and b_k. The energy-eigenvalue problem is determinate and is obtained as a compatibility determinant for the existence of solutions for a_0, b_0, etc. We first note that

$$\Delta_k^{(a)} = \Delta_k^{(b)} = \Lambda_k^{(1)} = \Lambda_k^{(2)} = 0, \quad k = -2, -3, \cdots.$$

In order to find the energy-eigenvalue problem, we put $k = -3$ and -2 in Eqs. (39) and obtain the following set of linear homogeneous equations for the determination of a_0, b_0, and

$$c_0 \equiv \frac{\beta}{2(\nu+2)}[\Gamma a_1 + (\nu - l + 1)b_1]:$$

$$\begin{aligned} \Gamma a_0 + (\nu - l)b_0 &= 0, \\ (\beta\Gamma + \lambda_j\mu/2\nu)a_0 + (\nu - l)c_0 &= 0, \\ (\beta - \lambda_j\mu/2\nu)b_0 + c_0 &= 0. \end{aligned} \tag{41}$$

The compatibility requirement for the nontrivial solutions of a_0, etc., is

$$\lambda_j\mu\Gamma^2 - 4\beta\nu\Gamma - \lambda_j\mu = 0. \tag{42}$$

One can continue this process and obtain higher-order compatibility determinants for the homogeneous equations which determine the coefficients a_1, b_1, etc. It can be shown that these compatibility determinants are zero if condition (42) is satisfied (see Appendix, Sec. 6). The eigenvalue equation (42) leads to a complicated higher-order algebraic equation in Γ (even in the equal mass case), which we can study adopting numerical methods. However, if we make the simplifying assumption that the interacting particles have equal masses ($\delta = 1$) and that the binding energy is small, then (42) reduces to a quadratic equation

$$\lambda_j\mu\Gamma^2 + 8\nu\Gamma + \lambda_j\mu = 0, \tag{43}$$

where λ_j is given by (37) and ν by (38). Note that Γ is positive and greater than unity [see Eq. (17)]. Thus we look for that root of (43) which is positive and greater than unity. This is possible only when $\lambda_j < 0$. In such a situation, when the potential is attractive, the above equation possesses an admissible solution,

$$\Gamma = [\nu + (\nu^2 - \bar{\lambda}^2)^{\frac{1}{2}}]/\bar{\lambda}, \tag{44}$$

where $\bar{\lambda} = \frac{1}{4}\mu\,|\lambda_j|$, $\mu = 1$, and we have the restriction $\bar{\lambda} < 1$, so that

$$\Gamma > 1, \quad \nu = 1, 2, \cdots. \tag{45}$$

We shall see later that this condition is also necessary for the convergence of the series for $G(\psi)$ and $F(\psi)$ [Eqs. (36)]. As a result, for the following states characterized by

$$j = \tfrac{1}{2}, \tfrac{5}{2}, \cdots, \quad I = \tfrac{1}{2},$$
$$j = \tfrac{3}{2}, \tfrac{7}{2}, \cdots, \quad I = \tfrac{3}{2},$$

we have discrete energy spectra. The discrete energies of the spectrum are given by

$$\epsilon = [1 - \bar{\lambda}^2/(n + j + \tfrac{1}{2})^2]^{\frac{1}{2}}, \tag{46a}$$

with $n = 0, 1, 2, \cdots$,

$$\begin{aligned} \bar{\lambda} &= g^2/32\pi, \quad I = \tfrac{1}{2}, \\ \bar{\lambda} &= g^2/16\pi, \quad I = \tfrac{3}{2}. \end{aligned} \tag{46b}$$

We thus obtain the $(\frac{1}{2}, \frac{1}{2})$, $(\frac{3}{2}, \frac{3}{2})$ states as possible bound states and also conclude that the $(\frac{3}{2}, \frac{3}{2})$ state is more strongly bound than $(\frac{1}{2}, \frac{1}{2})$. There may be many more excited bound states corresponding to various values of n. There is no experimental knowledge of these states at present.

Equation (46) further shows that the $S_{\frac{1}{2}}$ and $P_{\frac{1}{2}}$ nucleon states are degenerate. This is akin to the accidental degeneracy one encounters in the Dirac-hydrogen atom problem. The existence of the degenerate $S_{\frac{1}{2}}$ and $P_{\frac{1}{2}}$ nucleon states should be compared with the results for the bound-state solutions of the fully covariant Bethe–Salpeter equation by Rothe.[13] Rothe finds that not only are the $S_{\frac{1}{2}}$ (0_+-channel) and $P_{\frac{1}{2}}$ (1_--channel) states nondegenerate, but also that the solutions to the Bethe–Salpeter equation in $P_{\frac{1}{2}}$ state are abnormal in that they correspond to discrete negative-energy spectra. Further, these abnormal solutions are related to the normal $S_{\frac{1}{2}}$ Bethe–Salpeter solutions corresponding to a positive-energy discrete spectrum through the existence of a generalized MacDowell symmetry. This states that if $\varphi(\lambda, \epsilon)$ is a solution of Bethe–Salpeter equation with $E = \epsilon$ for the channel $l_\pm$, then $\varphi(\lambda, -\epsilon)$ is also a solution of the Bethe–Salpeter equation for $(l \pm 1)_\mp$ channel with the same coupling strength. This is a consequence of the full Lorentz symmetry of the covariant Bethe–Salpeter equation. The existence of an abnormal solution may be attributed to the presence of the relative time in the fully covariant equation, which may be regarded as a new degree of freedom in the equation. In the present case, as we work in the instantaneous interaction approximation, this eliminates the extra degree of freedom from our modified Bethe–Salpeter equation. As a consequence, no abnormal solutions are present in our problem. The MacDowell[19] symmetry in the covariant case removes the accidental degeneracy we note in our solutions.

[19] S. W. MacDowell, Phys. Rev. **116**, 774 (1959).

Finally we must check the convergence of our series for $G(\psi)$ and $F(\psi)$, given by (36). Equations (41) and (44) show

$$|a_0/b_0| \approx O(\bar{\lambda}),$$

where $\bar{\lambda} < 1$ and is given by (46b). This indicates a possible identification of F and G as the small and large components. The ratio $|a_k/b_k|$ are all found to be of the order of $\bar{\lambda}$. In addition we find, for example, from (39) and (41),

$$|b_1/b_0| \approx O(\bar{\lambda}^2).$$

This shows that the two series can be convergent only when $\bar{\lambda} < 1$, a condition which is also necessary for the existence of the physically admissible solutions of (43). Hence we have convergent eigensolutions only if the coupling is weak and an accidental degeneracy in the energy spectra of the pion–nucleon Bethe–Salpeter equation emerges as in the case of Dirac-hydrogen atom problem. Our result is similar to that obtained by Tiktopoulos,[20] who showed that the Bethe–Salpeter equation for the positronium bound state exhibits a Schrödinger-type degeneracy in the weak coupling limit.

It is interesting to note that if the pion–nucleon coupling $\bar{\lambda}$, given by (46b), is large, Eq. (46) leads to unphysical values of the total energy and the corresponding wavefunctions $G(\psi)$ and $F(\psi)$ both diverge. This is probably a feature of the noncovariant approximation used in our treatment whose full implications are not easily discernible. The nonexistence of Bethe–Salpeter solutions for large coupling could be due to the retention of only the ladder diagrams in the interaction kernel; this may not be a valid approximation for large coupling strength. This type of nonanalytic behavior of the wavefunction with coupling constant has also been noted by Datta[21] in different connections.

ACKNOWLEDGMENT

We would like to thank Dr. K. Datta for kindly going through the manuscript and for interesting comments.

APPENDIX

1. Derivation of Equation (13)

To obtain the angular-momentum eigensolutions, we first set the small components equal to zero. Now from standard theory of angular momentum we can write the two-component wavefunction $\psi_{j\mu}$ for $j = l + \frac{1}{2}$ as

$$\psi_{j\mu} = \sum_{ms=-\frac{1}{2},\frac{1}{2}} C(l, \tfrac{1}{2}, l + \tfrac{1}{2}; \mu - m_s, m_s) Y_l^{\mu - m_s} \chi_{m_s}, \tag{A1}$$

where χ denotes the spin-$\frac{1}{2}$ wavefunction. Setting $\mu = m + \frac{1}{2}$, we have

$$\psi_{j,m+\frac{1}{2}} = \left(\frac{l+m+1}{2l+1}\right)^{\frac{1}{2}} Y_l^m \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \left(\frac{l-m}{2l+1}\right)^{\frac{1}{2}} Y_l^{m+1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence we attempt to write $\varphi^+_{j,m+\frac{1}{2}}$ for $j = l + \frac{1}{2}$ as

$$\varphi^+_{j,m+\frac{1}{2}} = \begin{pmatrix} g_1(k)\left(\dfrac{l+m+1}{2l+1}\right)^{\frac{1}{2}} & Y_l^m \\ g_2(k)\left(\dfrac{l-m}{2l+1}\right)^{\frac{1}{2}} & Y_l^{m+1} \end{pmatrix}, \tag{A2}$$

where g_1 and g_2 are to be found from Bethe–Salpeter equations. Since conservation of parity in pion–nucleon system demands that there cannot be any transition from $l = j + \frac{1}{2}$ to $l = j - \frac{1}{2}$, we have that $\boldsymbol{\sigma} \cdot \mathbf{L}$ is a good quantum number; hence from (A2) we find easily that

$$g_1^2 = g_2^2,$$

leading to

$$g_1 = \pm g_2.$$

In our case we choose the normalization such that $g_1 = -g_2 = g$. Thus we get Eq. (13) for φ^+. The equation for φ^- can also be obtained by an argument parallel to that of Dirac equation.

2. The Recursion Relations[22]

$$\begin{aligned} xP_\nu^\mu &= (2\nu+1)^{-1}\{(\mu+\nu)P_{\nu-1}^\mu \\ &\quad + (\nu-\mu+1)P_{\nu+1}^\mu\}, \\ (1-x^2)^{\frac{1}{2}}P_\nu^\mu &= (2\nu+1)^{-1}\{P_{\nu+1}^{\mu+1} - P_{\nu-1}^{\mu+1}\}, \\ -(1-x^2)^{\frac{1}{2}}P_\nu^{\mu+1} &= (2\nu+1)^{-1}\{(\nu+\mu)(\nu+\mu+1)P_{\nu-1}^\mu \\ &\quad - (\nu-\mu)(\nu-\mu+1)P_{\nu+1}^\mu\}. \end{aligned} \tag{A3}$$

3. The $P_{n,l}^{(2)}$ Functions

The $P_{n,l}^{(2)}$ functions used are simply related to the Gegenbauer polynomials C_n^ν:

$$P_{n,l}^{(2)}(\cos\psi) = \frac{1}{n+1} \sin^l\psi C_{n-l}^{l+1}(\cos\psi). \tag{A4}$$

[20] G. Tiktopoulos, J. Math. Phys. **6**, 573 (1965).

[21] K. Datta, Ph.D. thesis, Brandeis University, 1966. See, however, W. Kohn and C. Majumdar, Phys. Rev. **138**, A1617 (1965) for a nonrelativistic model in which the analytic dependence of the bound-state energy on the coupling constant is smooth in the transition to the continuum.

[22] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949).

The corresponding relation with the associated Legendre polynomials is given by[23]

$$m!\, C_m^{\lambda} = \Gamma(\lambda + \tfrac{1}{2})(2\lambda)_{m\frac{1}{4}}(x^2 - 1)^{\frac{1}{4}-\lambda/2} P_{m+\lambda-\frac{1}{2}}^{\frac{1}{2}-\lambda}(x). \tag{A5}$$

4. Hecke's Theorem[23]

Let $F(x)$ be a function of the real variable x which is continuous for $-1 \leq x \leq 1$, and $Y_n(\xi)$ be any general surface harmonic of degree n. Then, for any unit vector η,

$$\int_{\Omega(\xi)} F(\xi, \eta) Y_n(\xi)\, d\Omega(\xi) = \lambda_n Y_n(\eta), \tag{A6}$$

where

$$\lambda_n = \frac{2\pi^{\frac{1}{2}(q+1)} n!\,(q-1)!}{\Gamma\left(\frac{q+1}{2}\right)(n+q-1)!} \times \int_{-1}^{1} F(x) C_n^{q/2}(x)(1-x^2)^{\frac{1}{2}(q-1)}\, dx. \tag{A7}$$

5. The Various Coefficients $L_k \cdots R_k$

$$L_k = \frac{(n+k+l+3)(n+k+l+2)(n+k+1)}{4(n+k+3)^2(n+k+2)} \times \mu_\epsilon(t/4 - 2\nu/\pi),$$

$$M_k = \frac{(n+k+1)(n+k+l+2)}{2(n+k+2)^2} \times \left[\beta - \frac{\nu}{n+k+2} + 2\mu_\epsilon(t/4 - 2\nu/\pi)\right],$$

$$N_k = \alpha - \frac{\nu}{n+k+1} + \mu_\epsilon(t/4 - 2\nu/\pi) \times \left\{1 + \frac{(n+k+l+1)(n+k-l)}{4(n+k+1)(n+k)} + \frac{(n+k+l+2)(n+k-l+1)}{4(n+k+1)(n+k+2)}\right\},$$

$$Q_k = \frac{(n+k+1)(n+k-l)}{2(n+k)^2} \times \left\{\beta - \frac{\nu}{n+k} + 2\mu_\epsilon(t/4 - 2\nu/\pi)\right\},$$

$$R_k = \frac{(n+k+1)(n+k-l)(n+k-l-1)}{4(n+k-1)^2(n+k)} \times \mu_\epsilon(t/4 - 2\nu/\pi).$$

6. The Compatibility Condition

Higher-order determinants, for example, for the determination of a_0, b_0, a_1, b_1, $c_1 \equiv [\beta/2(\nu+3)] \times [\Gamma a_2 + (\nu - l + 2)b_2]$, can be reduced to the following:

$$\begin{vmatrix} -\frac{1}{\Gamma}\left(\beta\Gamma + \frac{\lambda_j\mu}{2\nu}\right) & \beta & \beta\Gamma \\ \left(\beta - \frac{\lambda_j\mu\Gamma}{2\nu}\right) & \beta & \beta\Gamma \\ B_0(\nu, \Gamma) & \left(\beta - \frac{\lambda_j\mu\Gamma}{\nu+1}\right) & -\left(\beta\Gamma - \frac{\lambda_j\mu}{\nu+1}\right) \end{vmatrix} = (\beta/\Gamma)(\lambda_j\mu\Gamma^2 - 4\beta\nu\Gamma - \lambda_j\mu) \times (\lambda_j\mu\Gamma^2 + 2\beta(\nu+1)\Gamma + \lambda_j\mu),$$

where we have used (41) and (39).

The vanishing of the determinant is thus guaranteed by the condition (42). This process can be continued to any order.

[23] *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill Book Co., Inc., 1953), Vol. 2.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Theory of Observables

PLATON C. DELIYANNIS
Nuclear Research Center "Democritos," Aghia Paraskevi Attikis, Athens, Greece

(Received 20 November 1968)

An alternative formulation of a general theory of observables is presented, which contains as special cases the systems proposed by Segal and Mackey. The basic properties are developed, and the exact relations to the aforementioned systems are deduced.

INTRODUCTION

Since the work of von Neumann in the thirties on the foundations of classical quantum mechanics, which lay dormant for the next 20 years,[1-3] the subject has been approached in two distinct abstract ways, initiated by Segal[4] and Mackey.[5] These axiomatic systems have several common features, but diverge at certain points in an essential way—for example, in Segal's system, one obtains a profusion of states, while in Mackey's their existence has to be postulated.

The purpose of this work is to propose and study a system of axioms designed to cover both approaches and still be rich enough to produce a reasonable theory. Further, it is based upon primitive concepts which, in the writer's opinion, are closer to the operational use of the various terms, thus perhaps achieving a more universal validity. These concepts are (i) the possibility of forming functions of observables, which is supposed to interpret the application of functions to the measured values of an observable, and (ii) the description of states through expectation value functionals, since it is the expectation value of an observable that one is trying to evaluate by measurements. We restrict attention to continuous functions of observables, first, because this seems to be the minimal class of functions one can work with effectively and, second, in order to avoid possible ambiguities in case of uncertainties in measuring. In the first part, we describe the system obtained and deduce its basic properties, showing that it contains as special cases the systems of Mackey and Segal. The second part is devoted to the development of a special system in order to obtain the exact connection to the work of Mackey. The relation to Segal's system is also discussed there. A construction of "glueing together" a set of vector spaces, required in order to explore the existence of states, is given in Appendix A. Appendices B and C contain certain special results needed in the main parts.

I. THE GENERAL SYSTEM

A. The Axioms

We assume the existence of a nonempty set $\mathcal{M}$ of objects, which we shall call observables, and that the algebra $\mathcal{F}$ of continuous functions from the reals to the reals acts on $\mathcal{M}$ to produce elements of $\mathcal{M}$. We shall write $f(A)$ for the result of applying $f \in \mathcal{F}$ to $A \in \mathcal{M}$, and we assume the following:

Axiom 1: Let $f_i(A) = g_i(B)$, $i = 1, 2, \cdots$, and suppose that the supports of the set $\{f_i\}$ and those of the set $\{g_i\}$ form locally finite systems. Then

$$\left(\sum_{i=1}^{\infty} f_i\right)(A) = \left(\sum_{i=1}^{\infty} g_i\right)(B) \quad \text{and} \quad (f_1 f_2)(A) = (g_1 g_2)(B).$$

Axiom 2: Let $\circ$ denote composition of functions. Then, for any $f, g \in \mathcal{F}$ and $A \in \mathcal{M}$, we have

$$(f \circ g)(A) = f(g(A)).$$

Axiom 3: If $f(A) = f(B)$ for all bounded $f \in \mathcal{F}$, then $A = B$. If 0, 1 denote the constant functions $x \to 0$, $x \to 1$, x real, then $0(A) = 0(B)$ and $1(A) = 1(B)$ for any $A, B \in \mathcal{M}$.

Remarks: The purpose of Axiom 1 is to allow us to introduce consistently the structure of an algebra on $\{f(A) \mid f \in \mathcal{F}\}$ (see Sec. B). The reason for including infinite sums is technical, but as it is desirable to avoid introducing from the beginning any topology on the observables or on $\mathcal{F}$, we restricted attention to infinite summations under which $\mathcal{F}$ is closed for algebraic rather than topological reasons. For a sequence f_i whose supports form a locally finite system (i.e., a class of sets such that each point has a neighborhood intersecting finitely many of the sets in this class), this is quite clear.

We first show that the axiom systems of Mackey[5] and Segal[4] satisfy the above axioms.

[1] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Julius Springer-Verlag, Berlin, 1932).
[2] J. von Neumann, P. Jordan, and E. Wigner, Ann. Math. **35**, 29 (1934).
[3] J. von Neumann and G. Birkhoff, Ann. Math. **37**, 823 (1936).
[4] I. E. Segal, Ann. Math. **48**, 930 (1947).
[5] G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W. A. Benjamin, Inc., New York, 1963).

Theorem 1: The axiom system of Mackey satisfies the above axioms.

Proof: Recall that an observable in the Mackey system is a σ-homomorphism A from the Borel sets of the real line into a given logic $\mathfrak{L}$. Thus it is determined by its values on the interval $(-\infty, r)$ where r is rational. Also recall that $f(A)$ is defined as the σ-homomorphism $A \circ f^{-1}$ (composition). We write for convenience $\{f < a\}$ for $\{x \mid f(x) < a\}$. Now assume that $A \circ f_i^{-1} = B \circ g_i^{-1}$ for $i = 1, 2, \cdots, n$. Since

$$\left\{\sum_{i=1}^{n} f_i < r\right\} = \bigcup \left\{ \bigcap_{i=1}^{n} \{f_i < \mu_i\} \,\middle|\, \sum_{i=1}^{n} \mu_i < r, \mu_i \text{ rationals} \right\}$$

and A, B are σ-homomorphisms, we obtain

$$\begin{aligned}
&\left(A \circ \left(\sum_{i=1}^{n} f_i\right)^{-1}\right)(-\infty, r) \\
&\quad = \bigcup \left\{ \bigcap_{i=1}^{n} (A \circ f_i^{-1})(-\infty, \mu_i) \,\middle|\, \sum_{i=1}^{n} \mu_i < r, \mu_i \text{ rationals} \right\} \\
&\quad = \bigcup \left\{ \bigcap_{i=1}^{n} (B \circ g_i^{-1})(-\infty, \mu_i) \,\middle|\, \sum_{i=1}^{n} \mu_i < r, \mu_i \text{ rationals} \right\} \\
&\quad = \left(B \circ \left(\sum_{i=1}^{n} g_i\right)^{-1}\right)(-\infty, r)
\end{aligned}$$

and hence

$$\left(\sum_{i=1}^{n} f_i\right)(A) = \left(\sum_{i=1}^{n} g_i\right)(B).$$

Since

$$\sum_{i=1}^{\infty} f_i = \lim_{n \to \infty} \sum_{i=1}^{n} f_i$$

and similarly for $\sum_{i=1}^{\infty} g_i$, let

$$f_i(A) = g_i(B), \quad f = \lim_{i \to \infty} f_i, \quad g = \lim_{i \to \infty} g_i$$

and use the relation

$$\left\{\lim_{i \to \infty} f_i < r\right\} = \bigcup_{\mu < r} \bigcap_{n=1}^{\infty} \bigcap_{i > n} \{f_i < \mu, \mu \text{ rational}\}$$

to obtain $f(A) = g(B)$. Analogous arguments for the product yield Axiom 1. Axiom 2 follows from $(f \circ g)^{-1} = (g^{-1}) \circ (f^{-1})$. For Axiom 3, let $j = \sum_{i=1}^{\infty} f_i$, where $j(x) \equiv x$ and the f_i are bounded and continuous with supports forming a locally finite system, to obtain $j(A) = j(B)$, and hence $A = B$ [since $j(A) = A$ in Mackey's system]. Finally, if c is the function $x \to c$, we have

$$\begin{aligned}
c^{-1}(E) &= R, \quad \text{if} \quad c \in E, \\
&= \varnothing, \quad \text{if} \quad c \notin E,
\end{aligned}$$

where R is the real axis. Thus $c(A)$ $(= A \circ c^{-1})$ maps E on the largest (smallest) element in $\mathfrak{L}$ if $c \in E$ $(c \notin E)$. This is independent of A; hence Axiom 3 holds.

Theorem 2: The axiom system of Segal satisfies the above axioms.

Proof: This follows from the functional representation theorem of Segal. The only abstruse part is Axiom 1, which involves infinite sums. But as the spectra of A, B are compact and the supports of f_i, g_i form locally finite sets, the series becomes finite sums

$$\left(\sum_{i=1}^{\infty} f_i\right)(A) = \sum_{k=1}^{n} f_{i_k}(A)$$

(summing in Segal's sense), so that the axiom does hold.

B. The Algebraic Structure on $\mathcal{M}$

Axiom 1 allows us to introduce an algebraic structure on each set $\mathcal{F}(A) = \{f(A) \mid f \in \mathcal{F}\}$, A fixed in $\mathcal{M}$, in the obvious way. We define

$$\sum_{i=1}^{\infty} f_i(A) \quad \text{as} \quad \left(\sum_{i=1}^{\infty} f_i\right)(A)$$

whenever the supports of the functions f_i form a locally finite system, and we define $f_1(A) f_2(A)$ as $(f_1 f_2)(A)$ for any functions $f_1, f_2 \in \mathcal{F}$. By Axiom 1 this is consistent, and $\mathcal{F}(A)$ becomes an algebra with zero element $\ominus = 0(A)$ and unity $I = 1(A)$.

It is clear from Axiom 1 that the above algebraic structure is consistent on any intersection, $\bigcap_{i=1}^{n} \mathcal{F}(A_i)$.

We shall need the following:

Proposition 1: Let $f_n(A) = \ominus$, $n = 1, 2, \cdots$, with the supports of the f_n forming a locally finite set, and let $f = \sum_{n=1}^{\infty} f_n$. Then $f(A) = \ominus$.

Proof: Take any $h \in \mathcal{F}$ and consider a partition of unity, $\sum_{n=1}^{\infty} g_n = 1$, so that $h = \sum_{n=1}^{\infty} h_n$ with $h_n = h g_n$. We then have $f_n(A) + h_n(A) = h_n(A)$ since $f_n(A) = \ominus$, i.e., $(f_n + h_n)(A) = h_n(A)$. Hence

$$\sum_{n=1}^{\infty} (f_n + h_n)(A) = \left(\sum_{n=1}^{\infty} h_n\right)(A) = h(A)$$

or

$$\left(\sum_{n=1}^{\infty} f_n\right)(A) + h(A) = h(A) \quad \text{or} \quad f(A) = \ominus.$$

Proposition 2: For any $A \in \mathcal{M}$, we have $j(A) = A$, where $j(x) \equiv x$.

Proof: Since $f \circ j = f$, we have that $f(j(A)) = f(A)$ for all f.

C. The Spectrum of an Observable

We shall write $0 \leq f \subset E$ to mean that, for all x, we have $0 \leq f(x) \leq \mathfrak{X}_E(x)$, where $\mathfrak{X}_E$ is the characteristic function (indicator) of the set E.

Definition 1: An open set U is "A-null" if, for any $f \in \mathcal{F}$ such that $0 \leq f \subset U$, we have $f(A) = \ominus$.

Proposition 3: Let U, V, U_i $(i \in I)$ be open. If $V \subseteq U$ and U is A-null, so is V. If each U_i is A-null, so is their union.

Proof: The first is obvious. Now we can write $\bigcup_{i \in I} U_i$ as a countable union $\bigcup_{n=1}^{\infty} V_n$ of a locally finite system V_n, such that each V_n is contained in some U_i because the real line is separable; then all V_n are A-null. Now take a partition of unity

$$\sum_{n=1}^{\infty} f_i = 1$$

subordinate to V_n, i.e., $0 \leq f_n \subset V_n$. If

$$0 \leq f \subset \bigcup_{n=1}^{\infty} V_n,$$

we have

$$f = \sum_{n=1}^{\infty} f f_n,$$

so that $f(A) = \ominus$ by Proposition 1.

Corollary 3.1: If $f(A) = \ominus$, then $\{x \mid f(x) \neq 0\}$ is A-null.

Proof: We may assume that $f \geq 0$, since $f^2(A) = f(A)^2 = \ominus$ while $f(x)$ and $f(x)^2$ vanish for the same values of x. Let $U_n = \{x \mid f(x) > 1/n\}$. If $0 \leq g \subset U_n$, we define $h(x)$ as $g(x)/f(x)$ if $g(x) \neq 0$ and as 0 if $g(x)$ vanishes; then h is continuous and $g = hf$. Thus $g(A) = f(A)h(A) = \ominus$ and U_n is A-null. But

$$\{x \mid f(x) \neq 0\} = \bigcup_{n=1}^{\infty} U_n$$

and thus is A-null.

Definition 2: The "spectrum" σA of $A \in \mathcal{M}$ is the complement of the union of all A-null open sets.

Thus, σA is the smallest closed set whose complement is A-null.

Proposition 4: For any $f \in \mathcal{F}$, $A \in \mathcal{M}$, we have that $\sigma f(A) \subseteq \overline{f(\sigma A)}$.

Proof: Let U be open, $U \cap f(\sigma A) = \varnothing$; then $f^{-1}(U) \cap \sigma A = \varnothing$. If $0 \leq g \subset U$, then $0 \leq g \circ f \subset f^{-1}U$; hence $(g \circ f)(A) = \ominus$ because $f^{-1}U$ is an A-null set. Thus $g(f(A)) = \ominus$ and therefore U is an $f(A)$-null set, so that $U \cap \sigma(f(A)) = \varnothing$. Taking for U the complement of the closure of $f(\sigma A)$, we obtain the result.

Proposition 5: We have $A = \ominus$ iff $\sigma A \subseteq \{0\}$.

Proof: Since $\ominus = 0(A)$, we have $\sigma\ominus \subseteq \overline{0(A)} = \{0\}$. Suppose that $\sigma A \subseteq \{0\}$. Write

$$(-\infty, 0) \cup (0, +\infty) = \bigcup_{n=1}^{\infty} V_n,$$

where the V_n form a locally finite system with compact closures, and take a partition of unity

$$\sum_{n=1}^{\infty} f_n = 1 \quad \text{on} \quad \bigcup_{n=1}^{\infty} V_n.$$

Then $f_n(A) = \ominus$, while $j = \sum_{n=1}^{\infty} j f_n$ so that

$$A = j(A) = \sum_{n=1}^{\infty} j(A) f_n(A) = \ominus.$$

Corollary 5.1: If $0 \leq f \leq g$ and $g(A) = \ominus$, then $f(A) = \ominus$.

Proof: If $g(A) = \ominus$, then $\{x \mid g(x) \neq 0\}$ is A-null; but $\{x \mid f(x) \neq 0\} \subseteq \{x \mid g(x) \neq 0\}$, and hence is also A-null. Thus $f \mid \sigma A = 0$ or $f(\sigma A) = \{0\}$ and hence $\sigma f(A) \subseteq \{0\}$ or $f(A) = \ominus$.

Proposition 6: For any $f \in \mathcal{F}$ and $A \in \mathcal{M}$, we have $\overline{f(\sigma A)} \subseteq \sigma f(A)$.

Proof: Take $x \in \overline{f(\sigma A)}$; then, for any open interval U containing x, we have $U \cap f(\sigma A) \neq \varnothing$ or $(f^{-1}U) \cap \sigma A \neq \varnothing$. Thus there exists an $h \in \mathcal{F}$ with compact support such that $0 \leq h \subset f^{-1}U$ and $h(A) \neq 0$. Because there exists some $h \in \mathcal{F}$ with this property, and by Proposition 1, we may assume h to vanish on all but a single component of $f^{-1}U$, say W. Now write W as a union of a locally finite countable covering of open bounded sets, and let h_n be a partition of unity subordinate to it. Then some $h_n(A)h(A)$ is not $\ominus$, and we may take it to be h. By Theorem 25 of Appendix C, there exists a $g \in \mathcal{F}$ such that $0 \leq g \subset U$ and $h \leq g \circ f$. Then $g(f(A)) \neq \ominus$ and therefore U is not $f(A)$-null; hence, $U \cap \sigma f(A) \neq \varnothing$ and, as $\sigma f(A)$ is closed, we have $x \in \sigma f(A)$.

We restate Propositions 4 and 6 in the following way.

Theorem 3: For any $f \in \mathcal{F}$, $A \in \mathcal{M}$, we have $\sigma f(A) = \overline{f(\sigma A)}$.

We have also established in the previous arguments the following:

Theorem 4: For any $f \in \mathcal{F}$, $A \in \mathcal{M}$, the observable $f(A)$ is completely determined by the function $f \mid \sigma A$. The map $f(A) \to f \mid \sigma A$ is an algebraic isomorphism.

Definition 3: An observable is "bounded" if its spectrum is bounded, i.e., compact.

We shall write $\mathcal{B}(A)$ for the set of bounded elements in $\mathcal{F}(A)$, and $\|A\|$ for $\sup_{x \in \sigma A} |x|$.

Proposition 7: If $B = f(A)$, then $\|B\| = \sup_{x \in \sigma A} |f(x)|$.

Proof: We have

$$\|B\| = \sup_{y \in \sigma B} |y| = \sup_{y \in \sigma f(A)} |y| = \sup_{y \in f(\sigma A)} |y| = \sup_{y \in f(\sigma A)} |y| = \sup_{x \in \sigma A} |f(x)|.$$

The next statement needs no proof.

Proposition 8: The set $\mathcal{B}(A)$ is a subalgebra of $\mathcal{F}(A)$ complete under the norm $B \to \|B\|$ and closed under the action of $\mathcal{F}$.

D. The Cone of Positive Elements

Definition 4: An observable is "positive" if it is the square of some other observable.

We use $\mathcal{C}(A)$ to denote the set of all positive observables in $\mathcal{F}(A)$.

Proposition 9: The observable $f(A) \in \mathcal{C}(A)$ iff $f \mid \sigma A \geq 0$.

Proof: Let $f(A) = B^2$; then $\sigma f(A) = \sigma(B^2) = \overline{f^2(B)}$, hence $\sigma f(A) \subseteq [0, +\infty)$, and thus $f(\sigma A) \subseteq [0, +\infty)$. Conversely, let $f \mid \sigma A \geq 0$ and consider any h on the reals extending $f \mid \sigma A$ continuously with nonnegative values (Tietze's theorem). Then $f \mid \sigma A = (h^{\frac{1}{2}} \mid \sigma A)^2$ and $f(A) = (h^{\frac{1}{2}}(A))^2$ so that $f(A)$ is a square.

Theorem 5: The set $\mathcal{C}(A) \cap \mathcal{B}(A)$ is a proper cone in $\mathcal{B}(A)$, radial at I; the resulting ordering is Archimedean and I is a unit element.

Proof: Clearly, $f(A)$, $g(A) \in \mathcal{C}(A)$ implies $f(A) + g(A) \in \mathcal{C}(A)$ because of Proposition 9; similarly, $\lambda f(A) \in \mathcal{C}(A)$ for $\lambda \geq 0$. Thus $\mathcal{C}(A)$ is indeed a cone. Now let B and $-B$ be in $\mathcal{C}(A)$, $B = f(A)$, $-B = g(A)$, so that $f(A) + g(A) = \ominus$. Then $f(x) + g(x) = 0$ for all $x \in \sigma A$ and, as $f(x)$, $g(x)$ are nonnegative on σA, we obtain $f \mid \sigma A = g \mid \sigma A = 0$, i.e., $f(A) = g(A) = \ominus$, and $\mathcal{C}(A)$ is proper. Now take any $f(A) \in \mathcal{B}(A)$ and let k be a lower bound of f on σA. Then, for $0 \leq t \leq |1 - k|^{-1}$, we have $(1 - t) + tf(x) \geq 0$ on σA and, thus, $(1 - t)I + tf(A) \in \mathcal{C}(A)$; i.e., $\mathcal{C}(A)$ is radial at I. To show that the ordering is Archimedean, suppose that $-\epsilon I \leq f(A) \leq \epsilon I$ for all $\epsilon > 0$; then $|f(x)| \leq \epsilon$ on σA for all $\epsilon > 0$ and therefore $f(A) = \ominus$. Finally, for any $f(A) \in \mathcal{B}(A)$, we have $-\|f(A)\| I \leq f(A) \leq \|f(A)\| I$, so that I is a unit.

E. States

Let $\mathcal{B} = \bigcup_{A \in \mathcal{M}} \mathcal{B}(A)$.

Definition 5: A "state" is a map m from $\mathcal{B}$ into the reals such that $m \mid \mathcal{B}(A)$ is linear for any $A \in \mathcal{M}$, $m(A^2) \geq 0$ for any bounded A, and $m(I) = 1$.

Write $\mathcal{S}$ for the set of all states.

Proposition 10: If A is bounded, then $|m(A)| \leq \|A\|$.

Proof: Let $f(x) = \|A\| - x$ and $B = f(A)$; since $f \mid \sigma A \geq 0$, we have $g(A) = f(A)$, where $g(x) = \max(0, f(x))$. As $g \geq 0$, the function $h = g^{\frac{1}{2}}$ is in $\mathcal{F}$ and thus $B = h(A)^2$. Therefore $m(B) \geq 0$; hence $m(A) \leq \|A\|$. Replace A by $-A$ to obtain $m(A) \geq -\|A\|$.

Theorem 6: Let m be a state and $\mathcal{F}_0$ the set of all bounded elements of $\mathcal{F}$. Then the map $m_A : \mathcal{F}_0 \to R$, given by $m_A(f) = m(f(A))$, determines uniquely a regular finitely additive probability measure μ_A on the ring of sets generated by the open sets in R such that, for any bounded $f(A)$, we have

$$m(f(A)) = \int f(x)\, d\mu_A(x).$$

This measure is supported by σA.

Proof: For $f \in \mathcal{F}_0$, we have

$$|m_A(f)| = |m(f(A))| \leq \|f(A)\| = \sup_{x \in \sigma A} |f(x)| \leq \|f\|_\infty;$$

therefore the measure μ_A exists and is uniquely determined. Also, $\mu_A(R) = m(1(A)) = m(I) = 1$ and, for $f \geq 0$, we have $\mu_A(f) \geq 0$, since

$$f = g^2 \quad \text{and} \quad m(g(A)^2) \geq 0.$$

Finally, if the support of f is disjoint from σA, we have $f(A) = \ominus$ and, hence, $m_A(f) = 0$, so that the support of μ_A is in σA.

Definition 6: The probability measure μ_A is the "distribution of A in the state m."

Note: Thus $m(A)$ is the expectation value of A in the state m, the points of σA being the various probable values of A. It is possible that for an unbounded A the expectation exists in the sense of the integral

$$\int_{\sigma A} x \, d\mu_A(x).$$

The interpretation of μ_A as a probability distribution is justified because of the following:

Theorem 7: For any set E in the ring generated by the open sets, we have $\mu_{f(A)}(E) = \mu_A(f^{-1}E)$.

Proof: For any bounded $g \in \mathcal{F}$, we set $B = f(A)$ and obtain $m(g(B)) = m((g \circ f)(A))$, so that

$$\int g \, d\mu_B = \int (g \circ f) \, d\mu_A .$$

But this is equivalent to the above statement.

Observe that the value of $\mu_A(U)$ (for open U) is given by $\sup \{m(f(A)) \mid 0 \leq f \subset U\}$.

Remark: If we restrict attention to bounded observables only, then we see that automatically the measures associated to the states are countably additive because the spectra are now compact, and we obtain from Dini's theorem the countable additivity condition: If $f_1 \geq f_2 \geq \cdots$ and $\inf_n f_n(x) = 0$ for each x, then

$$\lim_{n \to \infty} \mu_A(f_n) = 0.$$

To study further the structure of $\mathcal{S}$, we introduce the set $\mathcal{C}$ of all maps $s: \mathcal{B} \to R$ such that $s \mid \mathcal{B}(A)$ is linear for each A and $|s(B)| \leq k \|B\|$ for all $B \in \mathcal{B}$ (k may depend upon s). Clearly $\mathcal{C}$ is a real vector space under pointwise operations, and

$$\|s\| = \sup_{B \in \mathcal{B}} \frac{|s(B)|}{\|B\|}$$

defines a norm on it.

Lemma 1: The set $\mathcal{S}$ is a strongly convex subset of the unit ball in $\mathcal{C}$.

Proof: If $m_i \in \mathcal{S}$ and $a_i \geq 0$ with $\sum_{i=1}^{\infty} a_i = 1$, then $m = \sum_{i=1}^{\infty} a_i m_i$ (pointwise operations) is evidently in $\mathcal{S}$. Further, $\|m\| \leq 1$ by Proposition 10.

The weak topology on $\mathcal{C}$ is obtained by introducing the basis of neighborhoods

$$N(s_0, \epsilon; A_1, A_2, \cdots, A_n) = \{s \mid |s(A_i) - s_0(A_i)| < \epsilon, \quad i = 1, \cdots, n\}$$

for the element s_0.

Proposition 11: The space $\mathcal{C}$, equipped with the weak topology, is locally convex Hausdorff. The unit ball is weakly compact, $\mathcal{S}$ is weakly closed in it and hence weakly compact.

The proof is standard and we omit it. By the Krein–Millman theorem we have the structure of $\mathcal{S}$.

Definition 7: A state is "pure" if it is an extreme point of $\mathcal{S}$.

Thus the states are obtained from the pure ones by mixing (convex combinations) and approximating (weak limits).

F. Existence of States

We shall now obtain a necessary and sufficient condition for the existence of states with desired properties. First, observe that, for a fixed $A \in \mathcal{M}$, the range of the map $m \to m(A)$ is contained in the interval $[\inf \sigma A, \sup \sigma A]$. Since the map is continuous in the weak topology on $\mathcal{S}$ for which $\mathcal{S}$ is compact, we see that the set $\{m(A) \mid m \in \mathcal{S}\}$ is a closed subinterval of $[\inf \sigma A, \sup \sigma A]$ (or perhaps empty).

Consider the following:

Axiom 4a: For any $a \in [\inf \sigma A, \sup \sigma A]$ there exists a state m such that $m(A) = a$.

First, we give an alternative formulation:

Axiom 4b: For any open non-A-null set U, there exists a state m such that $\mu_A(U) = 1$.

To prove the equivalence, we need the following:

Proposition 12: If Axiom 4b is valid, then $\|A\| = \sup_{m \in \mathcal{S}} |m(A)|$.

Proof: Since $|m(A)| \leq \|A\|$, we only need the reverse relation. Take $x_0 \in \sigma A$, $\epsilon > 0$, and let $U = (x_0 - \epsilon, x_0 + \epsilon)$; then we have a state m such that $\mu_A(U) = 1$.

Now,

$$|m(A)| = \left| \int x \, d\mu_A(x) \right|$$

$$= \left| \int_U x \, d\mu_A(x) \right| \geq |x_0 - \epsilon| \geq |x_0| - \epsilon.$$

Thus,

$$|x_0| \leq \epsilon + \sup_{m \in \mathcal{S}} |m(A)|, \quad \text{for any} \quad \epsilon > 0,$$

which proves the result.

Theorem 8: Axioms 4a and 4b are equivalent.

Proof: Assume Axiom 4a and take $x_0 \in (\sigma A) \cap U$. There exists an $f \in \mathcal{F}$ such that $0 \leq f \subset U$ and $f(x_0) = 1$. Then $1 = \sup f(\sigma A) = \sup \sigma f(A)$ and, hence, there is a state m such that $m(f(A)) = 1$. But then $\mu_A(U) = 1$. Now assume Axiom 4b, and let $[a, b]$ be the range of $m \to m(A)$. By considering $A - aI$, we may assume $a = 0$ without loss of generality, so that $m(A) \geq 0$ for any state m. Now, if σA contains both negative and positive points, there exists by Theorem 26 of Appendix C a function $f \in \mathcal{F}$ such that (i) $-1 \leq f(x) \leq 0$ and the value -1 is taken on σA, and (ii) $f(x) + 1 - d \geq cx$ for some $c, d > 0$. Then, if $B = f(A)$, we obtain $-B$ positive and $B + (1 - d)I - cA$ also positive; since $m(A) \geq 0$, we obtain

$$-1 + d \leq m(B) \leq 0$$

for all states m. By Proposition 12, we have $\|B\| < 1$, while by Proposition 7, we have $\|B\| = 1$. Thus we conclude that σA is nonnegative. But then, again by Proposition 12, we see that $\sup \sigma A = b$ and, thus, the two intervals $[a, b]$ and $[\inf \sigma A, \sup \sigma A]$ coincide. Axiom 4a follows.

Observe that Axiom 4a is equivalent to the apparently weaker statement: If $A \neq \ominus$, then there exists a state m such that $m(A) = \sup \sigma A$.

We shall now translate Axiom 4 into a form which involves only the algebraic structure of the sets $\mathfrak{B}(A)$ where $A \in \mathcal{M}$. We observe that as vector spaces they form a coherent system (see Appendix A), and we can apply the ideas developed there. The cones $\mathcal{C}(A) \cap \mathfrak{B}(A)$ are all radial at I, and it is obvious that, if $f(A) \notin \mathcal{C}(A)$, then there is a point $x_0 \in \sigma A$ at which $f(x_0) < 0$; thus the map $g(A) \to g(x_0)$ defines a positive functional on $\mathfrak{B}(A)$ for which $f(A)$ has a negative value. Thus the hypotheses of Theorem 24 in Appendix A are satisfied.

Theorem 9: Axiom 4 is equivalent to the statement: Every element in the sum $\sum_A (\mathcal{C}(A) \cap \mathfrak{B}(A))$ that is in $\mathfrak{B}(B)$ is already in $\mathcal{C}(B)$.

Proof: If this statement is valid, then the positive functionals on $\mathfrak{B}(B)$ extend. Now take any $a \in [\inf \sigma B, \sup \sigma B]$ and define a functional φ on the subspace $\mathcal{N}$ = multiples of I by $\varphi(cI) = c$; this is positive. We can extend φ to the subspace spanned by I and B, provided we assign to B a value between

$$\sup \{\varphi(A) \mid A \in \mathcal{N}, A \leq B\}$$

and

$$\inf \{\varphi(A) \mid A \in \mathcal{N}, B < A\}$$

[the ordering being the one induced by the cone $\mathcal{C}(B)$], according to Appendix B. But, since $A \in \mathcal{N}$ means $A = cI$, the relations $A \leq B$ and $B \leq A$ mean $c \leq \inf \sigma B$ and $\sup \sigma B \leq c$; thus, φ can be extended to a positive functional on the subspace of $\mathfrak{B}(B)$ spanned by I and B, provided we assign to B a value between $\inf \sigma B$ and $\sup \sigma B$. Call this extension ψ. Now ψ is positive on a subspace of $\mathfrak{B}(B)$ which contains the point I, at which the cone $\mathcal{C}(B)$ is radial; hence applying Theorem 3.3 of Ref. 6, we obtain an extension of ψ to $\mathfrak{B}(B)$. But then our hypothesis allows an extension to $\sum_A \mathfrak{B}(A)$, i.e., we obtain a state m such that $m(B) = a$. Now for the converse: Assume Axiom 4 and let B be an element in $\sum_A (\mathcal{C}(A) \cap \mathfrak{B}(A))$; this implies that $m(B) \geq 0$ for any state m. Now, we show that $\sigma B \geq 0$. If not, we find, as in the proof of Theorem 8, an element $B_1 \in \mathfrak{B}(B)$ such that $-1 + d \leq m(B_1) \leq 0$ for any state m, while -1 is in the spectrum of B_1. This contradicts Axiom 4, and, hence, the spectrum of B is nonnegative.

In the following, we assume Axiom 4.

Theorem 10: If $x_0 \in \sigma A$, there exists a pure state m at which A is measured exactly (i.e., its probability distribution has zero variance) such that $m(A) = x_0$.

Proof: The map $f(A) \to f(x_0)$ is positive on $\mathfrak{B}(A)$; hence, it extends to a state. Let $\mathcal{S}_0$ be the set of all its extensions. Clearly, $\mathcal{S}_0$ is a weakly closed subset of $\mathcal{S}$ and, thus, is weakly compact; therefore it has an extreme point m. We shall show that m is an extreme point of $\mathcal{S}$. Let $m = a_1 m_1 + a_2 m_2$, with $a_i \geq 0$ and $a_1 + a_2 = 1$; the restrictions of m_1, m_2 to $\mathfrak{B}(A)$ determine measures which must coincide with μ_A because this is evidently concentrated on $\{x_0\}$. Therefore $m_1, m_2 \in \mathcal{S}_0$ and, as m is extreme in $\mathcal{S}_0$, we have $m = m_1 = m_2$. Thus m is a pure state, and evidently A is measured exactly in it, since μ_A is concentrated on $\{x_0\}$.

[6] J. L. Kelley and I. Namioka, *Linear Topological Spaces* (D. Van Nostrand, Inc., New York, 1963).

II. SPECIAL SYSTEMS

A. Algebraic Structure

We now assume that $\mathcal{F}$ consists of all finite Borel-measurable functions from the reals to the reals and that their action on $\mathcal{M}$ satisfies Axioms 2, 3, plus the stronger version of Axiom 1 stated below. Observe that Theorem 1 is still valid.

Axiom 1': Let $f_i(A) = g_i(B)$, $i = 1, 2, \cdots$, and assume that

$$f = \sum_{i=1}^{\infty} f_i, \quad g = \sum_{i=1}^{\infty} g_i \in \mathcal{F}.$$

Then $f(A) = g(B)$, and $(f_1 f_2)(A) = (g_1 g_2)(B)$.

We aim at establishing an analogous theory and showing its exact relations to the Mackey system of axioms. Not all of the previous arguments are valid, because in several we made essential use of the continuity of the functions in $\mathcal{F}$. Whenever we omit a proof, it is because only trivial (if any) changes are necessary to adapt the previous argument.

First, observe that, as in Sec. I.B, we obtain on each $\mathcal{F}(A)$ a homomorphism of the algebraic structure of $\mathcal{F}$.

Proposition 13: If $f_n(A) = \ominus$ and $f = \sum_{n=1}^{\infty} f_n \in \mathcal{F}$, then $f(A) = \ominus$.

Proposition 14: For any A, $j(A) = A$.

Proposition 15: If $0 \leq f \leq g$ and $g(A) = \ominus$, then $f(A) = \ominus$.

Proof: We can now directly define $h(x)$ as $f(x)/g(x)$ for $g(x) \neq 0$ and $h(x) = 0$ for $g(x) = 0$, to obtain $h \in \mathcal{F}$ and $f = gh$; then $f(A) = g(A)h(A) = \ominus$.

The following result is very useful:

Theorem 11: An open set is A-null (according to Definition 1) iff $\mathfrak{X}_U(A) = 0$.

Proof: If $\mathfrak{X}_U(A) = 0$ and $0 \leq f \subset U$, then by Proposition 15 we obtain $f(A) = \ominus$ and thus U is A-null. Conversely, let U be A-null and consider a partition of unity $\sum_{n=1}^{\infty} f_n = 1$ on U (with all f_n zero outside U). Then $f_n(A) = \ominus$ and, as $\mathfrak{X}_U = \sum_{n=1}^{\infty} f_n$, we have by Proposition 13 that $\mathfrak{X}_U(A) = 0$.

Thus we can extend Definition 1 to the following:

Definition 8: A Borel set E is "A-null" if $\mathfrak{X}_E(A) = 0$.

By using Propositions 13 and 15, we have the following by standard arguments:

Theorem 12: The set of all A-null sets is a σ-ideal in the σ-ring of all Borel sets.

We can now prove the following:

Theorem 13: The observable $f(A) = \ominus$ iff $\{x \mid f(x) \neq 0\}$ is A-null.

Proof: As before, we may assume $f \geq 0$. Suppose $f(A) = \ominus$ and let

$$E_n = \{x \mid (n+1)^{-1} \leq f(x) < n^{-1}\}$$

so that $\mathfrak{X}_{E_n} \leq (n+1)f$; but then E_n is A-null and, as

$$\{x \mid f(x) \neq 0\} = \bigcup_{n=1}^{\infty} E_n,$$

we see that it is A-null. Conversely, if

$$E = \{x \mid f(x) \neq 0\}$$

is A-null, we obtain from the relation $f = f\mathfrak{X}_E$ that $f(A) = \ominus$.

B. Spectrum, Norm, and States

Retaining the definition of the spectrum, we then have the following:

Proposition 16: For any $A \in \mathcal{M}$, $\sigma A \subseteq \{0\}$ iff $A = \ominus$.

Proof: By Theorem 13, we have $A = \ominus$ iff $R - \{0\}$ is A-null; but this is open, and its being A-null means $\sigma A \subseteq \{0\}$.

We can now establish the analog of Theorem 4:

Theorem 14: For any $f, g \in \mathcal{F}$ and $A \in \mathcal{M}$, we have $f(A) = g(A)$ iff $f \mid \sigma A = g \mid \sigma A$ except on some A-null set.

Proof: We need only establish that $f(A) = \ominus$ iff $f \mid \sigma A = 0$ except on some A-null set. If $f \mid \sigma A = 0$ except on the A-null set E, then $f\mathfrak{X}_{\sigma A} = f\mathfrak{X}_{\sigma A}\mathfrak{X}_E$ and, since $\mathfrak{X}_{\sigma A}(A) = I$, we obtain $f(A) = f(A)\mathfrak{X}_E(A) = \ominus$. Conversely, if $f(A) = \ominus$, we have by Theorem 13 that f vanishes except on some A-null set and hence $f \mid \sigma A$ likewise.

It is unfortunate that Theorem 3 is not valid in the present context, because proposition 6 may fail if f is not continuous. However we have:

Theorem 15: For any $f \in \mathcal{F}$ and $A \in \mathcal{M}$, we have $\sigma f(A) \subseteq \overline{f(\sigma A)}$. If f is continuous, then equality holds.

Proof (after Gudder[7]): Consider any $x \in \sigma f(A)$ and any open U containing x; then $\mathfrak{X}_U(f(A)) \neq \ominus$ and hence $\mathfrak{X}_{f^{-1}(U)} \neq \ominus$ or $f^{-1}(U) \neq \varnothing$. This means that x is in the closure of the range of f, which implies that $\sigma f(A)$ is contained in the closure of the range of f. Now take $x \in f(\sigma A)$ and let

$$g(y) = f(y), \quad \text{for} \quad y \in \sigma A,$$
$$= x, \quad \text{for} \quad y \notin \sigma A,$$

so that $g \in \mathcal{F}$ and the range of g is contained in $f(\sigma A)$. Since $f \mid \sigma A = g \mid \sigma A$, we have $f(A) = g(A)$; thus the spectrum of $f(A)$ is contained in the closure of the range of g, hence in the closure of $f(\sigma A)$.

We also retain the definition of a bounded observable and of the norm. We shall now establish a formula for $\|A\|$ analogous to that of Proposition 7. The essential supremum of a function is understood relative to the σ-ideal of all A-null sets; we shall write it as A-essup.

Lemma 2: If $|f(x)| \leq M$ for all $x \in \sigma A$, then $\|f(A)\| \leq M$.

Proof: We have

$$\|f(A)\| = \sup_{y \in \sigma f(A)} |y| \leq \sup_{y \in \overline{f(\sigma A)}} |y| \leq \sup_{|y| \leq M} |y| = M.$$

Thus we obtain:

Lemma 3: If $\lim_{n \to \infty} f_n = f$ uniformly on σA, then

$$\lim_{n \to \infty} \|f_n(A) - f(A)\| = 0.$$

Theorem 16: Let $f \in \mathcal{F}$ and $A \in \mathcal{M}$. Then, $f(A)$ is bounded iff the function f is essentially bounded on σA; in such a case,

$$\|f(A)\| = A\text{-essup}_{x \in \sigma A} |f(x)|.$$

Proof: First, take f of the form $\sum_{n=1}^{N} a_n \mathfrak{X}_{E_n}$, with E_n pairwise disjoint, and let $B = f(A)$. Then $\sigma B \subseteq \{0, a_1, a_2, \cdots, a_N\}$ by Theorem 15. It is clear that $a_n \notin \sigma B$ iff E_n is A-null, because, for any small enough neighborhood U of a_n, we have $f^{-1}U = E_n$. Thus

$$\sup_{y \in \sigma B} |y| = \max \{|a_n| \mid E_n \text{ not } A\text{-null}\} = A\text{-essup}_{x \in \sigma A} |f(x)|.$$

[7] S. P. Gudder, Trans. Am. Math. Soc. **119**, 428 (1965).

Now we proceed to general f, and by the usual argument we assume $f \geq 0$. Take a sequence of f_n having the previous form such that $0 \leq f_n(x) \leq f(x)$, which converges uniformly and increasingly to $f(x)$. Then A-essup f_n converges to A-essup f, and $\|f_n(A)\|$ converges by the lemma to $\|f(A)\|$. Therefore we have the desired equality.

The results of Sec. I.D are carried over to the present context with trivial changes in the arguments.

We shall also retain the definition of a state. Theorem 6 holds, but now the ring of sets on which the probability measures are defined consists of all Borel sets. In fact, we have directly that $\mu_A(E) = m(\mathfrak{X}_E(A))$.

Axiom 4b takes the following form:

Axiom 4c: If E is not A-null, then there exists a state m such that $\mu_A(E) = 1$.

We shall omit the minor modifications necessary to adapt Sec. I.F.

C. Simple Observables

We shall now reduce the study of $\mathcal{M}$ to the study of a proper subset.

Definition 9: An observable is "simple" if it equals its square.

We shall write $\mathfrak{L}$ for the set of all simple observables, and $\mathfrak{L}(A)$ for the set $\mathfrak{L} \cap \mathcal{F}(A)$.

Proposition 17 (Mackey): The three relations $A = A^2$, $\sigma A \subseteq \{0, 1\}$, and $A = \mathfrak{X}_E(B)$, for some $B \in \mathcal{M}$ and some Borel set E, are equivalent.

Proof: From $A = A^2$ we obtain $(j - j^2)(A) = \ominus$, or that the spectrum of $(j - j^2)(A)$ is contained in $\{0\}$. Since $j - j^2$ is continuous, we see that it maps σA into $\{0\}$. But only the set $\{0, 1\}$ has this property; hence $\sigma A \subseteq \{0, 1\}$. Conversely, $\sigma A \subseteq \{0, 1\}$ implies $j^2 \mid \sigma A = j \mid \sigma A$ or $j^2(A) = j(A)$, i.e., $A^2 = A$. Now let $A = f(B)$, so that $f^2 - f$ is zero on σB except on some B-null set F; then

$$f(x)\mathfrak{X}_{F'}(x) = [f(x)\mathfrak{X}_{F'}(x)]^2$$

(where F' is the complement of F); hence $f(x)\mathfrak{X}_{F'}(x) = \mathfrak{X}_E(x)$ for some Borel set E. Since $\mathfrak{X}_F(A) = \ominus$, we have $f(B) = \mathfrak{X}_E(B)$. The converse is obvious.

Note: Observe that $\mathfrak{X}_E(B) = \mathfrak{X}_F(B)$ iff the sets E, F differ by a B-null set.

The connection between an observable A and simple observables is contained in the following:

Theorem 17: If $\mathfrak{X}_E(A) = \mathfrak{X}_E(B)$ for all Borel sets E, then $A = B$.

Proof: Clearly, we have $f(A) = f(B)$ for f of the form

$$\sum_{k=1}^{n} a_k \mathfrak{X}_{E_k},$$

where the E_k are pairwise disjoint. Now for a bounded f, we take a series of functions f_n of the above form converging to f. Then

$$f(A) = \sum_{n=1}^{\infty} f_n(A) = \sum_{n=1}^{\infty} f_n(B) = f(B)$$

and hence $A = B$.

Thus A is characterized by the map $E \to \mathfrak{X}_E(A)$ from the Borel sets into $\mathfrak{L}$. Observe that each $\mathfrak{L}(A)$ carries a natural structure of Boolean σ-algebra induced by the partial order: $B \leq C$ iff $B = \mathfrak{X}_E(A)$ and $C = \mathfrak{X}_F(A)$ imply that $E \subseteq F$ except on some A-null set (or equivalently, in view of Axiom 4c, $m(B) \leq m(C)$ for all states m).

Theorem 18: The map $E \to \mathfrak{X}_E(A)$ is a σ-homomorphism into $\mathfrak{L}(A)$. Conversely, if $E \to Q(E)$ is such a σ-homomorphism, there exists a unique $B \in \mathfrak{F}(A)$ for which $Q(E) = \mathfrak{X}_E(B)$.

Proof: The only thing to show for the first part is that $\mathfrak{X}_E(A)$ is the supremum of the elements $\mathfrak{X}_{E_n}(A)$ in $\mathfrak{L}(A)$ under the assumptions that

$$E = \bigcup_{n=1}^{\infty} E_n$$

and the E_n are pairwise-disjoint modulo the A-null sets. But this is clear because the unions are countable, and the A-null sets form a σ-ideal. For the converse, we write $Q(E) = \mathfrak{X}_{S(E)}(A)$, where $S(E)$ is unique modulo for some A-null set. All equalities of sets in the following will be understood modulo some A-null set. We shall show that $E \to S(E)$ is a σ-homomorphism (modulo A-null sets). Let $E \cap F = \varnothing$. Then, $Q(E) + Q(F) = Q(E \cup F)$ and, hence, by squaring, we have

$$Q(E) + Q(F) + 2Q(E)Q(F) = Q(E \cup F)$$

or $Q(E)Q(F) = \ominus$. Hence, $\mathfrak{X}_{S(E) \cap S(F)}(A) = \ominus$ and, thus, $S(E) \cap S(F) = \varnothing$. Furthermore, we obtain

$$\mathfrak{X}_{S(E)}(A) + \mathfrak{X}_{S(F)}(A) = \mathfrak{X}_{S(E) \cup S(F)}(A),$$

since $S(E) \cap S(F) = \varnothing$, and hence $\mathfrak{X}_{S(E \cup F)}(A) = \mathfrak{X}_{S(E) \cup S(F)}(A)$, or $S(E \cup F) = S(E) \cup S(F)$. We now show that in general $S(E \cap F) = S(E) \cap S(F)$. We have $S(E \cap F) \cup S(E \cap F') = S(E)$ and, as

$$S(E \cap F') \cap S(F) = \varnothing,$$

we get $S(E \cap F) \cap S(F) = S(E) \cap S(F)$; but

$$S(E \cap F) \subseteq S(F)$$

and thus $S(E \cap F) = S(E) \cap S(F)$. Finally, we show complete additivity: Assume

$$E_i \cap E_j = \varnothing, \quad \text{for} \quad i \neq j,$$

and compute. We have

$$\begin{aligned} \mathfrak{X}_{S(\cup E_n)}(A) &= Q\left(\bigcup_{n=1}^{\infty} E_n\right) = \sup_n Q(E_n) \\ &= \sup_n \mathfrak{X}_{S(E_n)}(A) = \mathfrak{X}_{\cup S(E_n)}(A); \end{aligned}$$

hence

$$S\left(\bigcup_{n=1}^{\infty} E_n\right) = \bigcup_{n=1}^{\infty} S(E_n).$$

Now, there exists a Borel function f such that $S(E) = f^{-1}(E)$, which will be uniquely determined modulo some A-null set. Then

$$Q(E) = \mathfrak{X}_{S(E)}(A) = \mathfrak{X}_{f^{-1}(E)}(A) = \mathfrak{X}_E(f(A)).$$

Since the ambiguity on f is restricted to an A-null set, the element $f(A)$ is uniquely determined.

We recapitulate below the results we have obtained this far.

A set $\mathfrak{L}$ is given as the union of Boolean σ-algebras $\mathfrak{L}_i$ (i varying in some index set); each $\mathfrak{L}_i$ is isomorphic to a quotient of the algebra of Borel sets on the real line by some σ-ideal; the minimum and maximum elements of the various $\mathfrak{L}_i$ are common to all, and the system of the $\mathfrak{L}_i$ is "coherent" in the sense that $\mathfrak{L}_i \cap \mathfrak{L}_j$ is a Boolean σ-subalgebra of $\mathfrak{L}_i$ and $\mathfrak{L}_j$, while on it the inherited structures coincide. An observable is obtained as a σ-homomorphism of the Borel sets into some member of the system.

Apropos, we remark that, starting with such a coherent system of Boolean σ-algebras, we can define observables as σ-homomorphisms of the Borel sets into some member of the system and the action of the Borel function f on the observable A to be $A \circ f^{-1}$ (composition); then Axioms 1, 2, and 3 hold, the proof being the same as in Theorem 1. The original set can then be recaptured as an isomorph of the set of simple observables.

With regard to states, we have the following characterization:

Theorem 19: Each state is uniquely determined by its restriction to the set of simple observables. If p_m is the restriction of m to $\mathfrak{L}$ and if μ_m is the additive measure $E \to p_m(\mathfrak{X}_E(A))$, then $m(f(A)) = \int f\, d\mu_m$. A map $m: \mathfrak{L} \to R$ is the restriction of a state iff $m \mid \mathfrak{L}(A)$ is an additive probability measure.

Proof: Only the second part needs a proof, as the first is contained in Theorem 6. Given the map m, observe that, if U is an open set disjoint from σA, then we have that $\mu_m(U) = m(\mathfrak{X}_U(A)) = 0$, so that, for a bounded $f(A)$, the integral $\int f(x)\, d\mu_m(x)$ exists and equals

$$\int_{\sigma A} f(x)\, d\mu_m(x).$$

To extend the functional thus defined to any bounded observable, we need the consistency relation: $B = f(A)$ implies

$$\int_{\sigma B} x\, d\nu(x) = \int_{\sigma A} f(x)\, d\mu(x),$$

where

$$\nu(E) = m(\mathfrak{X}_E(B)) \quad \text{and} \quad \mu(F) = m(\mathfrak{X}_F(A));$$

but

$$\nu(E) = m(\mathfrak{X}_{f^{-1}E}(A)) = \mu(f^{-1}E)$$

and, as ν, μ are concentrated on σB and σA, respectively, we have the desired equality of integrals. Thus m extends consistently to $\mathfrak{B}$. Linearity on each $\mathfrak{B}(A)$ as well as the remaining properties of states follows immediately from the properties of m.

This allows us to define, in the framework of simple observables, a state as a map from $\mathfrak{L}$ to R such that its restriction to each $\mathfrak{L}_i$ is an additive probability measure. Axiom 4c then takes the form:

Axiom 4d: If $A \in \bigcup \mathfrak{L}_i$ is not $\ominus$, there exists a state m such that $m(A) = 1$.

D. Relation to the System of Mackey

It is now clear that Mackey's system is obtained by imposing further restrictions on the coherent system $(\mathfrak{L}_i)$.

The first is that the given partial ordering on the $\mathfrak{L}_i$ determines a partial ordering on the union $\mathfrak{L} = \bigcup \mathfrak{L}_{i,j}$ i.e., that the relation "$A, B \in \mathfrak{L}_i$ for some i and $A \leq B$" is a partial ordering on $\mathfrak{L}$. It is further assumed that every sequence of disjoint elements of $\mathfrak{L}$ has a supremum (or, equivalently, that it is contained in some $\mathfrak{L}_i$). Finally, the definition of a state is more restrictive, as it is assumed to be countably additive; this, in terms of expectation-value functionals on the bounded observables, means a condition of the following form: If $A_1 \geq A_2 \geq \cdots \geq A_n > \cdots$ and the infimum is $\ominus$ [all A_i in some $\mathfrak{B}(A)$], then $\lim m(A_n) = 0$ as $n \to \infty$. It may be interesting to find a condition on $\mathcal{M}$ equivalent to the above assumption on the partial order of $\mathfrak{L}$.

E. Segal's System

It is not hard to verify that Segal's system is obtained from Axioms 1, 2, and 3 of Sec. I.A by further imposing:

(i) The coherent system $(\mathfrak{F}(A))$ is already a vector space, i.e., the construction of $\sum \mathfrak{F}(A)$ does not produce anything outside of $\bigcup \mathfrak{F}(A)$.

(ii) All observables are bounded.

It is not necessary to assume Axiom 4 because it follows from (i) that the condition of Theorem 9 is satisfied.[8]

APPENDIX A: SUMS OF VECTOR SPACES

1. The Sum of Vector Spaces

Let $(V_i)_{i \in I}$ be a family of vector spaces over some scalar field F; we assume the zero element to be common to all (write it 0), while we write addition as $+$.

Definition 10: The family $(V_i)_{i \in I}$ is "coherent" if, for any choice of $i_1, i_2, \cdots, i_n \in I$, the set $V_{i_1} \cap V_{i_2} \cap \cdots \cap V_{i_n}$ is a vector subspace of each V_{i_k}, and on it the various induced vector-space structures coincide.

Now consider the set V of all maps $a: I \to \bigcup_{i \in I} V_i$, such that (i) $a(i) \in V_i$ and (ii) all but finitely many values are zero, and equip it with pointwise operations (i.e., consider the restricted outer direct sum of the spaces V_i). Let $a \sim b$ mean: There exist distinct $j, k \in I$ such that, for $i \neq j, k$,

$$a(i) = b(i)$$

and, for some $x \in V_j \cap V_k$,

$$a(j) + x = b(j),$$
$$a(k) - x = b(k).$$

Obviously this relation is symmetric and reflexive. Its minimal extension to an equivalence relation $a \approx b$ is defined as follows: There exists a sequence $a_0, a_1, \cdots, a_n \in V$ such that $a = a_0 \sim a_1 \sim \cdots \sim a_n = b$; call such a sequence a "chain of length n joining a to b."

[8] S. Sherman, Proc. Am. Math. Soc. **2**, 31 (1951).

Lemma 4: The relation $a \approx b$ means: $a(i) + c(i) = b(i)$, where $c(i) = 0$ except for finitely many $i = i_1, i_2, \cdots, i_n$, while

$$\begin{pmatrix} c(i_1) \\ c(i_2) \\ \cdot \\ \cdot \\ \cdot \\ c(i_n) \end{pmatrix} = M \begin{pmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{pmatrix},$$

where the matrix M has n rows and m columns ($m < n < 2m$), its entries are 1, 0, or -1, and it contains in each column exactly one 1 and one -1.

Proof: We use induction on the length of a chain joining a to b. For length 1, this is true with $M = \binom{1}{-1}$. Suppose that it is true for all chains of lengths $\leq p$, and let $a_0, a_1, \cdots, a_{p+1}$ join a to b. We then have $a(i) + c(i) = a_p(i)$ with $c(i_k) = \sum_{j=1}^{m} M_{kj} c_j$, M of the required form, and $c(i) = 0$ for $i \neq i_1, i_2, \cdots, i_n$. Since $a_p \sim b$, we have, for $i \neq r, s$,

$$a_p(i) = b(i)$$

and, for some $x \in V_r \cap V_s$,

$$a_p(r) + x = b(r),$$
$$a_p(r) - x = b(s).$$

Then $a(i) + c(i) = b(i)$, for $i \neq r, s$ and for $a(r) + c(r) + x = b(r)$, $a(s) + c(s) - x = b(s)$, so that the matrix expressing the new $c(i)$'s has one more column of the desired form and at most two more rows, depending upon whether or not r or s (or both) is equal to one of the indices $i_1, i_2, \cdots, i_n$. In any case, the element x is added at the end of the column of the c_j. Thus the new matrix M has the required form.

Remarks: (i) It is clear that, in case M contains a $2 \times k$ submatrix of nonzero elements (so that the remaining elements in the corresponding columns are zero), we can replace it by a 2×1 submatrix of the form $\binom{1}{-1}$ and change the right-hand column accordingly. (ii) Since the total number of nonzero elements in M is exactly $2m$ and the number of rows is between m and $2m$, it follows that, if no row contains exactly one nonzero element, then all contain exactly two such nonzero elements. (iii) The relation $a \approx b$ involves only the spaces V_j, for which at least one $a(j)$ or $b(j)$ is not zero because the elements c_i are in intersections of these spaces.

Lemma 5: Let $a_1 \approx b_1$ and $a_2 \approx b_2$; then $a_1 + a_2 \approx b_1 + b_2$.

Proof: First suppose that $a_1 \sim b_1$ and $a_2 \sim b_2$, so that, for some $j, k, r, s \in I$, we have, for $i \neq j, k$,

$$a_1(i) = b_1(i)$$

and, for some $x \in V_j \cap V_k$,

$$a_1(j) + x = b_1(j),$$
$$a_1(k) - x = b_1(k);$$

and, for $i \neq r, s$,

$$a_2(i) = b_2(i),$$

and, for some $y \in V_r \cap V_s$,

$$a_2(r) + y = b_2(r),$$
$$a_2(s) - y = b_2(s).$$

Case 1: All j, k, r, s are distinct. Define an element $c \in V$ by

$$c(i) = a(i),$$
$$c(j) = a(j) + x,$$
$$c(k) = a(k) - x,$$

for $i \neq j, k$, so that $c \sim a$. Now $c(j) = b(j)$ and $c(k) = b(k)$, since j and k are distinct from r and s. And

$$c(r) = a_1(r) + a_2(r) = b_1(r) + b_2(r) - y,$$
$$c(s) = a_1(s) + a_2(s) = b_1(s) + b_2(s) - y,$$

which means that $c \sim b$. Thus $a \approx b$.

Case 2: Three indices are distinct, and suppose $k = r$. Define $c, d \in V$ by the following:

for $i \neq j, k$: $c(i) = a(i)$,
$$c(j) = a(j) + x,$$
$$c(k) = a(k) - x;$$
for $i \neq k, s$: $d(i) = b(i)$,
$$d(k) = b(k) - y,$$
$$d(s) = b(s) + y.$$

Then $c \sim a$ and $d \sim b$. Now, for $i \neq j, k, s$, we have $c(i) = d(i)$; also,

$$c(j) = a(j) + x = a_1(j) + a_2(j) + x$$
$$= b_1(j) + b_2(j) - y = d(j) - y,$$

while

$$c(k) = a_1(k) + a_2(k) - x = b_1(k) + b_2(k)$$
$$= d(k) + y$$

and

$$c(s) = a_1(s) + a_2(s) = b_1(s) + b_2(s) + y = d(s).$$

Thus $c \sim d$ and $a \approx b$.

The case where both indices coincide is trivial. Now suppose

$$a_1 = a_{01} \sim a_{11} \sim \cdots \sim a_{n1} = b_1$$

and

$$a_2 = a_{02} \sim a_{12} \sim \cdots \sim a_{m2} = b_2;$$

by repeating, if necessary, we may assume that $n = m$. Then

$$a_1 + a_2 = a_{01} + a_{02} \approx a_{11} + a_{12} \approx \cdots \approx a_{n1} + a_{n2} = b_1 + b_2.$$

Theorem 20: The set $V/\approx$ of equivalence classes is a vector space under the operations inherited from V.

Proof: By the previous lemma, addition is well defined. The relation $\lambda a \approx \lambda b$ is an immediate consequence of $a \approx b$, and the rest of the argument is standard.

Notation: We write $\sum_{i\in I} V_i$ for $V/\approx$. For any $a \in V$, we write either $\sum_{i\in I} a(i)$ or A for the equivalence class containing a.

Now let $x \in V_j$, $\hat{x}$ be the map $i \to \delta_{ij}x$ and X the class of the map $\hat{x}$.

Theorem 21: The map $x \to X$ is an isomorphism of V_j into $\sum_{i\in I} V_i$ preserving the intersections of distinct V_j.

Proof: It is clear that this map preserves the operations. Let $X = 0$, i.e., $x \approx 0$. By Lemma 4, there exists a $c \in V$ such that $c(i) = \delta_{ij}x$, where all $c(i) = 0$ except for

$$i = i_1, i_2, \cdots, i_n \quad \text{and} \quad c(i_k) = \sum_{r=1}^{m} M_{kr}c_r,$$

with M of the form stated in the lemma. If there is a row of M with one nonzero element only and if this row corresponds to the index $i_k = j$, then $x = 0$ and we have finished; if $i_k \neq j$, then $c(i_k) = 0$ and we reduce m by 1. If no row has a single nonzero element, then they all have exactly two such nonzero elements. Consider a row corresponding to $i_k \neq j$: If the nonzero elements are in the columns r, s, then $c_r = \pm c_s$, and we can replace c_r in the other equations by $\pm c_s$, thereby reducing the number of equations while retaining the form of M. We are thus led to the case $m = 1$, i.e., $x \sim 0$; we then have $x + y = 0$ and $-y = 0$ for some y; hence $x = 0$. This means that the map $x \to X$ is indeed an isomorphism. Finally, let $x \approx y$ and assume that x, y are not in the same V_j, since this case is already covered. Then the same argument as above applies since the column of the $c(i_k)$ contains all zeros except x and y. Thus we check the case $x \sim y$. This means that, for some $z \in V_j \cap V_k$,

$$\delta_{ir}x = \delta_{is}y,$$
$$\delta_{jr}x + z = \delta_{js}y,$$
$$\delta_{kr}x - z = \delta_{ks}y.$$

This gives immediately $x = y$.

Notation: We shall identify V_j with its image under the above map, or, equivalently, call this image V_j.

2. Linear Functionals

Definition 11: A family of maps $f_i: V_i \to W$ (any set) is "coherent" if $f_j \mid V_j \cap V_k = f_k \mid V_j \cap V_k$, for any $j, k \in I$.

Lemma 6: There is a one-to-one correspondence between coherent families $(f_i)_{i\in I}$ of maps into W and maps $f: \bigcup_{i\in I} V_i \to W$ determined by $f_i = f \mid V_i$.

Theorem 22: There is a one-to-one correspondence between coherent families of linear maps $(f_i)_{i\in I}$ into a vector space W and linear maps $f: \sum_{i\in I} V_i \to W$ determined by $f_i = f \mid V_i$.

Proof: Clearly, if f is linear, so is each f_i, and the family $(f_i)_{i\in I}$ is coherent. For the converse, suppose $a \sim b$. Since we have $a(i) = b(i)$ for $i \neq j, k$ while

$$a(j) + x = b(j) \quad \text{and} \quad a(k) - x = b(k), \qquad x \in V_j \cap V_k,$$

we also have

$$\begin{aligned}\sum_{i\in I} f_i(a(i)) &= \sum_{i\neq j,k} f_i(a(i)) + f_j(a(j)) + f_k(a(k)) \\ &= \sum_{i\neq j,k} f_i(b(i)) + f_j(b(j) - x) + f_k(b(k) + x) \\ &= \sum_{i\in I} f_i(b(i)) - f_j(x) + f_k(x) \\ &= \sum_{i\in I} f_i(b(i)),\end{aligned}$$

since $x \in V_j \cap V_k$ and the family of maps is coherent. Thus we may define $f(A)$ as $\sum_{i\in I} f_i(a(i))$ for any $a \in A$, since by the use of a chain we see that the value obtained is the same. Obviously, f is linear, and $f \mid V_i = f_i$.

Notation: We write $\sum_{i\in I} f_i$ for the linear map generated by the family $(f_i)_{i\in I}$.

3. Positive Linear Functionals

The scalars are now assumed to be the reals. Suppose that each V_i contains a convex cone C_i with vertex at the origin, and let C be the convex hull of the union $\bigcup_{i\in I} C_i$ in $\sum_{i\in I} V_i$.

Lemma 7: The set C is a convex cone, and $c \in C$ iff $c = \sum_{i\in I} c_i$, with $c_i \in C_i$.

Proof: The set $C' = \{\sum_{i\in I} c_i \mid c_i \in C_i\}$ is a convex cone containing all C_i; furthermore, each element of the form $\sum_{i\in I} c_i$ is in C, since it has the form

$$\sum_{i\in I} \lambda_i[(1/\lambda_i)c_i],$$

with $\lambda_i > 0$ and $\sum \lambda_i = 1$. Thus $C' = C$.

Notation: We write $\sum_{i\in I} C_i$ for C.

Theorem 23: A functional $\sum_{i\in I} f_i$ is positive relative to $\sum_{i\in I} C_i$ iff each f_i is positive relative to C_i.

Proof: Since

$$\left(\sum_{i\in I} f_i\right)\left(\sum_{i\in I} c_i\right) = \sum_{i\in I} f_i(c_i),$$

positivity of the f_i implies that of $\sum_{i\in I} f_i$. Conversely, since

$$f_j = \left(\sum_{i\in I} f_i\right)\Bigg| V_j \quad \text{and} \quad C_j \subseteq \sum_{i\in I} C_i,$$

positivity of $\sum_{i\in I} f_i$ implies that of f_j.

Lemma 8: Suppose that $E \in \bigcap_{i\in I} C_i$, and that each C_i is radial at E. Then $\sum_{i\in I} C_i$ is also radial at E.

Proof: Consider any $A, a \in A$ and let $a(i) = 0$ except for $i = i_1, i_2, \cdots, i_n$. Now E, as an element of $\sum_{i\in I} V_i$, is represented by any map $e: i \to \delta_{ij}E$ (any j). On each segment from E to $a(i)$, there exists by hypothesis an element $b(i) \in V_i$ which is in the cone C_i. Let B be the class of the map

$$b: i \to b(i), \quad \text{for} \quad i = i_1, i_2, \cdots, i_n,$$
$$b: i \to 0, \quad \text{for all other } i.$$

Since $b(i) \in C_i$, we have $B \in \sum_{i\in I} C_i$. Let $D = tE + (1 - t)B$ so that D is represented by the map

$$d: i \to tE + (1 - t)b(i), \quad \text{for} \quad i = i_1,$$
$$d: i \to (1 - t)b(i), \quad \text{for} \quad i = i_2, i_3, \cdots, i_n,$$
$$d: i \to 0, \quad \text{for all other } i.$$

Since all $d(i) \in C_i$, we have $D \in \sum_{i\in I} C_i$, which means that the cone is radial at the point E.

Without any restriction on the cones C_i, there may not exist coherent families of positive functionals on the family $(V_i)_{i\in I}$. It is clear that, in case $\sum_{i\in I} C_i$ has a radial point, positive nonzero functionals exist iff

$$\sum_{i\in I} C_i \neq \sum_{i\in I} V_i$$

(Theorem 3.2 of Ref. 6).

Theorem 24: Suppose that each C_i is radial at

$$E \in \bigcap_{i\in I} V_i$$

and that, for each $x \in V_i$, $x \notin C_i$, there exists a positive functional f on V_i such that $f(x) < 0$. Then the positive (relative to C_j) functionals on V_j extend to positive (relative to $\sum_{i\in I} C_i$) functionals on $\sum_{i\in I} V_i$ iff

$$\left(\sum_{i\in I} C_i\right) \cap V_j = C_j.$$

Proof: Suppose that this condition is satisfied, i.e., that the order induced on V_j as a subspace of $\sum_{i\in I} V_i$ is the original order. Then any functional on V_j, positive on C_j, is also positive relative to $\sum_{i\in I} C_i$. Since $E \in V_j$ and the cone $\sum_{i\in I} C_i$ is radial at E, any such functional extends to a functional on $\sum_{i\in I} V_i$, positive relative to $\sum_{i\in I} C_i$ (Theorem 3.3 of Ref. 6). Conversely, suppose that all such functionals extend, and take $\sum_{i\in I} c(i)$ in V_j but not in C_j; there exists a functional f on V_j which is positive and such that $f(\sum_{i\in I} c(i)) < 0$. But f extends to a positive functional on $\sum_{i\in I} V_i$ relative to $\sum_{i\in I} C_i$; hence

$$f\left(\sum_{i\in I} c(i)\right) \geq 0,$$

which is a contradiction.

It is not hard to find cases where the condition in the above theorem does not hold.

APPENDIX B: EXTENSION OF POSITIVE FUNCTIONALS

Let C be a cone in a real vector space V and $\leq$ the corresponding partial ordering. Assume that V has a unit $I \in C$, i.e., that for each $A \in V$, there is a constant $k > 0$ such that $-kI \leq A \leq kI$.

Lemma 9 (Segal): Suppose that W is a subspace of V containing the unit I and that f is a positive functional on W. Then, for any $x \notin W$, there exists a positive extension g of f to the subspace spanned by W and x.

Proof: Let $a = \inf\{f(y) \mid x \leq y \in W\}$ and $b = \sup\{f(z) \mid x \geq z \in W\}$. As I is a unit and belongs to W, both a and b are finite; also, since we have $z \leq x \leq y$, we have $b \leq a$. Now choose any c between a and b, and set $g(w + \lambda x) = f(w) + \lambda c$; g is obviously linear on $W + \{x\}$. Suppose that we have $w + \lambda x \geq 0$. In case $\lambda > 0$, we have $x \geq -(1/\lambda)w$, and hence $f(-(1/\lambda)w) \leq b \leq c$. In case $\lambda < 0$, we have $x \leq -(1/\lambda)w$ and hence $f(-(1/\lambda)w) \geq a \geq c$, so that in both cases, we have $f(w) + \lambda c \geq 0$, or $g(w + \lambda x) \geq 0$.

APPENDIX C: EXISTENCE OF FUNCTIONS

The following lemma is needed for the proof of Theorem 25.

Lemma 10: Given an interval U, a continuous function f, and a continuous function h, with compact support K such that $0 \leq h \subset f^{-1}U$, let

$$k(y) = \sup \{h(x) \mid f(x) = y\}.$$

Then, if a is on the boundary of U, we have $\lim_{y \to a} k(y) = 0$.

Proof: Otherwise there exists a sequence y_n converging to a and a $\delta > 0$ such that $k(y_n) > \delta$ for all n. Take $x_n \in f^{-1}y_n$, with $h(x_n) > \delta$ and a subsequence $x_{n(k)}$ converging to some x_0 (which is possible since K is compact). Then $h(x_{n(k)})$ tends to $h(x_0)$, and $f(x_{n(k)})$ tends to $f(x_0)$; this implies that $a = f(x_0)$ and $h(x_0) \geq \delta$. But $a \notin U$, hence $x_0 \notin f^{-1}U$ (because f is single-valued), and thus $h(x_0) = 0$, which is a contradiction.

Theorem 25: Under the hypotheses of Lemma 10, there exists a continuous g such that $0 \leq g \subset U$ and $g \circ f \geq h$.

Proof: Take a sequence ϵ_n converging to zero through positive values which are less than unity. There exists a δ_1 such that, if $|y - a| < \delta_1$, then $k(y) < \epsilon_1$. Join the point $(a + \frac{1}{2}\delta_1, \epsilon_1)$ to the point $(a + \delta_1, 1)$ by a segment $s_1(y)$. Then, in the interval $(a + \frac{1}{2}\delta_1, a + \delta_1)$, we have $s_1(y) \geq k(y)$. There exists now a $\delta_2 < \frac{1}{2}\delta_1$ such that, if $|y - a| < \delta_2$, then $k(y) < \epsilon_2$. Join the point $(a + \frac{1}{2}\delta_2, \epsilon_2)$ to the point $(a + \delta_2, \epsilon_1)$, and the point $(a + \delta_2, \epsilon_1)$ to the point $(a + \frac{1}{2}\delta_1, \epsilon_1)$ to obtain $s_2(y)$, which is thus $\geq k(y)$. Proceed in this way to obtain a numerical sequence δ_n, converging to zero, and a sequence of polygonal lines $s_n(y)$ over the intervals $(a + \frac{1}{2}\delta_n, a + \frac{1}{2}\delta_{n-1})$, such that $s_n(y) \geq k(y)$ in each interval. Join the lines together to obtain $g(y)$ in a neighborhood of a. Clearly, g is continuous, $g(a) = 0$, $g(a + \delta_1) = 1$, and $g(y) \geq k(y)$. Define g similarly at the other end of U, and set it equal to one in between and equal to zero outside U.

Theorem 26: Let K be a compact set on the real axis containing both positive and negative points. Then there exists a function f, continuous on K, and positive numbers c and d, such that $-1 \leq f(x) \leq 0$ for $x \in K$, $-1 \in f(K)$, and $f(x) + 1 - d \geq cx$ for $x \in K$.

Proof (after Sherman[8]): Let x_1 and x_2 be the infimum and the supremum of K, so that $x_1 < 0 < x_2$. Let $F_1 = (-\infty, \frac{3}{4}x_1] \cap K$ and $F_2 = [\frac{1}{2}x_1, +\infty) \cap K$; these are closed and disjoint sets. Consider a function f, continuous on K, and such that $f \mid F_2 = 0$, $f \mid F_1 = -1$ and $-1 \leq f(x) \leq 0$ for $x \in K$ (Tietze's theorem), which satisfies the first two requirements. Let $d = \min(\frac{1}{2}, -x_1/4x_2)$, and take $x \in F_2$. Then $f(x) + 1 - d = 1 - d \geq x/2x_2$ for all such x. For $x \notin F_2$, we have $f(x) + 1 - d \geq -1 + 1 - d \geq x_1/4x_2$; but $x \in K$, $x \notin F_2$ implies $x < \frac{1}{2}x_1$, so that we can take $c = 1/2x_2 > 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 11 NOVEMBER 1969

Erratum: Strong-Coupling Limit in Potential Theory. II

[J. Math. Phys. **9**, 1890 (1968)]

WILLIAM M. FRANK
Bar-Ilan University, Ramat Gan, Israel
and
U.S. Naval Ordnance Laboratory, Silver Spring, Maryland

(Received December 1968)

When the above article was published, the by-line was incomplete. It should read as above.